8.01 Overview and Introduction

Christian P. Whitman, The University of Texas, Austin, TX, USA

8.01.1 Introduction

Enzymes are amazing biological catalysts, which are characterized by their enormous rate accelerations and stringent substrate specificities. A compendium of estimated rate accelerations shows that the enzymatic reactions are 10^6-fold (carbonic anhydrase) to 10^{21}-fold (fructose-1,6-bisphosphatase) faster than the nonenzymatic reactions.[1,2] It is also impressive that most enzyme-catalyzed reactions take place in an aqueous environment at neutral pH values. Others proceed in extreme environments, that is, at high or low pH, in high salt concentrations, and/or at high or low temperatures. In addition to advancing our knowledge of basic biological processes, understanding how enzymes work and capitalizing on this knowledge impact medicine (drug design, development, and diagnostics), agriculture (herbicides and pesticides), industry (biocatalysts, biofuels, and household products), and the environment (bioremediation).

Since the publication of the first volume of *Comprehensive Natural Products Chemistry* in 1999, our understanding of enzyme mechanisms has advanced significantly. We can now dissect and describe mechanisms in a level of detail not previously accessible. This is due to a combination of factors (and confluence of events) including advances in instrumentation, enhanced computing power, and improvements in recombinant DNA technology coupled with ease of purification. Site-directed and random mutagenesis experiments are now more frequently done using kits so that the identification of critical residues (in and outside the active site) is fast. The availability of so many genomes and genetic contexts has facilitated mechanistic and structural studies, functional assignments, and an understanding of how enzymes evolved. Crystal structures of soluble enzymes, which initially required heroic efforts and extraordinary luck, have almost become so routine that a crystal structure is sometimes considered to be part of the characterization process along with the kinetic parameters. Although enzymes are largely studied *in vitro*, we can sometimes put these mechanisms in the context of the cell and examine how the mechanisms (and properties) are modulated by protein–protein interactions and allosteric molecules. Finally, single-molecule enzymology has come of age so that we can measure the dynamic behavior of individual molecules in real time.

We are now poised to tackle more challenging problems and uncover unprecedented biochemistry and biocatalysts in nature. These reactions are frequently found in biosynthetic pathways of natural products in plants or exotic microorganisms. A limited understanding of the genetics of these organisms makes the characterization of the enzymes comprising the pathways challenging. A second challenge will be the mechanistic and structural characterization of large, and sometimes 'floppy', macromolecular enzyme complexes. Membrane-bound enzymes also pose a challenge. A third major challenge will be the study of enzymes in the cell. Such studies have implications for how enzymes function in their natural environment and enzymes that cannot be isolated. Finally, bioinformatic and computational studies will likely flourish and perhaps make the assignment of function to unknown gene products less problematic.

The chapters in this volume showcase representative examples of the advances in many areas. The major questions are summarized, discussed, and sometimes answered. Future areas of interest are also outlined. The work discussed in these chapters generally builds on and expands classical enzymological techniques, and shows just how remarkable and versatile enzymes are. However, it will become clear that much remains to be discovered in the area of enzyme mechanisms, and the answers to so many questions are still unknown.

8.01.2   Chapter Summaries

The first two chapters discuss the evolution of enzymes. Copley traces the evolution of specific and efficient modern enzymes from their rudimentary beginnings (the so-called generalist stage), whereas Khersonsky and Tawfik focus on the role of promiscuity in the evolution of new enzyme activities. Enzymes with broad specificity and poor efficiency characterized the generalist stage. As organisms increased in complexity, new folds were explored and these basic scaffolds became adorned with adjoining domains and coenzymes. This process resulted in specific and efficient enzymes whose activities could be regulated. Selective pressures on these catalysts from the cellular and external environments contributed to their catalytic efficiency, stringent specificity, and exquisite regulation. Most recently, the environmental presence of anthropogenic compounds prompted bacteria to assemble pathways and evolve enzymes so that these compounds could be used as carbon sources. In addition to the obvious intellectual interest in how enzymes evolved, an understanding of this process has ramifications for efforts to develop biocatalysts.

Catalytic promiscuity, an enzyme's ability to catalyze low-level activities in addition to its physiological one, has emerged as a fairly widespread trait that challenges the 'one enzyme one substrate' dogma. In this regard, the chapter by Khersonsky and Tawfik makes an important contribution to our understanding of this phenomenon and its ubiquity. The authors cite some early accounts of such activities, but they indicate that these activities were viewed more as isolated incidences or idiosyncrasies. The importance of catalytic promiscuity to the evolution of new activities is now being recognized and cataloged.

One of the major questions in this area is how the same active site and catalytic machinery can be so specific for its physiological substrate but yet still be promiscuous. The authors outline five scenarios to address this question including conformational plasticity, the same but suboptimal interactions, different protonation states, different subsites, alternative cofactors, and the assistance of water. Various models are also discussed to explain how low-level catalytic activities can spawn new activities.

The next six chapters focus on enzymes that are involved in the transformation of small molecules. The series begins with a chapter summarizing catalytic strategies used by bacterial dehalogenases to break a carbon–halogen bond in mostly anthropogenic compounds. The following chapter describes three classes of guanidine-modifying enzymes, hydrolases, dihydrolases, and amidinotransferases, found in the pentein superfamily. The focus of the chapter by Richards *et al.* is the movement of the ammonia from the glutaminase site to a second site in various enzymes through a molecular tunnel, and the origin and mechanics of these molecular tunnels. Tonge *et al.* outline three sets of enzyme-catalyzed reactions, that is, carbon–carbon bond formation and cleavage, oxidation–reduction, and hydration–dehydration reactions, in the context of fatty acid biosynthesis and catabolism. Solanapyrone synthase, the first identified 'Diels–Alderase', along with two more recently discovered Diels–Alderases and several suspected ones is the subject of the chapter by Oikawa. The series ends with a chapter by Brandao and Hengge that describes phosphatases and sulfatases, which are responsible for the transformations of phosphate and sulfate esters, respectively.

The chapter by Polearends and Whitman updates our current understanding of the catalytic strategies used by microbial dehalogenases. Structural and mechanistic studies have uncovered three major strategies: intra-molecular displacement of the halide by the substrate hydroxyl group; ester formation between the enzyme and substrate (to release the halide) followed by ester hydrolysis: and the hydrolytic dehalogenation of the substrate. The last strategy is the most recently discovered one and is used by *cis*- and *trans*-3-chloroacrylic acid dehalogenase. Interestingly, both enzymes are tautomerase superfamily members, and use similar mechanisms to activate a water molecule for a 1,4 conjugate addition to the substrate. The resulting species collapses to release the halide and product. The structural homology and the conservation of key functional groups suggest that the two enzymes diverged from a common ancestor.

The pentein superfamily members share a similar fold, but are diverse in sequence and function. The proteins in this superfamily are grouped into noncatalytic and catalytic ones, the latter functioning as the guanidine-modifying enzymes. These enzymes modify free arginine as well as arginine found in proteins (and other guanidine derivatives). The resulting products can impact several biological processes and signaling pathways.

The functions of the guanidine-modifying enzymes further subdivide the family into three distinct groups: hydrolases, dihydrolases, and amidinotransferases. Hydrolases catalyze the hydrolysis of guanidine derivatives to form ureido compounds, whereas dihydrolases catalyze a hydrolysis reaction to yield a primary amine, ammonia, and bicarbonate (or carbon dioxide). The amidinotransferases transfer an amidino group from one substrate to an amine. Although these are distinct reactions, they are characterized by common structural and mechanistic themes.

The catalytic mechanism of L-arginine:glycine amidinotransferase (AGAT) is a representative example. AGAT catalyzes the first step in the synthesis of creatine, which along with phosphocreatine, is an important energy reservoir in the body. The catalytic machinery is the same as that found in the hydrolases, but the substrate is repositioned in the active site so that a different C^ζ–N bond is broken. An 'activated urea' covalent intermediate is formed and used in the subsequent reaction with glycine.

Glutamine-dependent amidotransferases catalyze the hydrolysis of ammonia from the amide group of glutamine and transport it via a molecular tunnel to a second site where the ammonia is used in a carbon–nitrogen bond formation reaction. The tunnel prevents the release of ammonia into the cellular environment. The cellular concentrations of free ammonia (and other small molecule metabolites) must be tightly controlled due to its toxicity. The transport tunnel represents one metabolic strategy to control the utilization of potentially toxic molecules or reactive intermediates. Although an effective strategy, Richards *et al.* note that it raises questions about the coordination of activities at the different sites, the transfer of ammonia through the tunnel, and the molecular events that led to the evolution of these tunnels. The chapter summarizes the current thinking about answers to these questions within the framework of the glutamine-dependent amidotransferase group of enzymes.

The synthesis and breakdown of fatty acids involves a cyclic series of reactions where the enzymes catalyzing these reactions fall into three major superfamilies: the thiolase superfamily (synthases and thiolases), the short-chain dehydrogenase reductase superfamily (oxidoreductases), and the crotonase superfamily (hydratases and dehydratases). The enzymes comprising these three superfamilies include a significant portion of those found in microbial and mammalian metabolism. Many of the enzymes in microbial metabolism are potential drug targets.

One striking example is the *Mycobacterium tuberculosis* enoyl-acyl carrier protein (ACP) reductase (InhA), which catalyzes the last reaction in fatty acid elongation. This enzyme is the target of isoniazid, one of the major drugs used for treatment of tuberculosis. Isoniazid is activated by KatG, a mycobacterial catalase–peroxidase, to a species that reacts with the NAD^+ coenzyme of InhA. The resulting adduct is a potent inhibitor of InhA. Efforts are under way to make more potent inhibitors that do not require KatG activation.

Since the discovery of the first 'Diels–Alderase' in the biosynthesis of the phytotoxin solananpyrone, two more have been identified and characterized, lovastatin nonaketide synthase and macrophomate synthase, and many more are suspected. The Diels–Alder reaction, the reaction between a 1,3-diene and an alkene to form a six-membered ring, is an important reaction in organic synthesis due to its versatility and stereoselectivity. With appropriately substituted dienes and alkenes, the Diels–Alder reaction can produce a variety of substituted ring systems. The reaction does not generally require catalysis, so the rate enhancement is achieved by entropic factors.[3] The macrophomate synthase crystal structure provides an insight into how this phytotoxin might be generated in the active site and how catalysis occurs.[3] The enzyme converts the substrates into the Diels–Alder substrates, which are not released from the active site, but are forced into a reactive conformation that maximizes overlap of the π-orbitals. The interactions between substrate and enzyme are mostly hydrogen bonds and electrostatic contacts. With the characterization of more Diels–Alderases, it will be interesting to compare and contrast strategies.

The hydrolysis of phosphate and sulfate monoesters is critical for all living organisms and is catalyzed by phosphatases and sulfatases, respectively. In addition to their fundamental importance, these enzymes produce rate accelerations that are among the most prodigious. Moreover, despite the resemblance between the phosphoryl and sulfuryl groups and the mechanistic similarity in the reactions of phosphate and sulfate monoesters, very distinct and highly specific enzymes have evolved to carry out these two reactions. The phosphatase mechanisms are generally understood, whereas the sulfatase mechanisms are not clearly understood. Much, however, remains to be learned about the biological roles of these enzymes.

The next three chapters discuss the enzymology of two of the three major biological polymers, that is, DNA and polysaccharides. The first chapter considers the mechanistic and structural basis for DNA polymerase selectivity, which is fundamental for the accurate transfer of genetic information and organism survival. The following chapter discusses glycosyltransferases, which are responsible for the degradation and synthesis of glycosides, and as such play a number of important roles. The final chapter in the series describes the structure, biosynthesis, regulation, and secretion of alginate, a polysaccharide that forms a protective capsule around various bacteria.

DNA-dependent polymerases are responsible for the synthesis of new DNA from deoxyribonucleotide triphosphates and an existing DNA template. It is well known that this process is critical for the preservation and transfer of an organism's genetic information. The accuracy (or fidelity) of DNA polymerases ranges from low (10^1) to very high (10^6) and has consequences for the organism's survival. The origin of this fidelity has been the object of much study. Tsai *et al.* note that the free energy difference between correct and incorrect base pairing is estimated to be 1–3 kcal mol^{-1}. This difference translates into a fidelity of 10^1–10^2 if the polymerase were to have no selectivity. Hence, DNA polymerases must possess an ability to enhance the selectivity. The chapter reviews the structural, kinetic, and computational studies that have been carried out in an attempt to understand the mechanistic origin of the enhanced selectivity.

In the broadest sense, glycosyltransferases are defined as enzymes that catalyze the transfer of glycosyl residues from a donor to an acceptor. If the acceptor is a water molecule, the enzyme is defined as a glycosidase. The chapter by Withers *et al.* provides an account of the reaction mechanisms for glycosidases and, more generally, glycosyltransferases, based on the accumulated structural and mechanistic work. The glycosidases fall into two categories: those that result in retention of the substrate's anomeric configuration and those that result in inversion. The observed stereochemical outcomes result from single or double displacement mechanisms that involve acid/base chemistry and oxocarbenium ion-like transition states.

In contrast, the mechanisms for the remaining glycosyltransferases are not as well understood. A better understanding of the mechanisms is critical because these enzymes play a central role in antibiotic production, attaching the carbohydrate moiety to the aglycone of a natural or nonnatural antibiotic molecule. The enzymes can be classified as retaining or inverting and show one of the two observed folds. However, the fold and the stereochemical outcome are not correlated. Thus far, the inverting glycosyltransferases appear to use a single displacement mechanism that involves acid/base chemistry via an oxocarbenium ion-like transition states. The retaining glycosyltransferases are mechanistically more complex. Examples are discussed for all of these enzymes.

Alginate is a linear polysaccharide composed of two monosaccharides, β-D-mannuronate and α-L-guluronate. The relatively simple structure of alginate and the fact that the genes for its biosynthesis have been identified make it a model system for studying the complex process by which polysaccharides are made, modified, and secreted. One of the more interesting enzymes in the biosynthesis of alginate is mannuronan epimerase, which converts β-D-mannuronate to α-L-guluronate in the alginate polymer. A major mechanistic question is how the enzyme abstracts the C5 proton, which is adjacent to a carboxylate group and has a pK_a value estimated to be greater than 30. Evidence gathered to date suggests that the enzyme forms a glycal with unsaturation at the C4–C5 position concomitant with transient cleavage of the glycosidic bond.

An intervening chapter delineates the enzymatic basis for the bacterial resistance to five classes of antibiotics: the β-lactams (e.g., penicillin), the glycopeptides (e.g., vancomycin), the aminoglycosides (e.g., streptomycin), the macrolides (e.g., erythromycin), and the quinolones (e.g., ciprofloxacin). The classes represent the major antibiotics used clinically today. Starting with the β-lactams and ending with the fluoroquinolones, the authors discuss the mechanisms and molecular basis for resistance, and show how resistance is not a matter of if, but when.

The three major mechanisms for antibiotic resistance are well known and involve the alteration of the target, the modification of the drug (e.g., hydrolysis of the β-lactam ring), or the exportation of the compound from the cell. Resistance to these compounds was observed shortly after their introduction. Much effort has been invested in deciphering these mechanisms for each class of antibiotics.

One example is the bacterial response to the challenge presented by the glycopeptide vancomycin – the bacteria simply altered the target. As noted by the authors, vancomycin is a clamp that binds to the D-alanine–D-alanine terminus of the peptide stem of a cell wall precursor, thereby preventing its incorporation into the cell

wall. Changing the D-alanine–D-alanine dipeptide to a D-alanine–D-lactate moiety decreases the affinity and renders the drug less effective. Because the D-alanine–D-alanine tail is removed in the cross-linking reaction, the change is inconsequential to the bacteria.

Four chapters are devoted to the roles of metal ions and radicals in enzyme-catalyzed reactions. The first chapter discusses copper-containing enzymes, which facilitate a number of redox reactions, frequently activating oxygen for insertion into a C–H bond. Hernick and Fierke describe the mechanisms for a wide variety of metal-dependent hydrolases with an emphasis on common catalytic themes. Nonheme iron-dependent dioxygenases are a fascinating family of enzymes responsible for the oxidative cleavage of the highly stable aromatic carbon bond. The final chapter in this series discusses the radical S-adenosyl methionine (SAM) superfamily, focusing on the structure and properties of the iron–sulfur clusters and how these properties relate to catalytic activity.

Binuclear copper proteins are involved in many critical biological processes. How copper mediates electron flow in enzymatic reactions and the identity of the copper oxygen species has been studied intensely, as detailed in the chapter by Messerschmidt. One of the best-studied copper-containing enzymes is dopamine β-monoxygenase, which catalyzes the conversion of dopamine to norepinephrine, and is representative of the copper hydroxylases. The enzyme uses two copper atoms in catalysis – one serves as a site for O_2 binding and activation and the other atom functions as an electron transfer site. A copper (II) superoxide species is proposed to be the activated form of oxygen that abstracts the hydrogen atom from dopamine. The formation of the product results in a two-electron oxidation of dopamine, a four-electron reduction of oxygen to water, and the incorporation of one atom of oxygen at the benzyl carbon of dopamine.

Metal-dependent hydrolases are critical in protein, carbohydrate, and nucleotide metabolism. In addition to the intellectual contributions, there is much interest in these hydrolases because many of them are potential targets for drug intervention. Hernick and Fierke detail the physical and chemical properties that make metals suitable for hydrolysis reactions including their Lewis acidity, electronic configuration, and exchange properties. Although a number of metal ions meet these criteria, zinc is most commonly found in metal-dependent hydrolases. Along with amino acid side chains that participate in proton transfer reaction and stabilize reaction intermediates, the presence of the metal ion facilitates the hydrolytic reaction.

Histone deacetylases (HDACs) are representative and timely examples of metal-dependent hydrolases. Histone acetylation regulates gene transcription making HDACs potential targets for anticancer agents. As a result, their mechanisms have received much scrutiny. The enzymes are divided into four classes where class I, II, and IV are metal-dependent hydrolases. The zinc ion-binding residues are conserved across the three classes. The mechanism involves the coordination of the substrate to the metal ion (for polarization of the carbonyl group of the acetyl moiety) followed by general base activation of the metal-coordinated water molecule for attack at the carbonyl carbon. The resulting tetrahedral intermediate is stabilized by the metal ion. General acid protonation facilitates the breakdown of this intermediate and leads to product release. More recent work suggests that the reaction might proceed through a single bifunctional general acid–base catalyst.

Nonheme iron-dependent dioxygenases, the first major topic of the chapter by Bugg, mediate a large number of oxidative carbon–carbon bond cleavages in bacterial, mammalian, and plant pathways. The dioxygenases involved in the oxidative cleavage of catechol have been intensely studied. Degradation pathways for aromatic compounds generally converge on this substrate or structurally similar ones, and are processed by an intradiol or an extradiol dioxygenase. The catalytic mechanisms of these enzymes along with the factors that control extradiol versus intradiol specificity are discussed in the chapter.

The other major class of nonheme iron-dependent dioxygenases is the α-ketoglutarate-dependent dioxygenases. These enzymes catalyze the oxidative decarboxylation of an α-keto acid to form succinate (when using α-ketoglutarate) and an iron (IV) oxo species, which is then used to carry out a hydroxylation reaction. These reactions are generally found in biosynthetic reactions (e.g., clavaminate synthase), but a few are found in catabolic pathways (e.g., p-hydroxyphenylpyruvate dioxygenase). The contrasts between these dioxygenases and those discussed above are highlighted in the chapter.

Enzymes in the radical SAM superfamily carry out a diverse set of reactions including sulfur insertion in biotin and lipoate synthases, DNA repair, tRNA modification, and are involved in the synthesis of molybdopterin, heme, thiamine, and many natural products. Shepard and Broderick discuss the common themes in these reactions and differentiating details in the chapter. Despite the diversity of the reactions, they all begin with a

series of steps that promotes homolytic S–C5′ bond cleavage and generates the 5′-deoxyadenosyl radical intermediate. The individual reactions then take separate mechanistic paths and give rise to the reactions listed above.

The final three chapters are devoted to useful methodologies in enzymology. In the first of these chapters, Furdui and Anderson describe techniques to identify and characterize short-lived intermediates in enzymatic reactions. In the following chapter, Frantom and Blanchard describe very recent work involving bisubstrate analogs. Bisubstrate or, more generally, multisubstrate analogs can be highly specific for an enzyme and are very useful for the differential inhibition of enzymes that share a common ligand such as ATP or coenzyme A.[4] Finally, Hopmann and Himo describe a powerful computational approach to model the active site of an enzyme and to use this model to evaluate different mechanistic scenarios.

Detection and characterization of an intermediate in an enzyme-catalyzed reaction can be especially challenging if the intermediate has a short lifetime or is labile. However, the identification of a putative intermediate is critical in establishing a reaction pathway. The chapter by Furdui and Anderson describes examples where rapid chemical quench and stopped flow techniques have been particularly helpful in defining the mechanism. Such was the case for 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase and UDP-GlcNAc enolpyruvyltransferase, two examples described in some detail by the authors. The former enzyme is involved in aromatic amino acid biosynthesis and is the target of the herbicide glyphosate (Roundup), and the latter enzyme makes a component for the peptidoglycan strand used in cell wall biosynthesis and is the target of the antibiotic fosfomycin. In both, kinetically competent intermediates were isolated and characterized.

In another example, the reaction catalyzed by 3-deoxy-D-*manno*-2-octulosonate-8-phosphate (KDO8P) synthase, the intermediate proved to be too labile to be isolated and characterized. For this, the authors turned to mass spectrometry, which offers two advantages over conventional chemical quenching: catalysis is terminated or quenched by desolvating the enzyme. In this way, a broader range of intermediates and enzyme complexes in the reaction can be identified. Electrospray ionization time-of-flight mass spectrometry also proved useful in delineating the phosphorylation kinetics of kinase reactions. In one example, the authors identified the sequential phosphorylation of the six tyrosine residues in fibroblast growth factor receptor by the appearance of the mono-, di-, and other multiphosphorylated species. The relative intensities of the species as a function of time provide the reaction rates.

The members of the Gcn5-related *N*-acetyltransferase (GNAT) superfamily catalyze the acetylation of a multitude of amine functional groups using acetyl coenzyme A (acetyl CoA). Acetylation plays important roles in diverse processes including gene expression by the modification of the ϵ-amino groups of specific lysine residues in histones and antibiotic resistance by the acetylation of aminoglycosides. Because these enzymes all use acetyl CoA, a high degree of specificity (and thus inhibition) for one enzyme can be achieved by combining an acetyl CoA analog with an individual substrate analog into a single molecule. If the resulting molecule, the bisubstrate analog inhibitor, proves to be a potent inhibitor, it can be used as an active site probe to characterize the enzyme's mechanism as well as a potential therapeutic agent.

The authors note that much has been learned about the structure and function of enzymes using multisubstrate analog inhibitors. However, a major frustration has been the inability to convert these compounds into clinically useful drugs. A major hurdle is the delivery of the compound to the active site of the target enzyme in the cell. Two approaches might circumvent this hurdle. One approach focuses on charge neutralization, as charged species do not seem to be readily taken up into the cell. The second approach is to mask the multisubstrate analog as a prodrug, which, after entering the cell, is processed to release the inhibitor.

Computational approaches to evaluate different mechanistic proposals for an enzyme have made great strides in the past 10 years. The chapter by Hopmann and Himo describe one such approach and its application to three different enzymatic reactions involving the transformation of an epoxide. The procedures and parameters to make a model of the active site are presented first and are followed by discussions of limonene epoxide hydrolase, soluble epoxide hydrolases, and haloalcohol dehalogenase. The results generally support the currently accepted mechanism for each enzyme but provide new insights into their regioselectivities.

The final enzyme discussed in the chapter, haloalcohol dehalogenase, catalyzes the formation of epoxides from halohydrins. However, the reverse reaction, the ring opening of epoxides with nonhalide nucleophiles such as cyanide, azide, and nitrite, has been exploited commercially to generate β-substituted alcohols that are potential synthons. For this reason, the regioselectivity of the enzymatic ring opening reaction is of

considerable interest. The enzymatic reaction is more complex than the one in solution because it is modulated by the active site environment and interactions between substrate and active site residues. The computational analysis of the regioselectivity of the reaction identified three elements (e.g., relative orientations of substrate and nucleophile, steric effects, electrostatic stabilization) in the active site that might be responsible for the observed outcomes. Subsequently, the contributions of these elements were evaluated by *in silico* mutagenesis of various active site groups.

References

1. R. Wolfenden; M. J. Snider, *Acc. Chem. Res.* **2001**, *34*, 938–945.
2. R. Wolfenden, *Chem. Rev.* **2006**, *106*, 3379–3396.
3. T. Ose; K. Watanabe; T. Mie; M. Honma; H. Watanabe; M. Yao; H. Oikawa; I. Tanaka, *Nature* **2003**, *422*, 185–189.
4. A. D. Broom, *J. Med. Chem.* **1989**, *32*, 2–7.

Biographical Sketch

Christian P. Whitman received his Ph.D. degree from the University of California, San Francisco (UCSF) in 1984 with Professor George L. Kenyon. At UCSF, he worked on the catalytic mechanism of mandelate racemase. This enzyme would later pique the interests of Professors John Gerlt, John Kozarich, Greg Petsko, and others, and expand our under-standing of divergent evolution to include enzymes that catalyze different reactions but are derived from a common ancestor. He then worked as a postdoctoral fellow with Professor Kozarich at the University of Maryland in College Park, where he tackled mechanistic and evolutionary questions about the enzymes comprising the β-ketoadipate pathway. He was appointed as an assistant professor in the Division of Medicinal Chemistry, College of Pharmacy at the University of Texas at Austin in 1987, and rose through the ranks to become Professor and Division Head in 1998. His research interests focus on the answers to two fundamental questions: how do enzymes work? and how do they evolve?

8.02 Evolution and the Enzyme

Shelley D. Copley, University of Colorado at Boulder, Boulder, CO, USA

8.02.1 Introduction

Enzymes catalyze almost every metabolic reaction in extant cells. A few unusually facile reactions, such as cyclization of L-glutamate γ-semialdehyde to form pyrroline-5-carboxylate in the proline biosynthesis pathway and decarboxylation of 2-amino-3-oxo-4-phosphonooxybutyrate in the pyridoxal phosphate (PLP) synthesis pathway, do not require acceleration to satisfy the demands of the cell. For all other reactions, catalysis is required because the rates of nonenzymatic reactions are very slow. Modern enzymes are marvelous catalysts. They accelerate reactions by up to 20 orders of magnitude,[1] prevent side reactions of reactive intermediates, and catalyze stereoselective and stereospecific reactions. Further, they are often exquisitely regulated by small molecule ligands.

Enzymes have played a key role in the adaptation of microbes to an extraordinary range of environmental niches and in the competition between species that has led to ever more efficient and well-regulated metabolic systems. The evolutionary processes that shaped the prodigious catalysts of extant organisms are of great interest to evolutionary biologists interested in understanding the origin and diversification of life, and to protein engineers interested in evolving novel catalysts for industrial uses.

Catalysts played an important role in the emergence of life on Earth nearly 4 billion years ago. Catalysis by mineral surfaces and small molecules enabled the emergence of a proto-metabolic network that, in turn, enabled the emergence of the RNA world. The first macromolecular catalysts may have been ribozymes, an idea first proposed by Carl Woese[2] that gained credence with the discovery of catalytic RNAs by Cech[3] and Altman.[4] Subsequently, ribozymes generated by *in vitro* evolution methods have been shown to catalyze a wide range of reactions involved in metabolism, including amino acid activation;[5] formation of coenzyme A (CoA), nicotinamide adenine dinucleotide (NAD), and flavin adenine dinucleotide (FAD)

from 4′-phosphopantetheine, nicotinamide mononucleotide (NMN), and flavin mononucleotide (FMN), respectively;[6] peptide bond synthesis;[7] and aldol condensation.[8] It is possible that ribozymes, in conjunction with catalytic auxiliaries such as metal ions, organic cofactors, and peptides, could have catalyzed all of the reactions required to both maintain metabolism and replicate genetic information.

This chapter will focus on catalysts in a later era, after the advent of genetically encoded proteins, and specifically on enzymes involved in metabolism in microbes. The earliest protein enzymes were probably generalists with broad substrate specificity[9] and consequently rather poor catalytic efficiencies. We can speculate about the properties of the earliest enzymes based upon reconstructions of the proteome of the last universal common ancestor (LUCA) and studies of extant generalist enzymes; the second section of this chapter addresses these issues. The third section describes mechanisms by which the structural diversity of enzymes seen in modern life arose. The fourth section discusses mechanisms by which enzymes with higher catalytic efficiency, stricter substrate specificity, and more sophisticated regulation emerged from inefficient progenitors. The fifth section examines how enzyme evolution is intimately tied to both the physiological conditions within cells and the environmental conditions, which are often changeable, in which microbes exist. The final section describes examples of enzyme evolution in the modern era caused by the introduction of anthropogenic compounds that exert new selective pressures on microbes.

8.02.2   The Earliest Enzymes: Getting the Basic Chemistry in Place

8.02.2.1   Enzymes in the LUCA

Proteins emerged before the LUCA, since rRNA and tRNA genes, as well as the genetic code itself, are conserved in all known forms of life. The emergence of the LUCA marks the point at which vertical transmission of genetic information became possible. Before the LUCA, life may have consisted of communities of proto-organisms that shared metabolites and genetic information.[10] It is difficult to know what the genetic and metabolic capabilities of these proto-organisms were, as we can obtain only a rather fuzzy picture of the LUCA itself. The LUCA clearly had DNA, ribosomes, proteins, and a well-developed metabolic network. Variability in gene structure and strategies for DNA replication in the three kingdoms of life suggests that replication and transcription processes were refined after the LUCA. Further, the structure of the cell wall had not yet been firmly established in the LUCA, as different strategies were ultimately adopted in different kingdoms of life. Bacteria and eukaryotes build their membranes from fatty acids linked to *sn*-glycerol-3-phosphate through an acyl linkage, whereas Archaea build their membranes primarily from isoprenoid lipids linked to *sn*-glycerol-1-phosphate through ether linkages.

Estimating the metabolic capabilities of the LUCA is a formidable problem because gene loss and gene gain, both by horizontal gene transfer and by emergence of new genes via gene duplication and divergence, have been rampant on an evolutionary time scale. **Figure 1** illustrates the difficulty. It is often possible to generate multiple scenarios that account for the presence or absence of genes in various lineages. In some scenarios, an ancestral gene may have been lost in some lineages, while in others, a gene that is not in fact ancestral may have spread by horizontal gene transfer between kingdoms.

Various authors using different data sets, algorithms, and assumptions about the frequencies of gene loss and gene gain have come to differing conclusions about the proteome of the LUCA. The number of universal
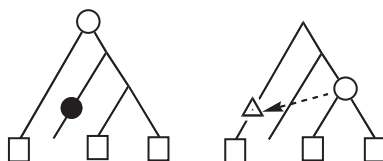
Figure 1   Patchy patterns of gene presence in extant organisms can obscure the evolutionary history of a gene and make it difficult to predict whether the gene was present in the LUCA and was subsequently lost in some lineages (left) or originated later and was transferred to another lineage (right). (○), first appearance of a gene; (●), gene loss; (△), gene acquired by horizontal gene transfer; (□), gene acquired from an ancestor.

protein domains found in all forms of life is estimated to be 190 by Abeln and Deane,[11] 219 by Lee *et al.*,[12] and 140 by Ranea *et al.*[13] Since these universal domains are primarily involved in transcription and translation, it has been suggested that the LUCA had only rudimentary metabolism, and relied upon amino acids and nucleotides supplied by a primordial soup.[13] This suggestion is at odds with the striking conservation of core metabolic pathways for the synthesis of amino acids, sugars, nucleotides, lipids, and cofactors in all domains of life. The core of metabolism must have been laid down by the LUCA, since it is inconceivable that nearly identical pathways would have emerged independently after the divergence of the three kingdoms of life.

Recently, more sophisticated analyses that deal explicitly with the impact of the frequencies of gene loss and gene gain on the predicted proteome of the LUCA have been developed. The results obtained depend on the topology of the tree and the value chosen for the relative frequency of gene gain and loss, as well as technical details of the algorithms. Thus, even the most sophisticated approaches produce results with some uncertainty. Nevertheless, interesting conclusions emerge. Mirkin *et al.* estimated that the LUCA contained 572 genes.[14] In a later study with a larger dataset, Ouzounis *et al.* estimated that the LUCA contained 669 genes.[15] Genes encoding enzymes for the synthesis of amino acids, nucleotides, sugars, fatty acids, and cofactors are predicted to have been present in the LUCA in both cases. It is quite remarkable that a few hundred enzymes capable of catalyzing reactions in core biosynthetic pathways emerged within the relatively short period between the end of the Late Heavy Bombardment about 3.9 billion years ago[16] and the appearance of life possibly as early as 3.8 billion years ago.[17]

8.02.2.2   Properties of Early Generalist Enzymes

The first enzymes were probably generalists that catalyzed similar reactions using a variety of substrates.[9] A relatively small number of generalist enzymes might have sufficed to support the metabolic network of the first proto-organisms. For example, extant pathways for the synthesis of guanosine monophosphate (GMP) starting from 5-phosphoribosyl-1-pyrophosphate (PRPP) involve 12 steps. The reaction shown in **Figure 2** (activation of a carboxylate via phosphoryl transfer from adenosine triphosphate (ATP), followed by attack of a nucleophilic amine) occurs 5 times. Similar reactions occur in other metabolic pathways, as well. Thus, a single nonspecific catalyst might have accelerated several physiologically significant reactions.

The hypothesis that the earliest enzymes were generalists is supported by the evolutionary history of PLP-dependent enzymes. PLP must have been available in the LUCA, since it is used by enzymes in all kingdoms of life. PLP facilitates a number of transformations of amino acids. Each of these reactions begins with a common step, attack of the amino group of a substrate upon an 'internal' aldimine formed between PLP and a lysine on the protein, resulting in formation of an 'external' aldimine. The cofactor serves as an electron sink for delocalization of electrons during cleavage of one of three bonds in the substrate. The orientation of the substrate determines the subsequent steps.[18] The bond to be cleaved is oriented by interactions with the active site to be orthogonal to the ring of the cofactor, allowing maximal overlap of the highest occupied molecular orbital (HOMO) of the labile bond and the lowest unoccupied molecular orbital (LUMO) of the cofactor. Based on the orientation of the substrate, the result can be transamination, decarboxylation, racemization, β–γ elimination, or β-replacement. Primordial PLP enzymes might have provided little more than a lysine to attach to PLP and a rudimentary binding site for substrates, which would have allowed acceleration of multiple
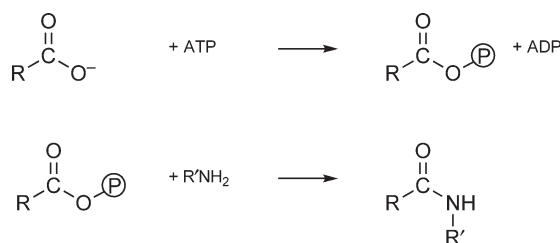

Figure 2   Activation of a carboxylate by formation of an acyl phosphate, followed by attack of an amine to form an amide, occurs 5 times in the 12 steps needed to synthesize GMP starting from PRPP.

reactions using a range of substrates. This hypothesis is consistent with the phylogenetic tree of the α-family of PLP enzymes,[19] which suggests that a nonspecific primordial enzyme diverged into reaction-specific subfamilies, and that substrate specificity emerged later within these subfamilies. Notably, many of these enzymes are found in all domains of life, suggesting that substantial divergence of PLP enzymes had occurred even before the LUCA.

Unfortunately, we cannot access specific sequence information about enzymes from the LUCA to allow resurrection of very ancient enzymes and evaluation of their properties. Phylogenetic reconstruction of ancestral protein sequences does not allow us to look back billions of years; beyond a certain point, there is too much ambiguity due to sequence divergence to allow a trustworthy estimation of the ancestral sequence.

A catalytic antibody (33F12) that catalyzes aldol additions and condensations provides a striking demonstration of the potential for catalysis of multiple reactions by a generalist enzyme.[20] A lysine in the active site of the antibody can form an enamine with several different aldehydes and ketones. This enamine can then react with several different aldehyde and ketone acceptors. Because the active site admits so many structurally different substrates, the antibody catalyzes over 100 different reactions with accelerations of 10^5–10^7 over the rate of the uncatalyzed reaction (see **Figure 3**). The structural basis for the broad specificity of the antibody was revealed by X-ray crystallography.[20] The catalytic lysine resides at the bottom of a deep hydrophobic pocket (see **Figure 4**) that accommodates structurally diverse substrates. Early generalist proteins may have used this strategy, providing binding interactions primarily for functional groups near the site at which chemistry needs to happen, while the rest of the substrate was allowed to protrude into the solvent or into a capacious cavity.
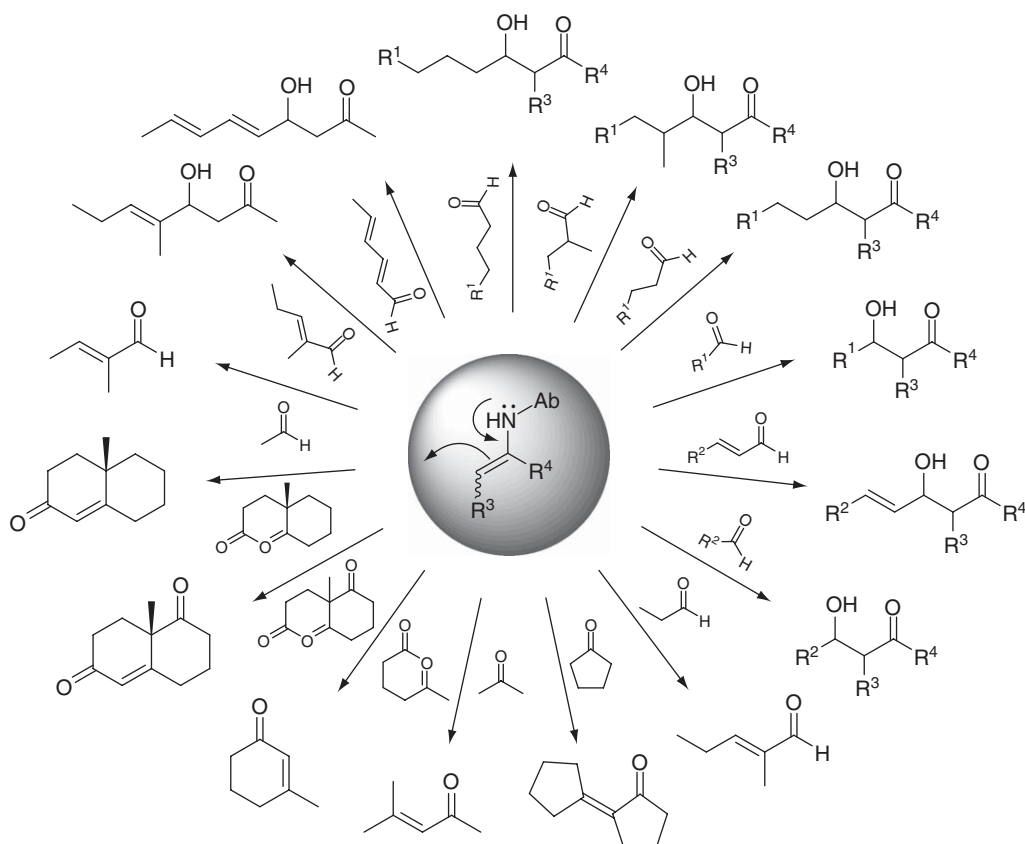

Figure 3 Broad substrate specificity in an antibody aldolase allows catalysis of over 100 different aldol additions and condensations. R[1], 4-acetamidobenzyl; R[2], 4-nitrobenzyl; R[3] and R[4], various alkyl groups. Reproduced with permission from C. F. Barbas, III; A. Heine; G. Zhong; T. Hoffman; S. Gramatikova; R. Björnestedt; B. List; J. Anderson; E. A. Stura; I. A. Wilson; R. A. Lerner, *Science* **1997**, *278*, 2085–2092.
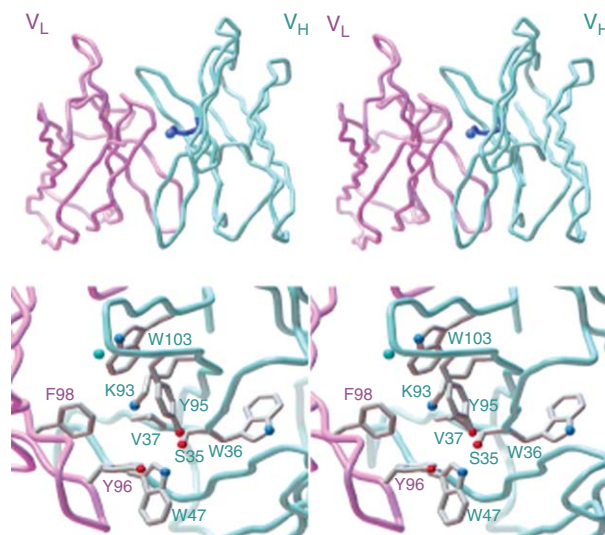
Figure 4 Stereo views of the active site pocket of the antibody aldolase 33F12. The light and heavy chains are shown in pink and blue, respectively. The catalytic lysine (K93) is shown in dark blue in the top panel. The bottom panel shows a view from the top of the protein. Reproduced with permission from C. F. Barbas, III; A. Heine; G. Zhong; T. Hoffman; S. Gramatikova; R. Bjornestedt; B. List; J. Anderson; E. A. Stura; I. A. Wilson; R. A. Lerner, *Science* **1997**, *278*, 2085–2092.

An alternative mechanism for achieving broad substrate specificity is to use a conformationally plastic region to fold around structurally different substrates. This strategy is used by microsomal P450 2B4. Crystal structures are available for the unliganded form of the enzyme and for the enzyme with two different ligands, 4-(4-chlorophenyl)imidazole (CPI) and bifonazole.[21] About two-thirds of the protein is similar in the two liganded complexes; these regions maintain the overall fold, the position of the heme, and the surface for binding electron transfer partners. However, the active site closes around the two ligands in strikingly different ways (see **Figure 5**). Of 10 residues in the active site with CPI, only 4 are in the active site with bifonazole. Five regions of the protein are particularly plastic. These 'plasticity regions' (labeled PR1–5 in **Figure 5**) show dramatically different conformations in the two enzyme–substrate complexes.
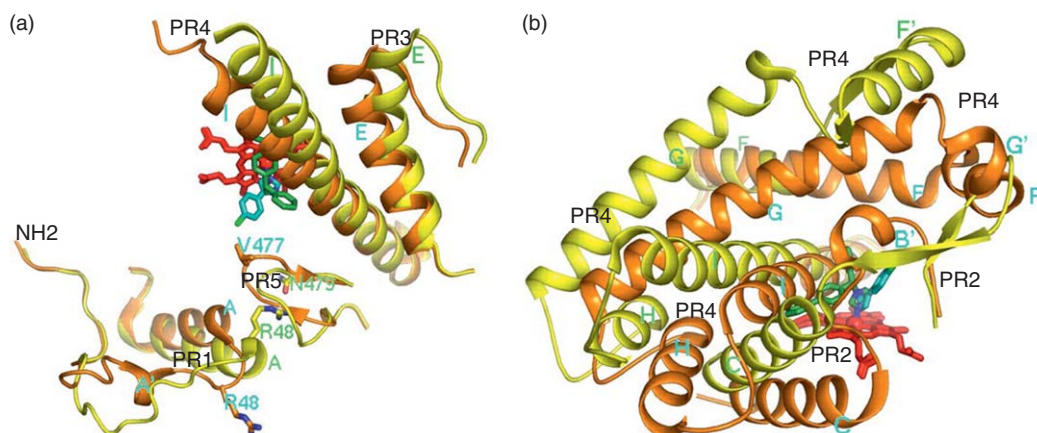

Figure 5 Two views of an overlay of structurally plastic regions of microsomal cytochrome P450 2B4 in complex with CPI (orange) and with bifonazole (yellow). The heme is shown in red, CPI in green, and bifonazole in cyan. Note that the ligands are bound in strikingly different orientations. Five 'plasticity regions' labeled PR1–PR5 show significantly different conformations in the two complexes. Reproduced with permission from Y. Zhao; M. A. White; B. K. Muralidhara; L. Sun; J. R. Halpert; C. D. Stout, *J. Biol. Chem.* **2006**, *281*, 5973–5981.

It is important to note that conformational flexibility can occur in the substrate as well as the enzyme. Flexibility can allow a substrate to be tucked into an active site; rotation around single bonds can position the reactive part of the substrate in proximity to catalytic groups, while allowing other parts of the molecule to avoid steric clashes. However, substrate flexibility could have been detrimental for early enzymes, as well. If a flexible substrate binds in an active site in a manner that doesn't orient the molecule properly with respect to the catalytic groups, a potential substrate becomes an inhibitor.

Although most enzymes involved in metabolism are quite specific, generalist enzymes still exist when broad specificity is advantageous. For example, broad specificity is important for detoxification enzymes such as glutathione *S*-transferases and cytochrome P450s because of the wide range of toxins to which organisms are exposed. The CYP3A subfamily, the most abundant P450 subfamily in human liver, is responsible for biotransformation of about 50% of the drugs metabolized in the liver. Kinetic constants for CYP3A4 and CYP3A5 have been collected for more than 50 substrates.[22] Values for k_{cat}/K_M vary enormously, from $20 \, \text{mol}^{-1}\text{s}^{-1}$ for N-dechloroethylation of ifosfamide[23] to $4.6 \times 10^6 \, \text{mol}^{-1}\text{s}^{-1}$ for oxidation of simvastin hydroxy acid to its 3′,5′-dihydrodiol metabolite[24] (both values for CYP3A4). Thus, generalist enzymes may accelerate many reactions, but do not necessarily provide efficient catalysis for all of them.

A rare example of a broad-specificity enzyme that catalyzes reactions in core metabolism is phosphoribosylisomerase A (PriA), an enzyme first identified in *Streptomyces coelicolor* and subsequently in *Mycobacterium tuberculosis*. This enzyme catalyzes two different Amadori rearrangements (see **Figure 6**) involved in histidine and tryptophan biosyntheses that are catalyzed by specialized enzymes in most organisms.[25,26] Kinetic parameters have not been reported for the two substrates. The assignment of function is based on genetic studies showing that deletion of *priA* in *S. coelicolor* creates a strain that is auxotrophic for both histidine and tryptophan. Furthermore, expression of PriA complements strains of *E. coli* that are auxotrophic for histidine and strains that are auxotrophic for tryptophan. It has been proposed that the broad-specificity enzyme might represent an ancestral 'generalist' state that persisted in certain lineages, but that in most lineages, gene duplication and subsequent divergence created specialized HisA and TrpF enzymes. This enzyme offers an unusual opportunity for investigation of the catalytic efficiency and regulation of a bifunctional enzyme that catalyzes similar reactions in two different metabolic pathways.

Early generalist enzymes evolved to generate families and superfamilies of specific enzymes via gene duplication followed by division of the original functions between the duplicate genes.[27] More specific enzymes were advantageous for several reasons. First, the cost of broad specificity can be the loss of catalytic power
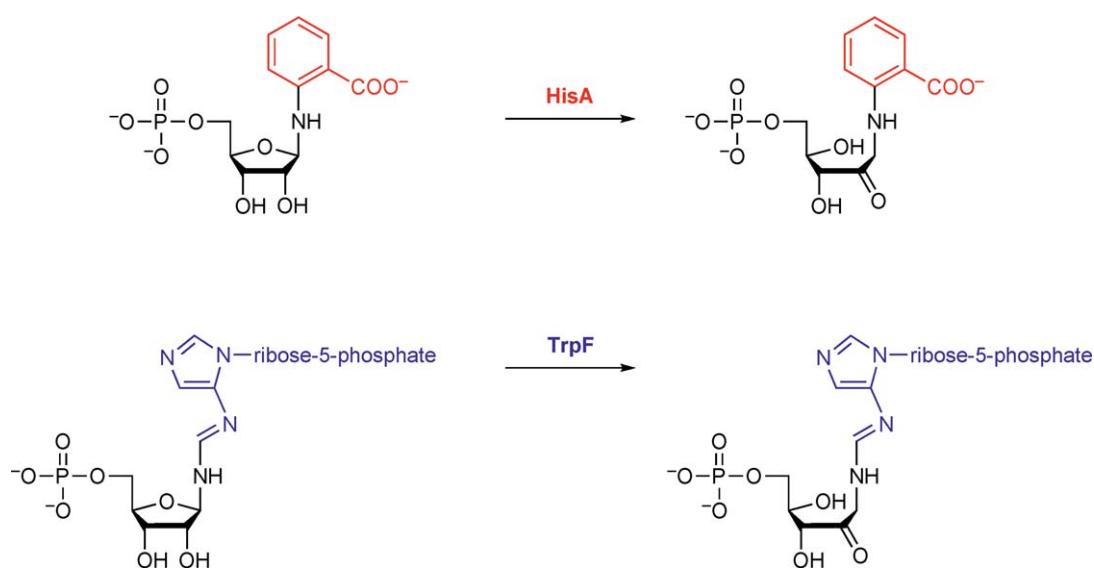
Figure 6   Amadori rearrangements catalyzed by HisA [*N*-(5′-phospho-L-ribosyl-formimino)-5-amino-1-(5′-phosphoribosyl)-4-imidazolecarboxamide isomerase] and TrpF (phosphoribosylanthranilate isomerase).

because substrates cannot be oriented optimally with respect to the active site machinery without a specific substrate-binding site. Broad substrate specificity also creates a potential for catalysis of undesirable reactions when certain substrates cannot be excluded from the active site. Furthermore, catalysis of reactions in more than one pathway by a generalist enzyme makes it difficult to optimize fluxes toward products that may be needed in very different quantities. Finally, each substrate of a broad-specificity enzyme is essentially an inhibitor of all other reactions catalyzed by the enzyme. These inefficiencies provided selective pressure for the evolution of more specific enzymes.

8.02.3   Exploring Fold Space

8.02.3.1   Protein Domains in the LUCA and in Extant Life

A substantial amount of protein fold space had been explored by the LUCA. As described above, the LUCA probably contained a few hundred genes, some of which may have encoded evolutionarily related proteins. Lee *et al.* estimate that the LUCA contained 219 protein families based upon the identification of homologous protein families found in at least 70% of the species in each kingdom.[28] Based upon structural data and a more stringent requirement that a domain be found in 90% of species from all three kingdoms and in at least 70% of Archaeal and eukaryotic species, Ranea *et al.* estimate that the LUCA contained 140 domains.[13] (Domains are defined as regions of continuous polypeptide chain that are compact semi-independent folding units.) It appears that the LUCA had discovered on the order of 200 domain structures. For comparison, the CATH[29] and SCOP[30] databases currently identify 1084 and 1086 domains among structurally characterized proteins in the Protein Data Bank (PDB) database, respectively. Clearly, exploration of new protein folds continued after divergence of the three kingdoms of life.

The rapidly expanding structural database shows us that a few folds are extremely common, whereas others are quite rare. Indeed, orphan structures found only in a single protein are known (e.g., the protein encoded by the TM0875 gene of *Thermotoga maritima*).[31] Goldstein has summarized three factors, functionality, designability, and evolutionary dynamics, that have contributed to this distribution of folds.[32] Some folds may be common because they are very well suited for certain functions. For example, the TIM barrel, one of the most common folds found in enzymes, is an excellent structure for catalysis. The mouth of the barrel provides a scaffold from which catalytic and substrate-binding groups can be projected into a central cavity. Thus, the functional versatility of this fold has likely contributed to its frequent use in different catalytic contexts. 'Designability' refers to the likelihood of evolution finding a particular fold, which depends on factors such as how easy it is to find a particular fold, whether a sequence can fold efficiently without being trapped in alternative structures that occupy local minima in the energy landscape, and whether mutations can occur without destroying the fold. Highly designable proteins are predicted to have many contacts between residues that are distant in the primary sequences and help the protein to find the correct fold. Interestingly, protein folds that date back to the LUCA have higher contact densities than average proteins,[33] suggesting that designability favored the early emergence of certain protein folds. Finally, the evolutionary dynamics of protein evolution have been influenced by contingency. The existing collection of proteins does not represent the optimal set of a completely sampled protein universe, but rather the end result of processes in which the first fold (or occasionally folds) that were found to perform a certain function may have been recruited and optimized, and subsequently been available for further evolutionary experimentation, whereas other folds that might have served just as well never had a chance because their initial rudimentary capabilities were dwarfed by those of folds that had already undergone some optimization.

Whether all structural domains have been discovered at this point is debated. Zhang *et al.* claim that the PDB is complete with respect to single-domain protein structures. These authors performed a computational study of structures predicted for a set of homopolypeptides.[34] The side chain in the homopolypeptides was modeled using just the C^{β}-atom. Although this is not realistic, it simplifies the computational problem and removes the complexities of individual side-chain interactions. An *ab initio* approach was used in which folding was governed by hydrogen-bonding interactions involving the backbone, a uniform attractive potential between the side chains, and simple steric exclusion. One hundred and fifty different patterns of secondary structures (alpha helices and beta sheets separated by loops) were assigned, and a large number of structures were

generated and clustered according to structural similarity. For 200-residue homopolypeptides, the top 100 clusters of structures for each of the 150 patterns could be structurally aligned with PDB structures of proteins of fewer than 150 residues with an average root mean square deviation (RMSD) of 4 Å. Conversely, all of the structures in a representative set of proteins with fewer than 150 residues from the PDB were found in the set of homopolypeptide structures. Zhang *et al.* claim that the consistency between the computer-generated structures and the PDB suggests that both sets are likely to be complete representations of the protein universe.

However, other perspectives suggest that there are likely to be as yet unidentified protein folds. As of 2004, the Pfam database (version 10.0) contained 6190 domains; only about a third were associated with a protein of known structure. Some of these may be cases in which highly divergent or completely unrelated sequences adopt an already discovered fold. Some may not be amenable to structural characterization, either because they are too large for characterization by nuclear magnetic resonance spectroscopy (NMR), or contain disordered regions that interfere with crystallization. However, there may indeed be unidentified protein domain structures remaining to be discovered.

8.02.3.2   Mechanisms for Accessing New Folds

Two possible mechanisms for acquisition of new folds have been described: (1) structural drift and (2) passage from an ancestral fold to a new fold through a structurally ambiguous intermediate that can adopt both folds. Structural drift results from the sequential effects of insertions and deletions (indels). A hypothetical pathway for transformation of an all-beta protein to an all-alpha protein as a result of sequential indels is shown in **Figure 7**.[35] Each structure in the figure corresponds to an actual protein. Some of the proteins are evolutionarily related (based upon significant pairwise sequence identity) and others are not, but simply serve as models
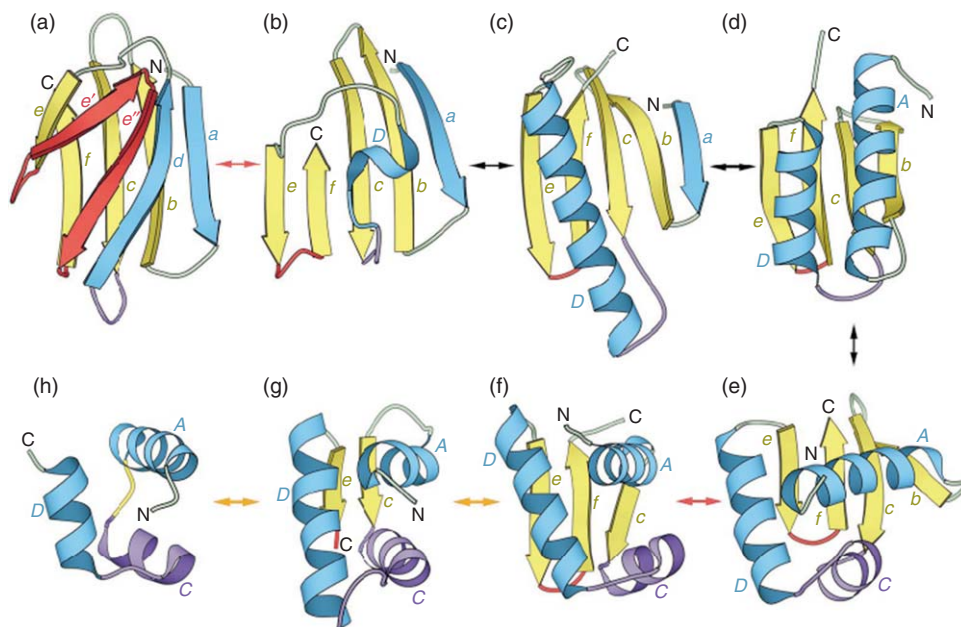

Figure 7   A hypothetical pathway by which an all beta sheet protein (a) might evolve into an all alpha-helical protein (h) by a series of insertions and deletions. Beta sheets are designated by lower case letters and alpha helices by upper case letters. Red arrows indicate a clear evolutionary relationship, orange arrows indicate a possible evolutionary relationship, and black arrows indicate no evolutionary relationship. (a) C-terminal domain of *Bacillus licheniformis* α-amylase (1BPL); (b) C-terminal domain of *Pseudomonas stutzeri* G4-amylase (2AMG); (c) γ-subunit of glycogen phosphorylase kinase N-terminal domain (1PHK); (d) N-terminal signaling domain of sonic hedgehog (1VHH); (e) C-terminal domain of catabolite activator protein (1CGP); (f) N-terminal domain of biotin repressor (1BIA); (g) C-terminal domain of ribosomal protein L11 (1FOW); (h) DNA-binding domain of HIN recombinase (1HCR). Reproduced with permission from N. V. Grishin, *J. Struct. Biol.* **2001**, *134*, 167–185.

for structural changes that could have taken place along this hypothetical pathway. A deletion of two beta strands (shown in red) in the first structure could cause beta strand d to fold into a helix. An insertion in this helix could lengthen it, as shown in the third structure. Substitution of beta strand a with a helix would give the fourth structure. Insertion of a helix in the loop shown in purple and reorientation of helix A could give the fifth structure. The final three structures are formed by successive deletions of beta strands. Although this pathway does not represent an actual evolutionary trajectory between the two end member structures, it makes the point that profound structural changes can result from the cumulative effect of sequential indels, each of which causes only a modest perturbation of the structure.

A second mechanism for evolution of new folds is passage from an ancestral fold to a new fold via a bridge sequence that can fold into two alternative structures.[32] **Figure 8** depicts a situation in which sequences in the red region adopt one structure, whereas sequences in the blue region adopt a different structure. Sequences in the purple region have the ability to adopt both structures, although one may be more heavily populated than the other. Silent mutations that occur as a result of neutral drift may move a protein into the region of structural ambiguity and allow it to exist as a mixture of two structures. If the novel structure provides a useful function, there will be selective pressure to maintain the sequence in the region where both structures can be adopted. Ultimately, gene duplication could allow one protein to diverge back to the original fold and the other to diverge toward the new fold.

The plausibility of this scenario for evolution of novel folds through structurally ambiguous intermediates is supported by known proteins that exist as a mixture of two different structures. The equilibrium can be perturbed toward one conformation or the other by environmental conditions or by the presence of proteins that bind to only one conformation. The most striking example is lymphotactin, a bifunctional human chemokine that binds glycosaminoglycans on cell surfaces and also stimulates XCR1, a G protein-coupled receptor on target leukocytes. Lymphotactin exists as a mixture of two forms under physiological conditions; the equilibrium between the two forms can be shifted by alterations of temperature or salt concentration. By choosing conditions favoring one or the other form, Tuinstra *et al.*[36] were able to solve the structures of the two forms, which are remarkably different (see **Figure 9**). Ltn10, the most stable form at 10 °C and 200 m mol l^{-1} NaCl, is monomeric and adopts a canonical chemokine fold. Ltn40, the most stable form at 40 °C in the absence of salt, is dimeric and adopts a previously unknown fold. Nearly every hydrophobic contact and hydrogen bond differs in the two forms. Further, every residue in the core of Ltn10 is exposed on the surface of Ltn40; the transition between forms has been described as the protein being turned inside out. Variants that were locked in each of the structures were constructed by site-directed mutagenesis. These mutant proteins were used to show that the Ltn10 form activates XCR1, and the Ltn40 form binds glycosaminoglycans. Thus, the structural plasticity of the protein enables it to perform two physiological functions using two different structures.
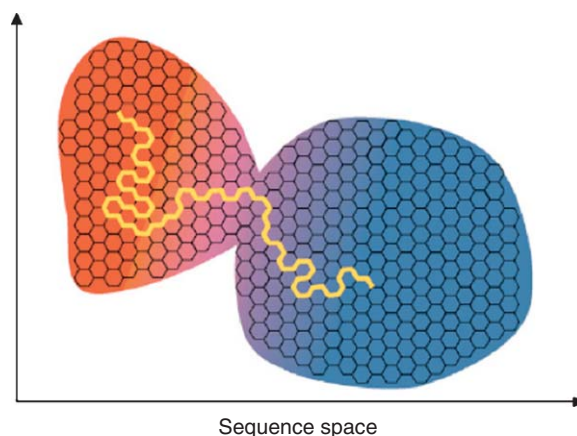
Sequence space

Figure 8 Neutral drift may allow a protein with a certain fold in the red region to access a region in which sequences can adopt the fold of both the red and blue regions, and then to cross into the blue region in which a different fold is found. Reproduced with permission from R. A. Goldstein, *Curr. Opin. Struct. Biol.* **2008**, *18*, 170–177.
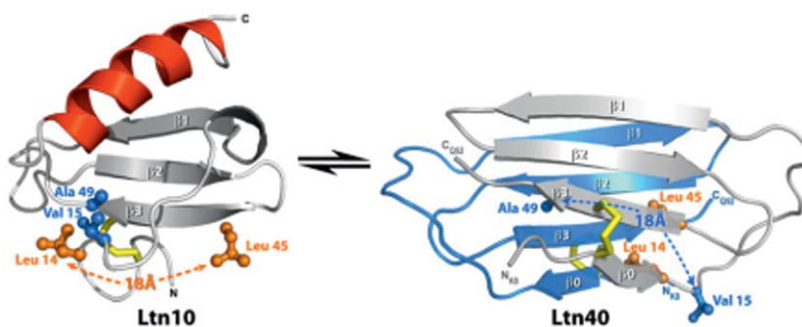
Figure 9   Structures of the two forms of lymphotactin, a human chemokine. Ltn10 activates XCR1 on leukocytes, whereas Ltn40 binds to glycosaminoglycans on cell surfaces. Reproduced with permission from R. L. Tuinstra; F. C. Peterson; S. Kutlesa; E. S. Elgin; M. A. Kron; B. F. Volkman, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5057–5062.

8.02.3.3   Novel Associations of Domains and Subunits Provide New Opportunities for Catalysis, Substrate Binding, and Regulation

In addition to the discovery of new folds by exploration of fold space, diversification of enzyme structures occurred by novel associations between structural units, either by domain fusion or by noncovalent association between subunits. Both strategies can lead to similar results, generating new active site clefts for catalysis, mechanisms for communication between active sites in bifunctional enzymes, and, in some cases, mechanisms for channeling of a product from one active site to another. Indeed, two enzymes may be fused in some organisms but exist as an oligomeric protein in others. An example is the bifunctional enzyme imidazole glycerol phosphate synthase, which liberates ammonia from glutamine at one active site and uses the ammonia to convert phosphoribulosylformimino-aminoimidazole carboxamide ribonucleotide-phosphate to D-*erythro*-imidazole-glycerol-phosphate and aminoimidazole carboxamide ribonucleotide at a second active site. This enzyme consists of two fused domains in fungi and plants,[37] but is a heterodimer in which each activity is found in a different subunit in bacteria.[38] Both types of enzymes exhibit channeling and interactive site communication.

Fusion of protein domains has provided a rich source of structural novelty. Fifty-four percent of Archaeal proteins, 57% of bacterial proteins, and 67% of eukaryotic proteins contain more than one domain. Most multidomain proteins contain two domains, but a few metazoan proteins, most of which are involved in cell adhesion or signaling, have 30–50 repeated domains.[39] Clearly, domain fusion has been a major force in shaping the proteomes of extant life.

An analysis of 40 genomes representing all three kingdoms of life identified 1307 pairwise combinations of 783 domains.[40] Only a small proportion of domains (1%) has been used extensively in combinations with other domains. In multidomain enzymes, the P-loop nucleotide triphosphate hydrolase domain and the Rossman fold (which binds NAD^+/NADH) are the most commonly found. The P-loop hydrolase domain is associated with 47 and 21 different domain partners in bacteria and Archaea, respectively, and the Rossman fold with 29 and 18 domain partners in bacteria and Archaea, respectively. The repeated use of these domains reflects the utility of the reactions they facilitate – phosphoryl transfer and hydride transfer. By combining these domains with a domain that provides substrate specificity, a wide range of reactions can be accomplished without a need to reinvent something that already works well.

Interfaces between domains or subunits are common locations for active sites. In a set of 178 structurally unrelated enzymes containing enzymes representing all classes in the EC classification, catalytic residues are contributed by more than one domain in 35, and by more than one subunit in 19.[41] Substrate-binding residues were not included in this analysis, so the percentage of enzymes with active sites at interfaces between domains or subunits is certainly higher. An example of an active site at the interface of two subunits is that of *trans*-3-chloroacrylic acid dehalogenase (see **Figure 10**). The enzyme is a trimer of $\alpha\beta$ heterodimers. Catalytic residues are provided by both subunits: Glu52 in the α-subunit acts as a general base, whereas Pro1 of the β-subunit acts as a general acid.[42] Remarkably, in a few cases, an active site is formed at the interface of three
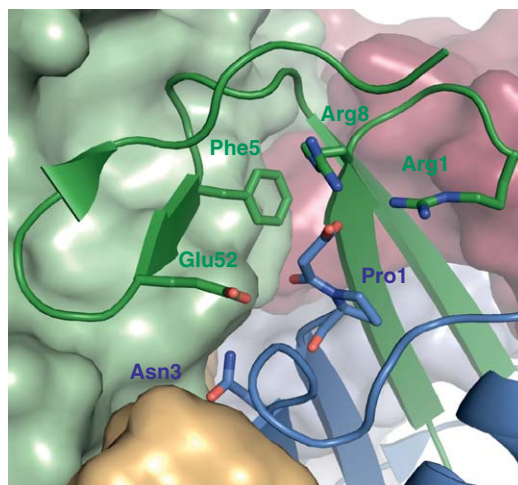
Figure 10   The active site of *trans*-3-chloroacrylic acid dehalogenase. A covalent adduct between 3-bromopropiolate and Pro1 is found in the active site. Glu52, Phe50, Arg8, and Arg11 are contributed by the α-subunit, and Pro1 and Asn39 by the β-subunit.

subunits. Adenylosuccinate lyase is a homotetramer; each active site is formed at the interface of three subunits.[43] Ser295 and His171, the base and acid catalysts respectively, are provided by two different subunits, and substrate-binding residues are provided by all three subunits.

Domain fusion and association of subunits can facilitate complex reactions by assembling active sites that carry out individual reactions required for a multistep transformation. This strategy is particularly useful when one active site generates a highly reactive intermediate such as ammonia, CO, acetaldehyde, or carbamoyl phosphate. Release of such intermediates into solution can allow side reactions or loss by volatilization. The efficiency of the reaction can be improved by channeling the intermediate to the second active site. In a number of enzymes, molecular tunnels have evolved to conduct reactive intermediates between active sites.[44] **Figure 11** shows the tunnel in carbamoyl phosphate synthase, which connects three active sites over a distance of 100 Å. The small subunit (blue) hydrolyzes glutamine to release ammonia, which travels partway through the tunnel to an active site in the N-terminal domain of the large subunit (green). At this site, the ammonia attacks carboxyphosphate formed by phosphorylation of bicarbonate by ATP. The resulting carbamate travels through the tunnel to the third active site in the C-terminal domain of the large subunit (purple), where it is phosphorylated by ATP to provide carbamoyl phosphate.

Association of domains or subunits carrying multiple active sites also allows coordination of various activities carried out by the protein. For example, in imidazole glycerol phosphate synthase, binding of phosphoribulosylformimino-amino imidazole ribonucleotide phosphate at one active site stimulates the hydrolysis of glutamate at another active site 30 Å away by 4900-fold.[45] This strategy prevents generation of NH_3, an inconveniently volatile intermediate, until it is needed.

In many enzymes, domain fusion combines active sites that catalyze successive steps in a pathway. Chorismate mutase converts chorismate to prephenate. Prephenate is converted to phenylpyruvate, a precursor of phenylalanine, by prephenate dehydratase, and to 4-hydroxyphenylpyruvate, a precursor of tyrosine, by prephenate dehydrogenase. *E. coli* contains two bifunctional proteins in which chorismate mutase is fused to prephenate dehydratase and prephenate dehydrogenase (see **Figures 12(a) and 12(b)**). Interestingly, domain fusions occasionally connect active sites that catalyze nonsequential steps in a pathway. *E. coli* HisB (imidazole glycerol phosphate dehydratase/histidinol phosphate phosphatase) catalyzes two nonconsecutive steps in the pathway for synthesis of histidine (see **Figure 12(c)**). It is not obvious why this situation is advantageous. Possibly, undiscovered interactions with the enzyme that catalyzes the intervening step result in a co-localization of three active sites that enhance the efficiency of transfers between them.

Domain fusion can add a domain to an enzyme that facilitates interactions with a membrane, DNA, or another protein. PutA in enteric bacteria is an interesting example. The N-terminal domain of PutA is a
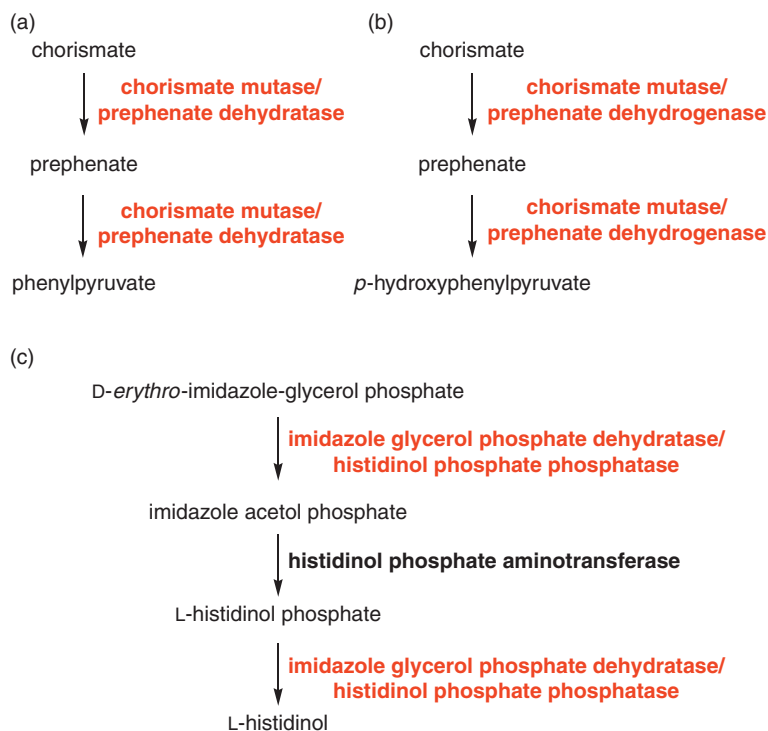
Figure 11   The structure of carbamoyl phosphate synthase, showing in gray mesh the tunnel that transfers ammonia released from glutamine by the small subunit (blue) to an active site in the N-terminal domain of the large subunit (green), where it reacts with carbonate and ATP to form carbamate. The carbamate then travels through the tunnel to a third active site in the C-terminal domain of the large subunit (purple), where it is phosphorylated by ATP to form carbamoyl phosphate. Reproduced with permission from A. Weeks; L. Lund; F. M. Raushel, *Curr. Opin. Chem. Biol.* **2006**, *10*, 465–472.

transcriptional regulator, the middle domain is proline dehydrogenase, and the C-terminal domain is Δ^1-pyrroline-5-carboxylate dehydrogenase. Together, the two enzymes catalyze the 4-electron oxidation of proline to glutamate. The redox state of the flavin in the proline dehydrogenase domain determines its location.[46] When the flavin is reduced by proline, the enzyme associates with the inner membrane, a location that makes sense because proline enters via a PutP (an Na^+-driven proline transporter). In the absence of proline, the flavin is oxidized, and a conformational change releases the protein from the membrane, allowing it to bind to DNA and repress the proline utilization genes *putA* and *putP*.[47]

8.02.4   Elaborating on the Basics

8.02.4.1   Structural Elaboration via Indels

Insertions and deletions (indels) are a common source of structural variation in protein superfamilies. Indels occur most often in loops and turns, as indels in these positions are less likely to disrupt folding than indels in the core of the protein. Insertions often provide novel structural elements that contribute to substrate binding, catalysis, or protein–protein interactions and confer novel characteristics to a diverging family. **Figure 13** illustrates the enormous variability due to indels in the loops surrounding the active site in the amidohydrolase superfamily.[48] The active sites in this superfamily are located at the mouth of an α/β barrel and contain one or two metal ions that activate water for nucleophilic attack on a wide range of substrates, including amino acids,

(a)

chorismate

| **chorismate mutase/ prephenate dehydratase**

prephenate

| **chorismate mutase/ prephenate dehydratase**

phenylpyruvate

(b)

chorismate

| **chorismate mutase/ prephenate dehydrogenase**

prephenate

| **chorismate mutase/ prephenate dehydrogenase**

p-hydroxyphenylpyruvate

(c)

D-*erythro*-imidazole-glycerol phosphate

| **imidazole glycerol phosphate dehydratase/ histidinol phosphate phosphatase**

imidazole acetol phosphate

| **histidinol phosphate aminotransferase**

L-histidinol phosphate

| **imidazole glycerol phosphate dehydratase/ histidinol phosphate phosphatase**

L-histidinol

Figure 12   Examples of bifunctional enzymes that catalyze consecutive steps (a and b) and nonconsecutive steps (c) in metabolic pathways.
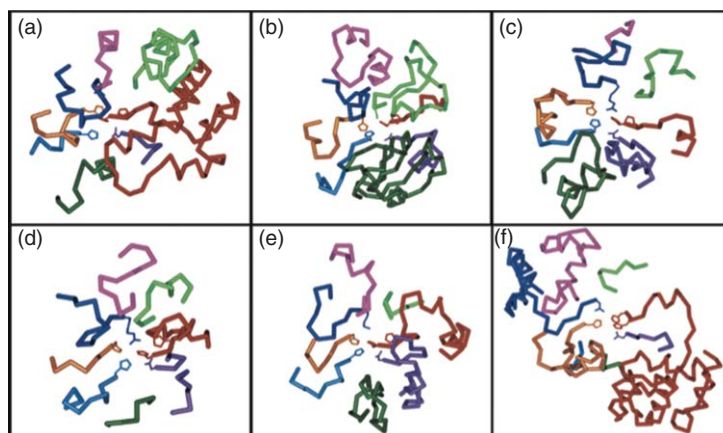
Figure 13   Variability in the conformations of the eight loops at the mouth of the α/β barrel in members of the amidohydrolase superfamily. Loops found in homologous positions are colored as follows; 1, red; 2, purple; 3, dark green; 4, bright blue; 5, orange; 6, dark blue; 7, pink; and 8, bright green. a) adenosine deaminase (PDB1A4M); b) D-aminoacylase (PDB 1M7J); c) dihydroorotase (PDB 1J79); d) isoaspartyl dipeptidase (PDB 1ONW); e) phosphotriesterase (PDB 1HZY); f) uronate isomerase (PDB 1J5S). Reproduced with permission from C. M. Seibert; F. M. Raushel, *Biochemistry* **2005**, *44*, 6383–6391.

nucleic acids, sugars, and organophosphate esters. Variability in these loops dictates substrate specificity as well as the particular mode of metal ion binding.

Indels can also lead to larger-scale structural diversity. The haloacid dehalogenase (HAD) superfamily includes the haloalkanoate dehalogenase, as well as ATPases, phosphonatases, phosphoesterases, and sugar phosphomutases. This superfamily facilitates group transfer reactions by forming a covalent adduct between an

aspartate residue and a carbon or phosphoryl group that will be transferred. Variations on the general architecture of the HAD fold include addition of one or more beta strands at either end of the core five-stranded beta sheet, addition of an α/β unit at one end of the central sheet, and domain-swapping in which a helix from one monomer crosses over to the other monomer in a dimeric enzyme.[49] Two distinctive structural features – a 'squiggle' (a single helical turn) and a flap (a beta hairpin motif) – are present in all HAD superfamily members (see **Figure 14(a)**). The squiggle and flap form a cap that helps bind substrates and sequester intermediates in the active site; the active site typically contains residues contributed by both domains. Two types of caps (designated C0/C1 and C2) occur at different positions and clearly resulted from different ancestral insertions. **Figure 14(b)** shows various structures found in C1 caps, and pathways by which insertions within the cap might have generated the striking structural diversity. Cap residues contribute to substrate recognition[50] and to catalytic activity, albeit in different ways. In phosphonatase, the cap domain provides a lysine residue that forms a Schiff base with the substrate, as well as histidine and methionine residues that bind a water believed to donate a proton during the reaction.[51] In L-2-haloacid dehalogenase, an Arg in the cap forms part of a halide-binding cradle that stabilizes the chloride leaving group.[52] Thus, structural diversity due to insertions has played a major role in the divergence of both substrate binding and catalytic mechanism in this superfamily.

8.02.4.2   The Power of Point Mutations

Point mutations are the most common source of genetic variation. Although most point mutations are neutral or deleterious, some are beneficial, giving rise to improved performance or novel characteristics. Point mutations in substrate-binding regions enabled evolution of ancestral generalist enzymes into highly efficient and specific enzymes, as well as evolution of enzymes to handle novel substrates encountered as environments changed. Point mutations can also alter catalytic residues, allowing catalysis of novel chemical reactions. The enolase superfamily exemplifies the mechanistic diversification that can be achieved within the context of an ancestral scaffold by point mutations. Members of this superfamily catalyze at least 13 different reactions.[53] Additional reactions may be possible, as enolase superfamily homologues that have no known function are found in the sequence databases. All members of the superfamily catalyze abstraction of a proton from a carbon alpha to a carboxylate. The resulting enolate is stabilized by interactions with an active site Mg^{2+}. The fate of the enolate depends on the array of catalytic residues in the active site, which is located between the mouth of a $(\beta\alpha)_8$ barrel and a cap domain. Residues involved in substrate binding and catalysis are contributed primarily by residues in the loops connecting the beta strands and following alpha helices. The four subgroups in the superfamily (see **Figure 15**) have different amino acids at specific positions at the ends of the beta strands. For example, in the enolase subgroup, a Lys at the end of the sixth strand abstracts the proton to form the enolate, and a general acid is found at the end of the second strand. In contrast, in the mandelate racemase subgroup, a His at the end of the seventh strand abstracts the proton, and a general acid is found at the end of the second, third, or fifth strand.

Evolution of a novel activity via point mutations generally requires multiple mutations, bringing up the question of the order in which mutations can occur. Mutations may be additive, synergistic, or antagonistic; consequently, some evolutionary trajectories will be more favorable than others. A rare example in which a fitness landscape has been comprehensively examined is an analysis of the trajectories for accumulation of five mutations in a β-lactamase that increase bacterial resistance to cefotaxime by 100 000-fold.[54] All 32 combinations of the five possible mutations were constructed, and the minimal inhibitory concentration (MIC) for cefotaxime was measured for *E. coli* DH5α carrying each variant. These data allowed estimation of the probabilities for each of the 120 trajectories by which five mutations can accumulate. The striking conclusion was that 102 of the trajectories are not accessible because they involve a mutation that decreases fitness of an evolutionary intermediate, even though the mutation is ultimately required for optimal fitness. Of the 18 accessible trajectories, only 10 account for >90% of the probability density.

Although we often focus on the effects of point mutations on substrate specificity and catalytic properties, point mutations can also affect protein expression and thermostability. A striking example comes from a study of 20 variants of human glutathione *S*-transferase M2-2 that vary only at position 210.[55] Among these 20 variants, expression levels varied by a factor of 10, and thermostability, as measured by half-life at 48 °C, varied by 500-fold.
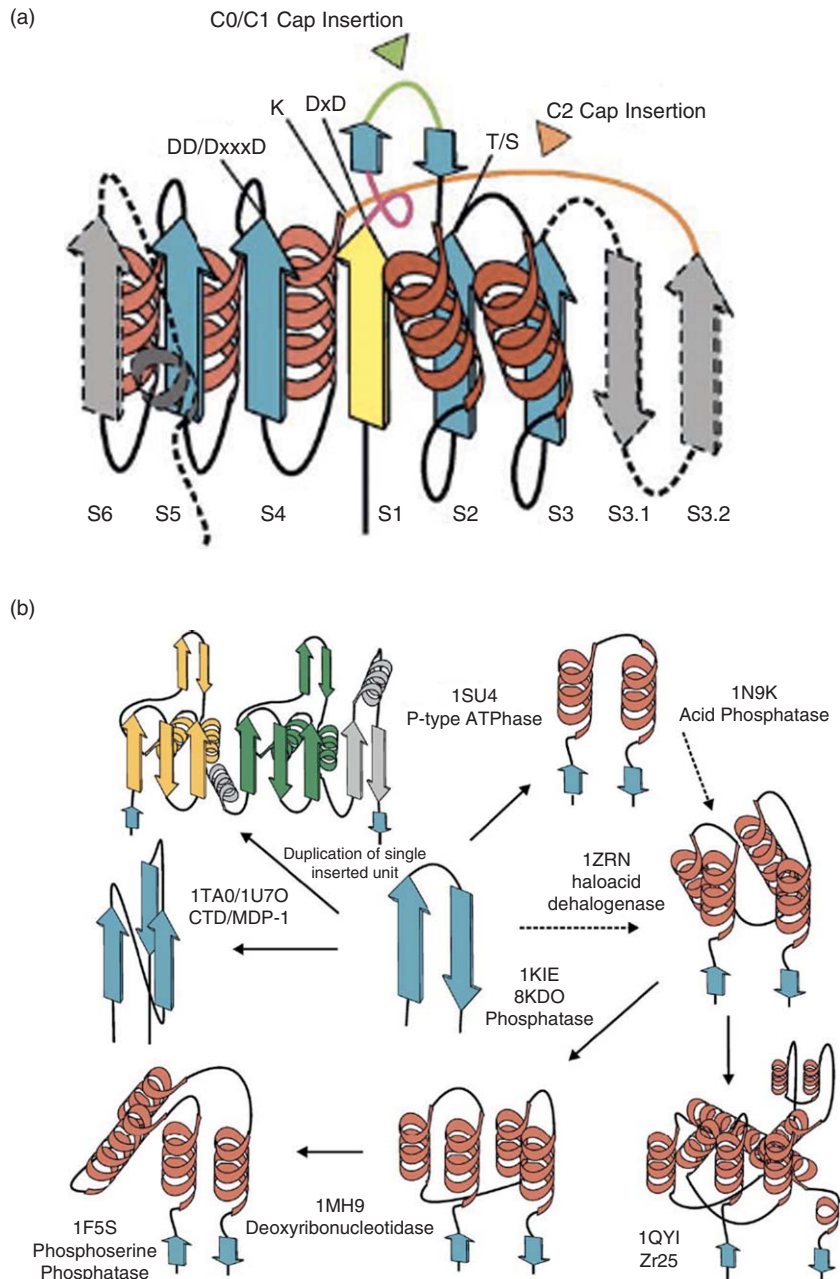
Figure 14   Elaboration of structure in the HAD superfamily by insertions. (a) The topology of the HAD fold, showing the location of the squiggle (pink) and flap (two short beta strands connected by a green loop) and additional beta sheets found in some but not all members of the superfamily (shown in gray); (b) Variation in C1 type caps, and possible pathways for evolution of larger caps from the basic unit in the center. Reproduced with permission from A. M. Burroughs; K. N. Allen; D. Dunaway-Mariano; L. Aravind, *J. Mol. Biol.* **2006**, *361*, 1003–1034.

In every enzyme family and superfamily, there are invariant residues that are required to maintain structure and/or function. However, for the majority of the residues, considerable variability is tolerated; enzymes with <30% sequence identity often have very similar structures and identical functions. Such sequence divergence occurs by neutral drift, a process by which mutations that do not affect the fitness of the organism accumulate over long periods of time.
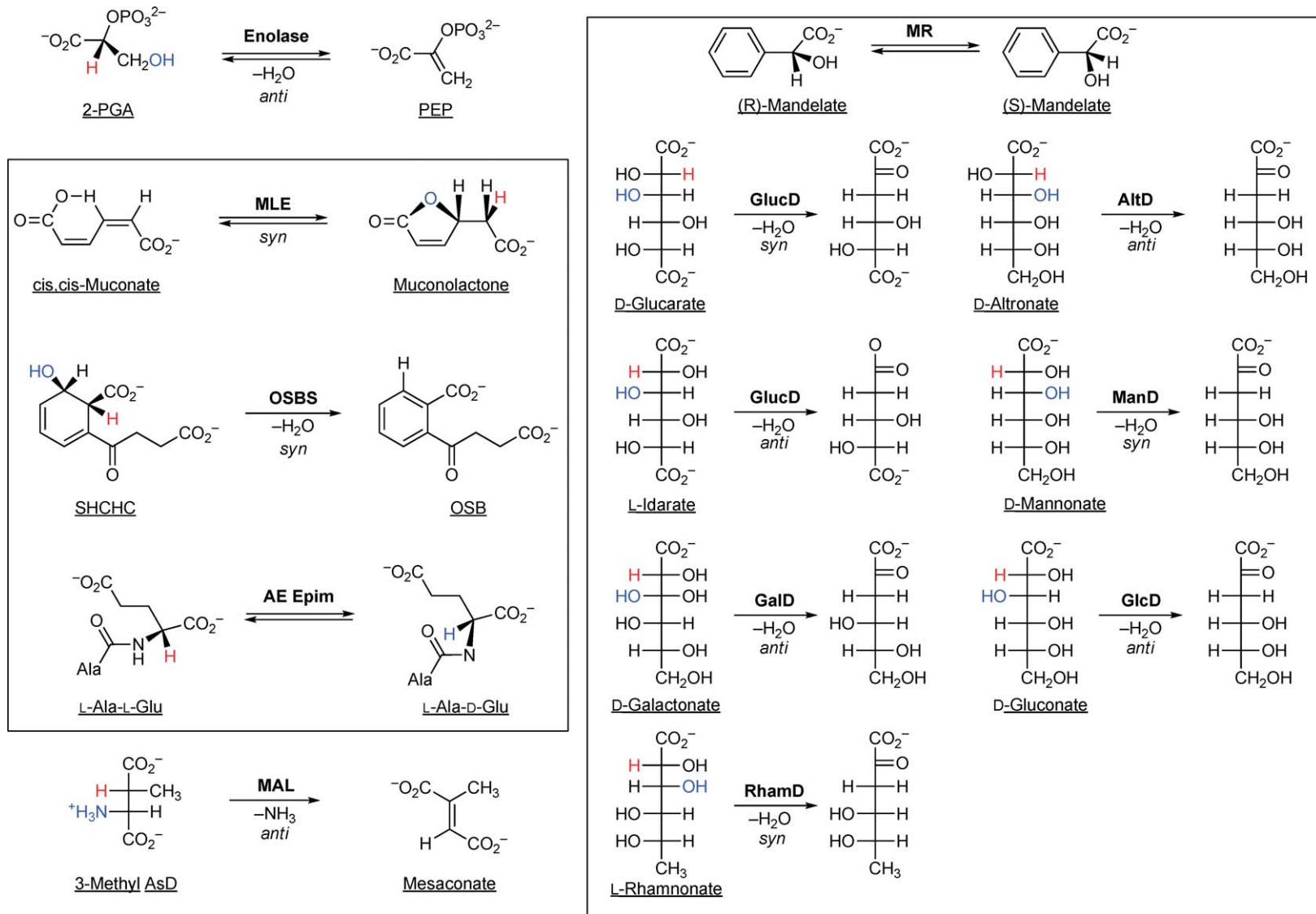
Figure 15 Reactions catalyzed by members of the enolase superfamily. Reproduced with permission from J. A. Gerlt; P. C. Babbitt; I. Rayment, *Arch. Biochem. Biophys.* **2005**, *433*, 59–70.

Although neutral drift has little effect on enzyme function, it can influence the potential for future evolution of novel enzymes in two ways. First, it can increase the stability of the protein, allowing the protein to accommodate later destabilizing mutations that confer a novel property.[56] Second, it can generate promiscuous activities that can be recruited to serve new functions if the environment changes. Enzymes that have diverged by neutral drift in different organisms would be expected to have different levels of promiscuous activities. For example, Gerlt and coworkers identified a relatively high-level promiscuous *N*-acyl amino acid racemase activity in *o*-succinylbenzoate synthase in *Amycolaptosis* sp. The k_{cat}/K_M for *N*-acetylmethionine was $3.7 \times 10^2 \, \mathrm{mol}^{-1} \, \mathrm{l} \, \mathrm{s}^{-1}$. The *E. coli* and *Bacillus subtilis* enzymes have *N*-acyl amino acid racemase activities that are more than four orders of magnitude lower.[57] Variability generated by neutral drift means that some organisms will have the potential to evolve a new enzyme from a promiscuous activity, whereas others will not because the promiscuous activity is too weak to be useful, or because mutations required to increase the level of the promiscuous activity destabilize the protein.

The effects of neutral drift on promiscuous activities were explored in variants of bacteria P450 BM3 by Bloom *et al.*[58] Variants generated by error-prone PCR[59] were screened to insure that they retained the ability to hydroxylate the substrate 12-*p*-nitrophenoxydodecanoic acid at a level at least 75% that of the parental enzyme. **Figure 16** shows the variability in the activity of 34 variants toward 12-*p*-nitrophenoxydodecanoic acid and five alternative substrates. Even within a relatively short mutational distance of the parental enzyme, increases or decreases in activity toward various promiscuous substrates of up to fourfold were found.

8.02.4.3   Multienzyme Complexes

Metabolic enzymes are generally believed to act independently, in contrast to regulatory and structural proteins, which exhibit multiple critical protein–protein interactions. However, a significant number of metabolic enzymes are bi- or even trifunctional, suggesting that association of enzymes involved in the same metabolic pathway may be advantageous. The cytoplasm is crowded, with very high concentrations of proteins, RNA, and metabolites. Maintaining enzymes that catalyze reactions in the same pathway in close proximity might improve access of enzymes to their substrates.

Multienzyme complexes are involved in fatty acid synthesis, nonribosomal peptide synthesis, polyketide synthesis, and decarboxylation of pyruvate, α-ketoglutarate, and branched-chain amino acids. The modular polyketide synthases are particularly impressive.[60] Delivery of the substrate to a succession of active sites in an order dictated by the order of genes in a biosynthetic cluster results in synthesis of elaborate natural products that are used primarily to interfere with growth of competitors in the environment. Evolutionary refinement of metabolic enzymes often enhances fitness by optimizing the ability to control metabolic fluxes and to grow under a range of environmental conditions. The selective pressure is obviously different in the case of the polyketide synthases. Since natural products are used to inhibit growth of competitors, which are likely to evolve mechanisms for resistance to chemical assault, selective pressure has fostered elaboration of structural diversity rather than higher efficiency. Indeed, polyketide synthases are not particularly speedy catalysts, but the combinatorial nature of the process makes evolutionary innovation due to deletion, addition, or rearrangement of catalytic modules particularly accessible.

In multienzyme complexes involved in fatty acid synthesis, nonribosomal peptide synthesis, and polyketide synthesis, substrates are covalently tethered to a phosphopantetheinyl group and shuttled from one active site to another. However, multienzyme complexes can also enhance function when substrates are not covalently tethered. Examples of channeling of intermediates between active sites connected by tunnels were described above. A different approach may be used by glutamate kinase, which produces glutamyl phosphate from glutamate and ATP. Glutamyl phosphate is very unstable; it is prone to cyclization to 5-oxoproline. It has long been suspected that this intermediate is channeled to the active site of the next enzyme, glutamyl phosphate reductase. *E. coli* glutamate kinase is a tetrameric enzyme (see **Figure 17(a)**). The active site is a large open crater. Marco-Marin *et al.* have modeled a postulated interaction between glutamate kinase and the next enzyme in the pathway (see **Figure 17(b)**) using the structure of the *T. maritima* glutamyl phosphate reductase.[61] The active sites of the dimeric glutamyl phosphate reductase can be modeled to neatly overlay the crater-like active sites of the two glutamate kinase active sites present on each face of the tetramer.
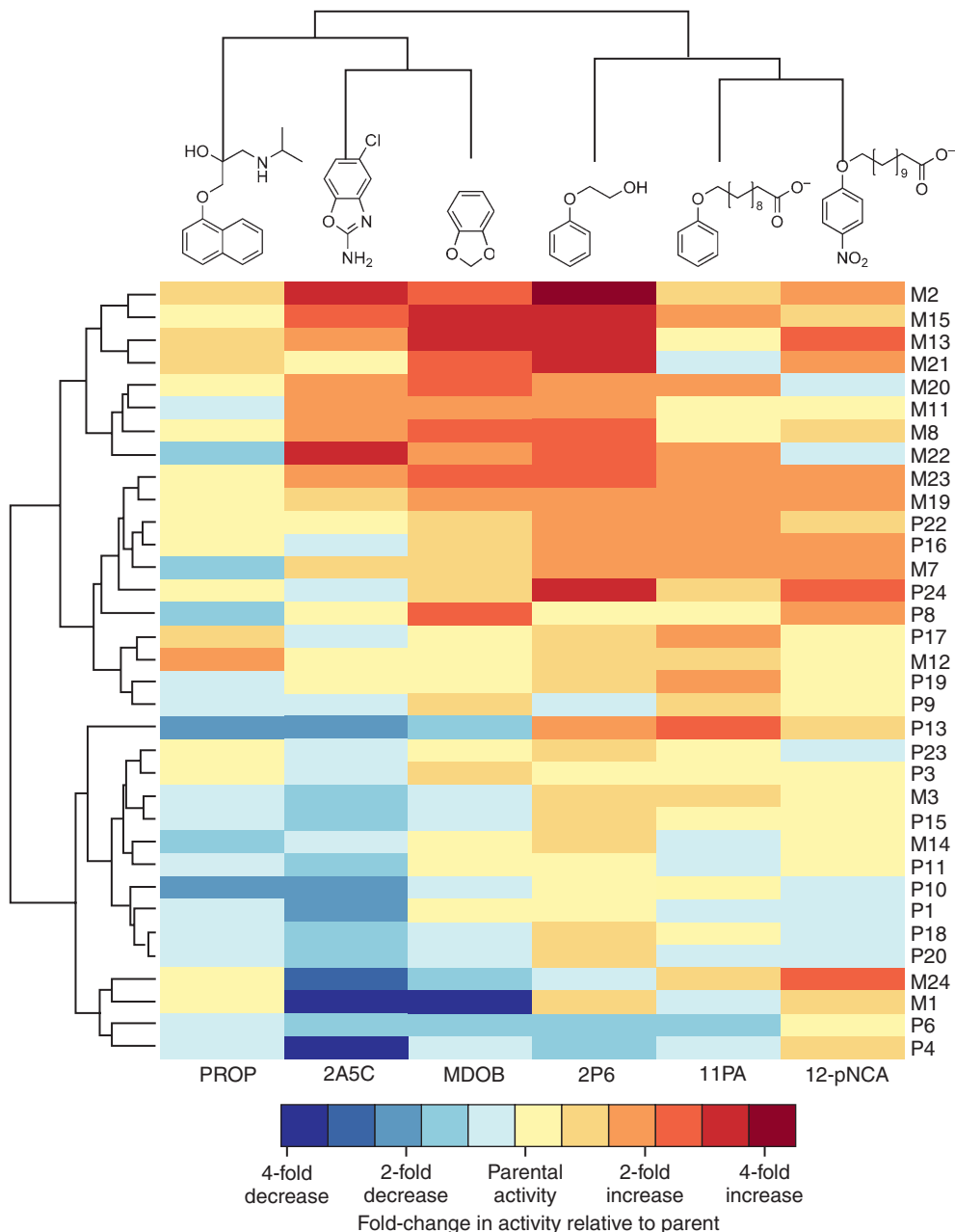
Figure 16 Variability in promiscuous activities caused by mutations in a cytochrome P450 enzyme that had been evolved to hydroxylate12-*p*-nitrophenoxydodecanoic acid (12-pNCA). The 34 variants listed on the right are clustered according to their activity profiles. PROP, propranolol; 2A5C, 2-amino-5-chlorobenzoxazole; MDOB, 1,2-methylenedioxybenzene; 2PE, 2-phenoxyethanol; 11PA, 11-phenoxyundecanoic acid. Reproduced with permission from J. D. Bloom; P. A. Romero; Z. Lu; F. H. Arnold, *Biol. Direct* **2007**, *2*, 17.

Recently, evidence for large-scale associations between enzymes involved in the *de novo* pathway for purine synthesis has emerged.[62] This pathway involves 10 steps that convert PRPP to inosine monophosphate. In higher eukaryotes, these steps are catalyzed by six enzymes; one is trifunctional and two are bifunctional. Versions of these enzymes tagged with green fluorescent protein (GFP) or orange fluorescent protein (OFP)
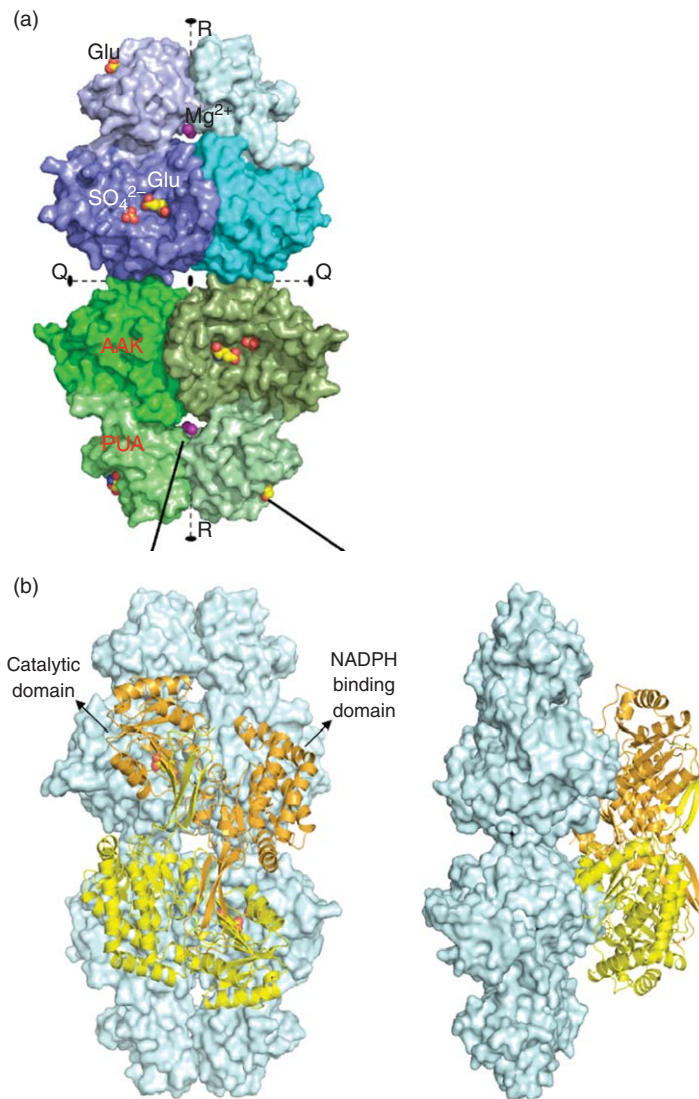
Figure 17 (a) Structure of *E. coli* glutamate kinase (2J5T). Two active sites marked by the positions of glutamate and SO_4^{2-} are visible in this view. Two additional active sites are located in comparable positions on the hidden face of the molecule. AAK and PUA mark the amino acid kinase domain and a domain named after pseudouridine synthases and archaeosine-specific transglycolases; (b) modeling of a complex between glutamate kinase (light blue) and the glutamyl phosphate reductase from *T. maritima* (yellow and orange) (PDB 1O20) that shows that the active sites of the tetrameric glutamate kinase line up nicely with those of the dimeric glutamyl phosphate reductase, possibly providing an opportunity for channeling of glutamyl phosphate from one active site to the other.

co-localize in clusters in living HeLa cells when purines are depleted from the medium, and disperse when purines are supplied. An *et al.* propose that a 'purinosome' forms when purine synthesis is needed, and that the association between the constituent enzymes may be controlled by post-translational modifications that respond to extracellular levels of purines. This demonstration of a purinosome in human cells raises a number of questions. Are multienzyme complexes involved in other metabolic pathways? Do similar complexes form in smaller bacterial cells in which co-localization may be less important? How are the protein–protein interactions mediated and regulated? Do such associations regulate the activities of the enzymes in the complexes? How do such multienzyme complexes affect fitness? Answers to these questions will provide insights into the evolutionary origins of such multienzyme clusters.

8.02.4.4   Regulation

The activities of many enzymes are exquisitely regulated by binding of ligands to allosteric sites. Such regulation is most common for enzymes that catalyze the first committed step in a pathway, and allows flux through a pathway to be adjusted in accord with environmental conditions. The earliest enzymes were probably not regulated at this level, both because they were likely to be generalists and therefore difficult to regulate appropriately, and because the structural intricacies involved in transmitting signals from a remote site to an active site had not yet developed. Selective pressure for enhanced metabolic efficiency fostered emergence of mechanisms for inhibition of activity when downstream metabolites are abundant, thereby allowing carbon and energy to be used for other purposes. Similarly, mechanisms evolved to increase activity when downstream metabolites are in short supply.

The aromatic amino acid synthesis pathway provides an example of how allosteric regulation can control metabolic fluxes even in a complex branched pathway. Phenylalanine, tyrosine, and trytophan are needed in very different quantities for protein synthesis, but are synthesized from a common precursor. The first reaction in the pathway, condensation of erythrose-4-phosphate with phosphoenolpyruvate, is catalyzed by three isozymes of 3-deoxy-D-arabinoheptulosonate-7-phosphate synthetase (AroF, AroG, and AroH) (see **Figure 18**). AroF is inhibited by tyrosine, AroG by phenylalanine, and AroH by tryptophan. Thus, flux through the pathway can be modulated independently by the three aromatic amino acids; a shortage of any leads to increased flux through the pathway. Further down the pathway, allosteric regulation of the first enzymes in the branches of the pathway committed to production of tyrosine, phenylalanine, and tryptophan
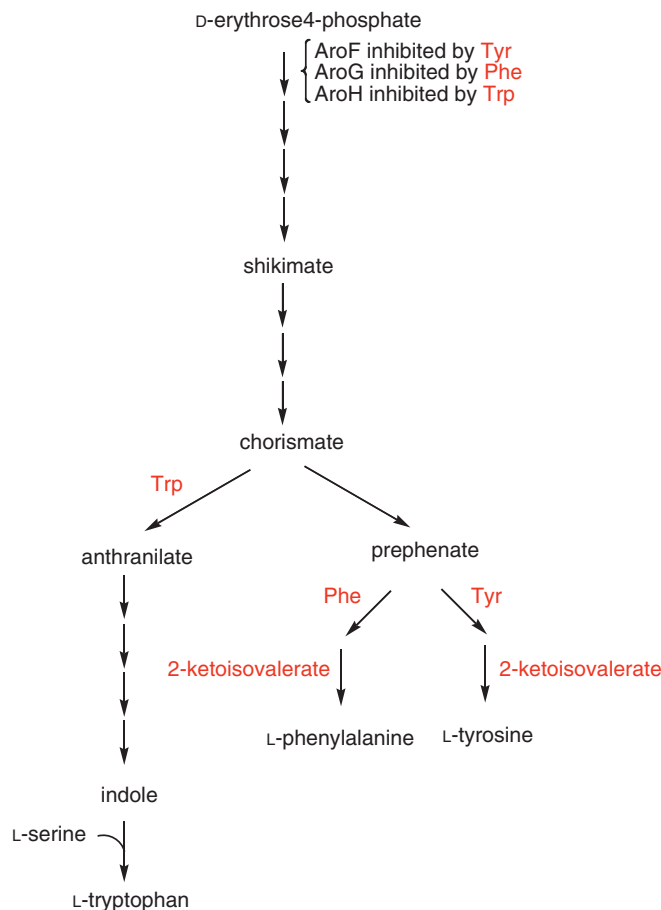
Figure 18   Allosteric regulation in the shikimate pathway for biosynthesis of aromatic amino acids in *E. coli*. Molecules that cause feedback inhibition are shown in red.

allows branches leading to amino acids that are not in short supply to be turned off, directing flux toward the amino acid that is needed.

Binding of an allosteric ligand to an enzyme results in conformational changes that alter the structure of the active site in a way that either enhances or inhibits activity. A fascinating example is the control of ribonucleotide reductase activity by its deoxyribonucleotide products. Ribonucleotide reductases convert ribonucleotide diphosphates or triphosphates to the corresponding deoxyribonucleotides. Class I enzymes, found in eukaryotes, aerobic bacteria, and a few Archaea, use a tyrosyl radical to initiate catalysis. Class II enzymes, found in microbes, use adenosylcobalmin to generate the thiyl radical, and Class III enzymes, which operate only in the absence of O_2 and are found in strict and facultative anaerobes, use a glycyl radical to generate the thiyl radical. The Class III enzyme may have been the ancestral form; Classes I and II enzymes may have evolved later after the appearance of O_2 in the atmosphere.[63] Within each class, a single enzyme reduces substrates containing adenine, uridine, guanine, and cytosine, an interesting feat of broad specificity. An 'activity' site stimulates activity when deoxyribonucleotide levels are low and inhibits activity when deoxyribonucleotide levels are high. The mechanism of control by this mechanism is not understood. The 'specificity' site manipulates enzymatic activity to provide balanced pools of nucleotides for DNA synthesis. When dATP is bound at the specificity site, the substrate-binding site accepts UDP and CDP. When dGTP is bound, the substrate-binding site accepts ADP, and when dTTP is bound, the substrate-binding site accepts GDP. The molecular logic of this arrangement is diagrammed in **Figure 19**. If dATP levels are high, production of dTTP is increased to balance levels of dATP and dTTP, which are needed in equal amounts for DNA synthesis. If levels of dGTP are high, levels of dCTP are raised indirectly by stimulating production of dATP, which stimulates production of dCTP. If levels of dTTP are high, levels of dATP are raised indirectly by stimulating production of dGTP, which stimulates production of dCTP. Thus, high levels of certain purine and pyrimidine nucleotides increase synthesis of both their cognate nucleotides and the other cognate pair, as well.

The structural basis of the allosteric regulation of ribonucleotide reductases is best understood in Class II enzymes. The specificity site and substrate-binding site are separated by a loop (Loop 2) that adopts different conformations depending on the effector and substrate that are bound (see **Figure 20**). Binding of dATP in the specificity site causes a conformational change that flips the side chain of Gln203 into the substrate-binding site in a conformation that favors binding of CDP. Binding of dGTP in the specificity site causes a different conformational change that positions Lys202 to interact with ADP in the substrate-binding site. When dTTP is bound in the effector site, the loop adopts a third structure in which the base of GDP in the substrate-binding site interacts with main-chain atoms of Loop 2 and a turn in the core barrel. It is curious that such a complex mechanism for orchestrating the activity of a single enzyme evolved rather than a set of four enzymes, each responsible for production of a single deoxyribonucleotide.

8.02.4.5 Beyond Structure – The Role of Motion

Evolution of an efficient catalyst requires tuning of the active site to optimally orient substrates with respect to catalytic groups. However, there is more to catalysis than this; protein motions promote catalysis in many enzymes. Protein motions that play a role in catalysis can be large or small, and can occur on timescales ranging from picoseconds to >1 second. Such motions can enhance enzyme activity by many mechanisms, including
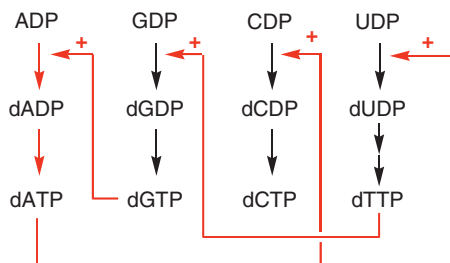
Figure 19 The regulatory circuit controlling the substrate specificity of Class II ribonucleotide reductases.
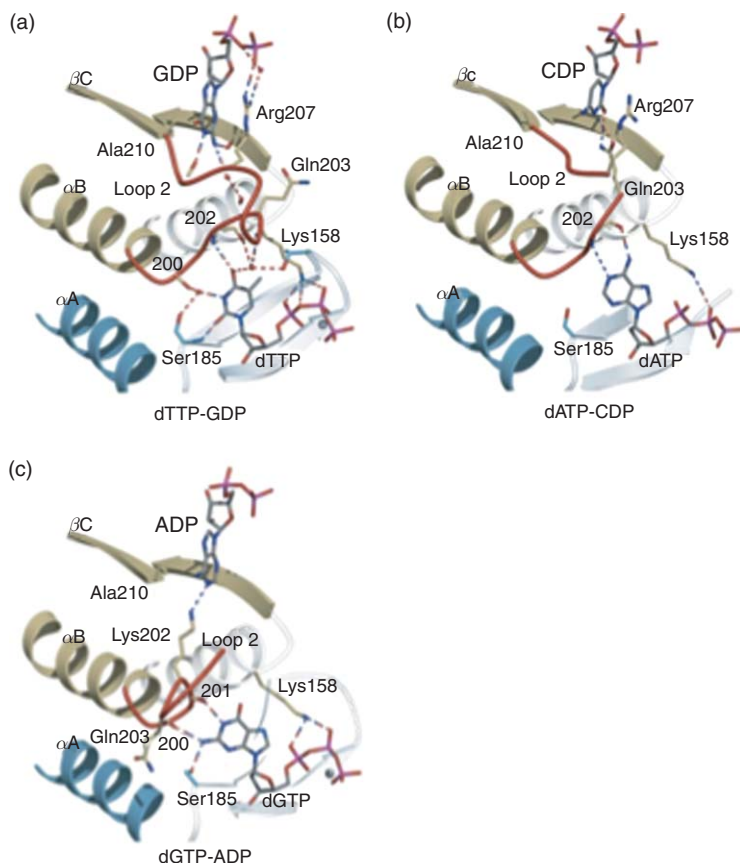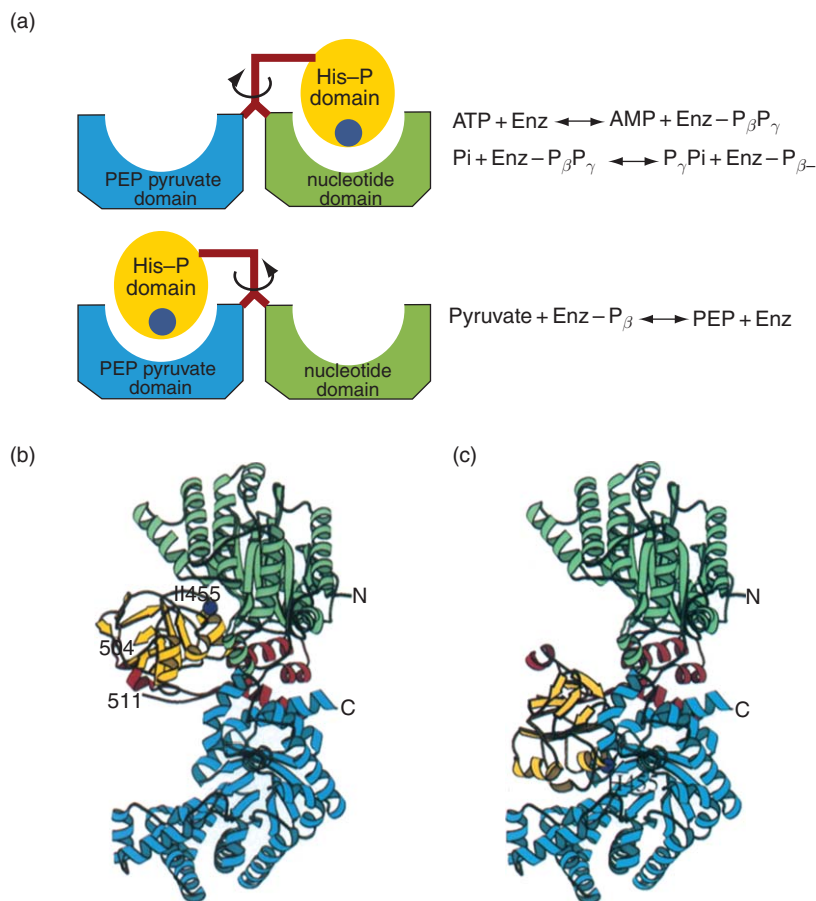
Figure 20 Allosteric interactions that adjust the specificity of the *T. maritima* Class II ribonucleotide reductase to allow generation of balanced pools of deoxyribonucleotides. Loop 2 is shown in red. The active site is at the top, and the specificity site is at the bottom right. (a) the dTTP–GDP complex; (b) the dATP–CDP complex; and (c) the dGTP–ADP complex. The structure of the dATP–UDP complex is similar to the dATP–CDP complex. Reproduced with permission from P. Nordlund; P. Reichard, *Annu. Rev. Biochem.* **2006**, *75*, 681–706.

bringing substrates together in optimal orientations for reactions to occur, bringing catalytic residues into close proximity to substrates, and controlling the production and reactivity of unstable intermediates.

Some enzymes undergo large conformational changes that reorient domains during the catalytic cycle. In pyruvate phosphate dikinase, a massive domain rearrangement shuttles a covalently bound phosphate from one active site to another 45 Å away (see **Figure 21**).[64] This is a remarkably complex way to transfer a phosphate group from ATP to pyruvate. There seems to be no chemical reason to invoke initial transfer to a histidine, followed by transfer to pyruvate, as many kinases transfer phosphate directly to an acceptor. It is possible that this strategy arose from an ancient module that transferred phosphate to multiple acceptors.

Less dramatic conformational changes are involved in cleft closure or closure of lids over active sites in many enzymes, including triosephosphate isomerase,[65] hexokinase,[66] phosphonatase,[67] and adenylosuccinate lyase.[43] Such motions sequester active sites, preventing escape of reactive intermediates or inadvertent reaction with the solvent, and often help to orient substrates and/or catalytic residues.

More subtle motions of loops or amino acid side chains are commonly involved in assembling catalytic sites around a substrate molecule. The importance of 'configurational adaptability' for ligand binding was recognized in 1950,[68] and this concept was elaborated in Koshland's theory of 'induced fit'.[69] Comparisons of enzyme structures in the absence and presence of ligands often reveal changes in structures in the liganded forms. Even very small motions can be important for catalysis. Studies of the effect of subtle perturbations in the positions of substrates bound to the active site of isocitrate dehydrogenase[70] illustrate this point. Isocitrate dehydrogenase catalyzes transfer of a hydride from isocitrate to nicotinamide adenine dinucleotide phosphate (NADP). If NADP is

(a)

His–P
domain

PEP pyruvate
domain

nucleotide
domain

$$ATP + Enz \longleftrightarrow AMP + Enz - P_\beta P_\gamma$$
$$Pi + Enz - P_\beta P_\gamma \longleftrightarrow P_\gamma Pi + Enz - P_{\beta-}$$

His–P
domain

PEP pyruvate
domain

nucleotide
domain

$$Pyruvate + Enz - P_\beta \longleftrightarrow PEP + Enz$$

(b)

Il455

504

511

N

C

(c)

N

C

Figure 21 Movement of domains in pyruvate phosphate dikinase during the catalytic cycle transfers the phosphate covalently bound to histidine from the nucleotide-binding site to the PEP/pyruvate-binding site. Green, nucleotide-binding domain; gold, phosphohistidine domain; blue, PEP/pyruvate-binding domain; (a) a schematic of the domain movement; (b) structure in which the phosphohistidine domain is docked at the nucleotide-binding domain; (c) a model of the structure in which the phosphohistidine domain is docked at the PEP/pyruvate-binding domain obtained by swiveling the phosphohistidine domain around residue 380. Reproduced with permission from O. Herzberg; C. C. Chen; G. Kapadia; M. McGuire; L. J. Carroll; S. J. Noh; D. Dunaway-Mariano, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2652–2657.

replaced by nicotinamide hypoxanthine dinucleotide phosphate (NHDP), k_{cat} decreases by more than 10^4, even though the relatively modest difference between the two cofactors is remote from the nicotinamide moiety. The hydrogen to be transferred from isocitrate is 2.7 Å away from C4 of the nicotinamide in Y160F isocitrate dehydrogenase complexed with Mg-NADP and isocitrate. (The Y160F form of the enzyme is active, but slow, allowing its structure to be determined by time-resolved Laue X-ray crystallography.) NHDP binds in a slightly different orientation, resulting in an increase in the distance between the hydrogen to be transferred from isocitrate and C4 of the nicotinamide by 1.55 Å. Thus, small structural changes that alter the distance and/or angle between reacting atoms can be critical for catalytic efficiency. Small conformational changes that alter the positions of catalytic residues when substrates are bound are common, and have certainly evolved due to the enhancement of reaction rates that can be achieved by optimizing the positions of catalytic residues around substrates.

A particularly elaborate set of protein motions controls chemical events at the active site of flavin mono-oxygenases that hydroxylate phenols.[71] The catalytic cycle begins with binding of the substrate (see **Figure 22**). After the substrate is bound, the enzyme undergoes a conformational change that moves the flavin cofactor into a position in which it can be reduced by NADPH. The reduced flavin reacts with O_2 to form C4a-hydroperoxyflavin, which transfers a hydroxyl group to the substrate. The catalytic cycle is completed by elimination of H_2O from the C4a-hydroxyflavin, which returns the flavin to the oxidized state, and release of
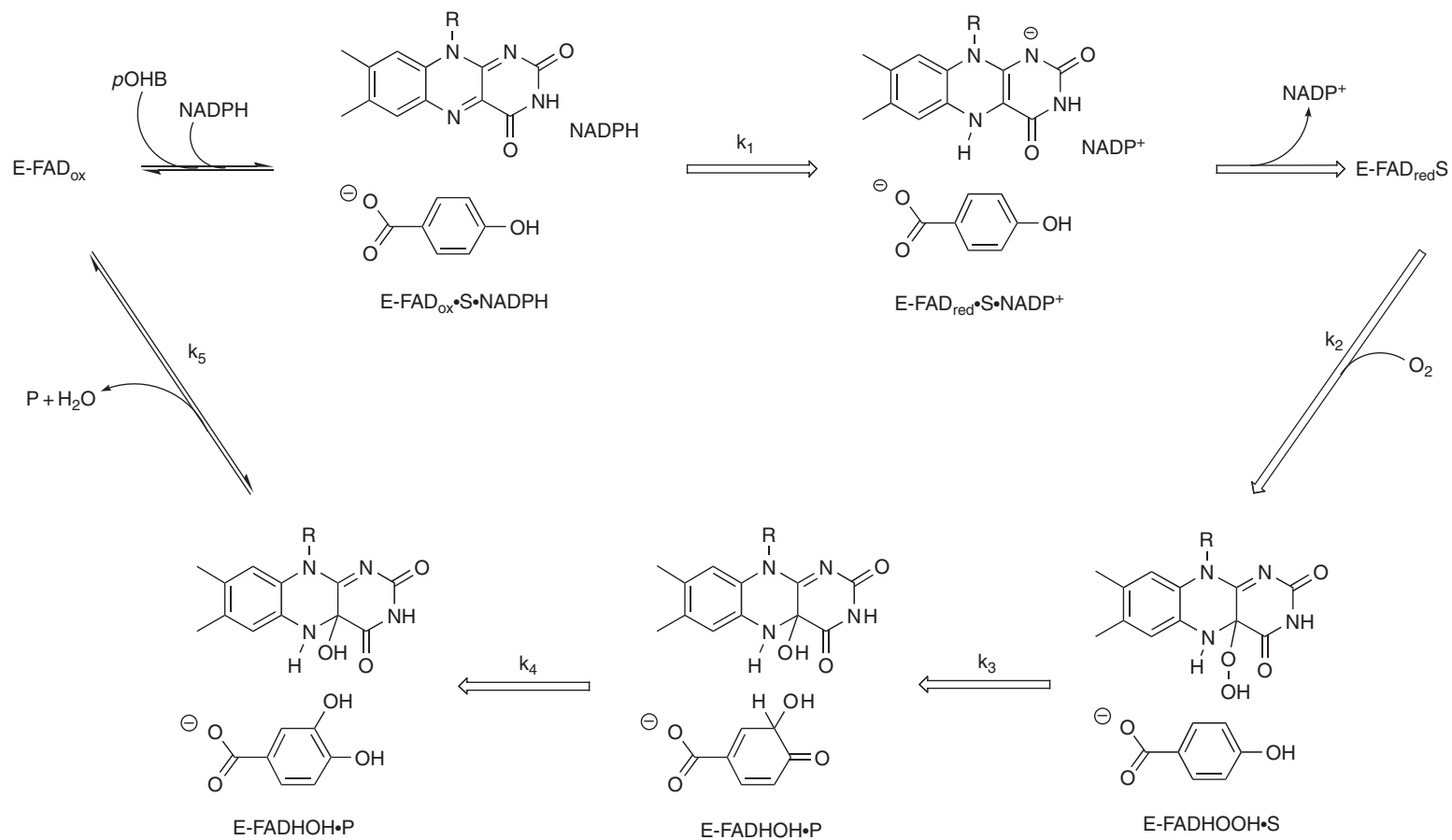
**Figure 22** The mechanism of *p*-hydroxybenzoate hydroxylase.

the product. The flavin motions are a critical part of the catalytic cycle. If the flavin is reduced in the absence of the substrate, it reacts with O_2, and the C4a-hydroperoxyflavin breaks down to H_2O_2 and oxidized flavin. The solution to this dilemma is to sequester the flavin until the substrate is bound, insuring that an acceptor for the hydroxyl group is present before the flavin is reduced and allowed to react with O_2. The sophisticated mechanism likely evolved to avoid waste of NADPH and generation of toxic H_2O_2 by the uncoupled reaction.

Protein motions can contribute to catalysis in even more subtle ways. Correlated motions within a protein can push two substrates together, or push a substrate and a catalytic group together, placing reacting atoms in a conformation that is poised to move toward the transition state due to the close distance between the reacting atoms, as well as optimal overlap of the HOMO and LUMO. Correlated motions can also open channels that allow substrates to bind to a buried active site.

Evidence for the importance of correlated motions that promote catalysis has emerged from a combination of structural, kinetic, and computational approaches. Enzymes that catalyze hydride transfer or electron transfer reactions have been studied most intensely. These reactions often occur via tunneling when rearrangement of the environment results in equalization of the energies of the particle to be transferred in the reactant and product wells. Protein motions that accomplish such energetic adjustments promote the reaction by allowing the particle to tunnel through the barrier.

An early experimental demonstration of the importance of protein vibrational motions came from studies of alcohol dehydrogenase from *Bacillus stearothermophilus*, a thermophile that grows at 65 °C.[72] Many thermophilic enzymes lose activity at moderate temperatures, an observation that suggested that protein motions that contribute to catalysis may be 'frozen out' at temperatures below the optimal growth temperature of the organism. Alcohol dehydrogenase catalyzes a hydride transfer reaction that occurs by a tunneling mechanism. Tunneling can be detected by measuring kinetic isotope effects. Under semiclassical conditions, the Schwain–Schaad relationship $(k_D/k_T)^{3.3} = k_H/k_T$ holds. However, if the reaction occurs by tunneling, the exponent in this relationship is larger than 3.3. For the *B. stearothermophilus* alcohol dehydrogenase, primary and secondary kinetic isotope effects are nearly constant between 60 and 30 °C, and the exponent is >10, providing strong evidence for tunneling. Below 30 °C, both primary and secondary isotope effects increase, but the exponent decreases toward the semiclassical limit. These results suggest that protein fluctuations accelerate the rate at temperatures experienced by the microbe by facilitating tunneling. Notably, below 30 °C, $\Delta H^{\neq}$ increases from 14.6 to 23.6 kcal/mol. Thus, protein motions can have dramatic effects on the rate of the reaction.

The role of protein motions in dihydrofolate reductase (DHFR) has been examined using molecular dynamics simulations as well as experimental studies of mutant enzymes. DHFR catalyzes transfer of hydride from NADPH to dihydrofolate via a tunneling mechanism.[73] Molecular dynamics simulations of complexes of the enzyme with both substrates show both correlated and anticorrelated motions (see **Figure 23**).[74] Mutations
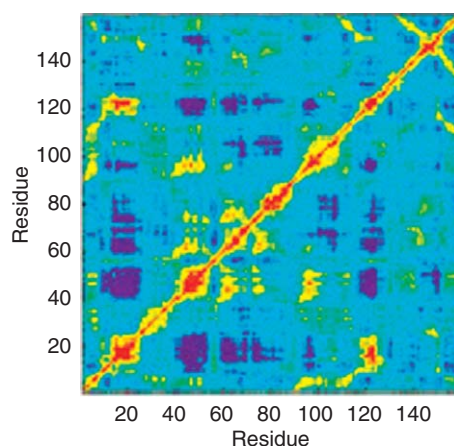
Figure 23   Residue–residue map of correlated motions in the complex of DHFR with dihydrofolate and NADPH. Red and yellow indicate regions of correlated motions, and dark blue regions of anticorrelated motions. Reproduced with permission from J. L. Radkiewicz; C. L. Brooks, III, *J. Am. Chem. Soc.* **2000**, *122*, 225–231.
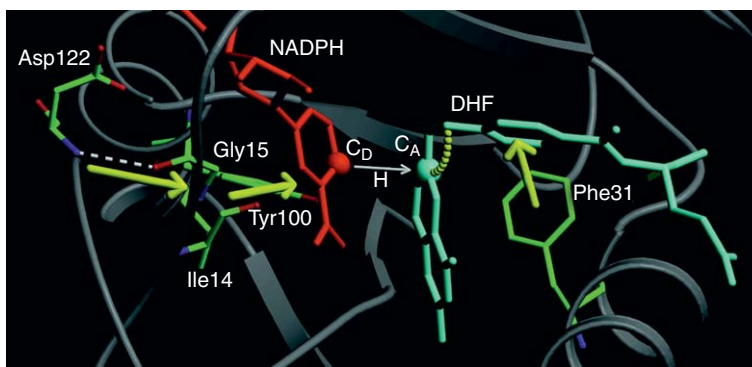
Figure 24   Coupled motions in DHFR are proposed to push NADPH toward dihydrofolate as the reaction proceeds. Reproduced with permission from P. K. Agarwal; S. R. Billeter; P. T. Rajagopalan; S. J. Benkovic; S. Hammes-Schiffer, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2794–2799.

in 11 residues had previously been shown to affect the rate of at least one of the first three steps in the catalytic cycle. Four of these residues line the active site, but seven are more remote. Notably, these seven residues occur in regions that are predicted to participate in anticorrelated motions.

Agarwal *et al.* calculated the average structure of DHFR at various stages along the reaction coordinate from the reactant to the product.[75] The positions of several side chains in the vicinity of the active site changed as the transition state was approached and the distance between the NADPH and the acceptor carbon of dihydrofolate decreased (see **Figure 24**). Gly15 and Ile14 shifted away from the backbone amide of Asp122, which forms a hydrogen bond with the backbone carbonyl of Gly15. The backbone carbonyl of Ile14 is hydrogen bonded to the carboxamide group of NADPH. This distance does not change, suggesting that Ile14 and NADPH move in concert toward the acceptor carbon of dihydrofolate. Thus, motions of the protein are correlated with motion of NADPH toward dihydrofolate, and are believed to promote hydride transfer by decreasing the distance between the donor and the acceptor. Notably, Asp122, Gly15, and Ile14 are absolutely conserved in DHFRs, suggesting that these residues are involved in an ancient function that is important for catalysis.

The evolution of rate-promoting correlated motions has been addressed using a computational approach that captures the essence of an evolutionary process in a model system.[76] Bagdassarian and coworkers simulated an enzyme as a system of 280 thermally fluctuating subunits arranged in five layers. The active site was modeled in the middle layer by one subunit designated C (the catalytic residue) and one designated S (the substrate) (see **Figure 25**). The surrounding subunits were labeled D or N, and the entire structure was

P	P	P	P	P	P	P	P	P	P
P	N	N	N	N	N	N	N	N	P
P	N	D_{25}	D_{26}	D_{27}	D_{28}	D_{29}	D_{30}	N	P
P	N	D_{19}	D_{20}	D_{21}	D_{22}	D_{23}	D_{24}	N	P
P	N	D_{13}	D_{14}	C_{15}	S_{16}	D_{17}	D_{18}	N	P
P	N	D_7	D_8	D_9	D_{10}	D_{11}	D_{12}	N	P
P	N	D_1	D_2	D_3	D_4	D_5	D_6	N	P
P	N	N	N	N	N	N	N	N	P
P	P	P	P	P	P	P	P	P	P

Figure 25   The middle layer of the model enzyme used by Bagdassarian and coworkers to examine the role of vibrations in promotion of catalysis. C is the catalytic subunit, and S the substrate. Molecular dynamics simulations were used to assess the catalytic efficiency of 'enzymes' that varied in the number of flexible and stiff linkages between neighboring subunits in the white box. P (phantom) and N (neutral) subunits were not varied during the simulation. Reproduced with permission from G. S. B. Williams; A. M. Hossain; S. Shang; D. E. Kranbuehl; C. K. Bagdassarian, *J. Theor. Comput. Chem.* **2003**, *2*, 323–334.

surrounded by a layer of spatially fixed phantom (P) residues. The N subunits were neutral in the sense that they were not modified by the genetic algorithm. Interactions between nearest neighbor subunits were modeled by harmonic interactions, and the spring constant was varied to represent stiff or loose springs. The C and S subunits were prevented from interacting until the distance between them, which was governed by fluctuations of the surrounding subunits, decreased to a defined length. At that point, the energy of a repulsive interaction (corresponding to an activation energy barrier) increased linearly until a specified distance was reached; at that point, a chemical 'hit' was scored. Molecular dynamics simulations were run and the number of chemical 'hits' achieved by systems with different values for the spring constants was scored, allowing a measure of the fitness of enzymes with varying degrees of stiffness. (Fitness is defined as the number of chemical hits during the simulation run.) A genetic algorithm was used to 'evolve' the initial system toward higher or lower fitness by altering spring constants between nearest neighbors in the set of C, S, and D subunits (with the exception of the interaction between the C and S subunits defined above). Among 30 000 variants, the fittest individual had a fitness of 253. An individual with all loose interactions had a fitness of 78, and one with all stiff interactions had a fitness of 96. Thus, a combination of loose and stiff interactions is optimal. Notably, the least fit individual had a fitness of only 16; some combinations of loose and stiff interactions actually decrease fitness. Furthermore, the distribution of stiff interactions was important; one individual with the same numbers of stiff and loose interactions as the fittest individual had a fitness of only 136. Considerable degeneracy was observed, suggesting that many combinations of stiff and loose interactions can provide the same catalytic fitness.

This model system was used to probe the mechanisms by which more fit individuals emerged during the evolutionary process.[77] The data suggest that catalytic fitness was improved in two stages. Initially, fitness was increased by shortening the C–S distance. Subsequently, fitness was improved by increasing the fraction of rate-promoting oscillations in the enzyme (i.e., anticorrelated motions in which C and S move toward and then away from each other, and motions in which one subunit is stationary and the other oscillates toward and away from it). Although this abstract model lacks many features of real enzymes, it reveals general principles about the evolution of rate-promoting motions that are intuitively reasonable and escape the anecdotal nature of experiments on individual proteins.

8.02.5 Enzyme Evolution in the Context of the Cell

The preceding sections have focused on the molecular aspects of evolution of enzyme efficiency, specificity, and regulation. However, it is important to remember that enzymes evolve in a cellular context. Natural selection acts upon the fitness of the organism, which is a complicated function of properties such as growth rate, robustness to environmental perturbations, ability to construct a protective biofilm, resistance to toxins, or synthesis of secondary metabolites that impair competitors. The catalytic capabilities of enzymes, as well as the repertoire of enzymes maintained by individual species of microbes, have been shaped by these environmental factors over billions of years.

8.02.5.1 The Influence of the Environment

Microbes inhabit environmental niches that span an enormous range of physical and chemical conditions. Microbes grow at pH values as low as 0, and as high as 11.[78] Extreme thermophiles grow at temperatures exceeding 80 °C; *Pyrolobus fumarii* grows at an astonishing 113 °C.[79] Halophiles grow in water that is nearly saturated with salt. Special adaptations are required for microbes to survive in such extreme environmental conditions, and these are reflected in their metabolic enzymes.

Most acidophiles and alkaliphiles maintain a cytoplasmic pH within two units of the environmental pH, but that still can leave substantial deviations from neutrality. Since the pK_a's of acids, bases, and nucleophiles in active sites are critical for catalysis, it is likely that residues in and near the active site must be changed to tune the pK_a's of catalytic residues to appropriate values in microbes whose cytoplasmic pH deviates from neutrality.

The high-temperature environments inhabited by thermophiles require adaptations to promote stability of proteins at temperatures that denature proteins from most organisms. A further challenge is that entropic

contributions to ΔG and $\Delta G^{\neq}$ are more substantial at high temperatures, altering the free-energy landscape within which the cell must operate.

Halophiles maintain a high internal osmotic pressure using KCl or osmolytes such as proline, amino acid derivatives, sugars, polyols, and methylated sulfur compounds.[80] High ionic strength in the cytoplasm, in particular, can profoundly effect interactions between enzymes and their substrates, as well as protein–protein interactions in general. Proteins from halophiles that have high cytoplasmic KCl concentrations contain an unusually high content of acidic residues.[81–83] Interactions between K^+ and the negative charges of Asp and Glu residues on the surface of proteins are believed to promote solubility in the cytoplasm. Notably, many halophilic enzymes do not function, and some do not even fold, in the absence of high salt concentrations, testifying to the effects of evolutionary processes that adapt enzymes to function in high salt conditions.

In addition to the challenges presented by extreme environments, many microbes inhabit environments that are changeable with respect to the availability of resources and/or physical characteristics such as temperature and pH. Features of enzymes and entire pathways that are optimal under one set of environmental conditions may not be optimal under a different set of conditions. Robustness to perturbations and the ability to adjust metabolic fluxes in response to environmental variability are features that are selectable and have resulted in evolution of control mechanisms that tune both gene expression and enzyme activity.

The relationship between environmental variability and fitness conferred by variant enzymes has not been examined, but studies of a regulatory protein illustrate principles that should be applicable to enzymes, as well. ArgR is a transcriptional regulator that controls the *arg* regulon.[84,85] The *E. coli* K12 version of ArgR, which has Pro at position 70, is a potent repressor of the *arg* regulon that allows tight control of arginine synthesis. The *E. coli* B version of ArgR, which has Leu at position 70, is a weak co-inducer that slightly increases expression of the *arg* regulon in the presence of arginine above a low constitutive level. In otherwise isogenic strains, the K12 version enhances fitness when arginine concentrations are high because the *arg* regulon can be repressed, and cells do not expend energy synthesizing unnecessary enzymes. The B version enhances fitness in the absence of arginine for reasons that are not clear. The relative fitness of the two strains in an environment in which arginine levels fluctuate turns out to depend on the frequency of the fluctuations. When the fluctuations are of small amplitude and high frequency, cells carrying the K12 version are favored because arginine is available most of the time, and it is advantageous to suppress the *arg* regulon. When the fluctuations are of large amplitude and low frequency, the cells spend a substantial amount of time under low arginine conditions, and cells carrying the B version are more fit. The net effect is that the composition of the population is governed by the frequency of environmental changes. Fluctuations can generate mixed populations that change in phase with the fluctuations. In real environments, fluctuations may not be cyclical, so the dynamics of the population may be even more complicated.

Changing environmental conditions can shape the course of enzyme evolution by temporarily removing the need for a particular enzyme. Microbes tend to lose genes that are not needed. If a habitat provides no selective pressure for retention of a metabolic pathway, one or more genes encoding enzymes in the pathway may be lost. If the environment subsequently changes and the pathway becomes important again, then a new version of the missing enzyme must be acquired, either through horizontal gene transfer or through recruitment of an enzyme with a sufficiently high promiscuous activity to supply the needed function. In some cases, this will result in evolution of a new enzyme that is not homologous to the enzyme that was lost. Many examples of convergently evolved enzymes that catalyze comparable reactions using different scaffolds and sometimes different mechanisms are known. For example, there are two structurally and mechanistically distinct classes of dehydroquinate dehydratases.[86] Type I enzymes have a parallel α/β barrel structure and use a Schiff base with an active site Lys to stabilize the carbanionic intermediate formed during dehydration of dehydroquinate. In contrast, Type II dehydroquinate dehydratases are comprised of a β sheet flanked by α-helices, and do not use a Schiff base to stabilize the carbanionic intermediate. Other examples of convergently evolved enzymes include pantothenate kinases,[87,88] ribose-5-phosphate isomerases,[89,90] and thymidylate synthetases.[91]

In the context of enzymes, various mechanisms allow adaptation to fluctuating environmental conditions. Most importantly, mechanisms for transcriptional control and allosteric regulation have evolved to modulate enzyme expression and activities and the consequent fluxes through metabolic pathways in response to environmental perturbations. Another mechanism for adaptation to variable environmental conditions is evolution of isozymes that catalyze a common reaction, but function optimally under different conditions.

For example, *E. coli* has three genes for fumarate hydratase (fumarase). The constitutively expressed FumA is the major fumarase under anaerobic and microaerophilic conditions. FumB provides minor activity under anaerobic conditions. Both enzymes lose activity at high O_2 levels due to oxidation of their iron–sulfur clusters.[92,93] FumC, which lacks an iron–sulfur cluster, is the major fumarase under aerobic conditions.[94] The availability of three isozymes means that fumarase activity can be maintained regardless of the concentration of O_2.

An additional strategy that enhances metabolic adaptability is evolution of alternative enzymes that synthesize the same product using different substrates (see **Table 1**). Alternative enzymes allow bacteria to take advantage of reactants that may be in sufficient supply only occasionally in a fluctuating environment. For example, *E. coli* contains two enzymes that catalyze oxidative decarboxylation of coproporphyrinogen III to form protoporphyrinogen IX in the heme biosynthesis pathway. HemF uses O_2; HemN is a radical *S*-adenosylmethionine enzyme that catalyzes a complex O_2-independent reaction. HemN contains an O_2-sensitive Fe–S cluster.[95] HemN is active only under anoxic conditions, and HemF is active only when O_2 is available, but *E. coli* is able to synthesize heme in either case.

The environment inhabited by a microbe is profoundly influenced by other microbes. Microbial communities are often densely populated and quite complex. The human gut microbiome consists of at least 15 000 species of bacteria, with significant variability between individual humans.[96] A sample of farm soil contained more than 3000 bacterial species,[97] and an endolithic community in ocean-floor basalt contained on the order of 440 bacterial species.[98] Competition for resources is a fundamental driver of evolutionary processes, leading to highly efficient enzymes and exquisitely regulated metabolic pathways. However, cooperative interactions between microbes also drive evolutionary processes. Consequently, the fitness of a microbial species will be strongly dependent on the nature of the microbial community. For example, selective pressure for evolution of a pathway for degradation of a particular compound may be present in an environment in which some microbes excrete that compound as a waste product, but not in others. The interdependence of microbial metabolisms in anaerobic degradation of complex organic matter to methane is illustrated in **Figure 26**. Fermentative bacteria break down complex organic matter into organic acids and alcohols. Other bacteria, termed syntrophic bacteria, convert organic acids and alcohols into acetate, formate, CO_2, and H_2. These compounds feed methanogenic Archaea. The enzymes and metabolic pathways in each partner in the ecosystem have evolved in response to the opportunities presented by other microbes in the system, as well as to competition from other microbes that can utilize similar carbon sources.

8.02.5.2   Gene Duplication and Constraints on Enzyme Evolution Imposed by the Need to Maintain an Original Activity

The existence of families, superfamilies, and suprafamilies of enzymes that have diverged from a common ancestor testifies to the importance of gene duplication in producing the current enzymatic repertoire. Gene duplication allows copies of a gene encoding a generalist enzyme to diverge toward two enzymes with specialized activities. Gene duplication also allows evolution of novel activities originating from promiscuous activities of enzymes. Promiscuous activities occur adventitiously as a result of the assemblage of catalytic residues and cofactors in active sites. Many promiscuous reactions accomplish transformations similar to the normal reaction, but using alternative substrates. However, this is not always the case. For example, *o*-succinylbenzoate synthetase from *Amycolaptosis* sp. has a promiscuous *N*-acylamino acid racemase activity,[57] and tetrachlorohydroquinone dehalogenase from *Sphingobium chlorophenolicum* has a promiscuous maleylacetone isomerase activity.[99] Promiscuous activities serve no particular function in the organism, and thus are not subject to selection.[100] Promiscuous activities provide an important reservoir of novel catalytic activities; even though promiscuous activities are generally orders of magnitude less efficient than well-evolved activities[101–104], they may enhance reaction rates by orders of magnitude relative to uncatalyzed reactions.[103,104] Thus, if the environment changes, a promiscuous activity may provide an important function that enhances fitness or even allows survival.

The relative timing of the gene duplication and mutations that enhance an initially poor activity during evolution of a new enzyme has been debated for decades. The most widely known model for divergent

Table 1 Examples of alternative enzymes in *E. coli* that catalyze formation of a common product using different reactants

Gene	Enzyme	Reactions
mqo	Malate:quinoneoxidoreductase	Malate + Q → **oxaloacetate** + Q
mdh	Malatedehydrogenase	Malate + NAD → **oxaloacetate** + NADH
asnA	Aspartate-ammonia ligase	Asp + ATP + NH$_3$ → **Asn** + AMP + PP$_i$
asnB	Asparaginesynthetase B	Asp + ATP + Gln → **Asn** + AMP + Glu
metE	Cobalamin-independent homocysteinetransmethylase	Homocysteine + MTHF → **Met** + THF (Cob independent)
meth	Cobalamin-dependent homocysteinetransmethylase	Homocysteine + MTHF → **Met** + THF
purN	GAR transformylase 1	GAR + FTHF → **FGAR** + THF
purT	GAR transformylase 2	GAR + ATP + formate → **FGAR** + ADP + P$_i$
hemF	O$_2$-dependent coproporphyrinogen III oxidase	Coproporphyrinogen III + 2 O$_2$ → **protoporphyrinogen** IX + 2 H$_2$O$_2$ + 2 CO$_2$
hemN	O$_2$-independent coproporphyrinogen III oxidase	Coproporphyrinogen III + 2 SAM → **protoporphyrinogen** IX + 25′deoxyadenosine + Met + 2 CO$_2$

Variable reactants in each case are underlined, and the common product is shown in bold.
FGAR, 5′-phosphoribosyl *N*-formylglycineamide; FTHF, N^{10}-formyl-tetrahydrofolate; GAR, 5-phosphoribosylglycineamide; hCys, homocysteine; MTHF, 5-methyl-THF; Q, quinone such as ubiquinone; SAM, *S*-adenosylmethionine; THF, tetrahydrofolate.
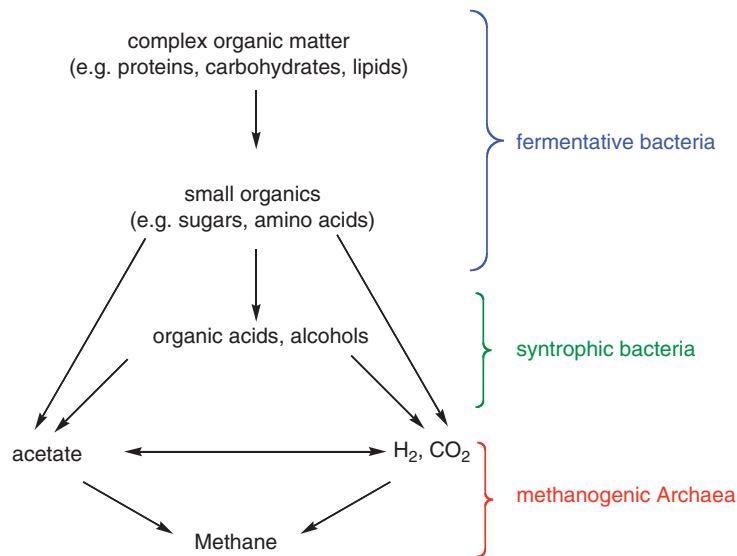
Figure 26 The evolutionary niche occupied by microbes is created by the other microbes in the environment. Adapted from T. Kosaka; S. Kato; T. Shimoyama; S. Ishii; T. Abe; K. Watanabe, *Genome Res.* **2008**, *18*, 442.

evolution of novel functions was proposed by Ohno, who suggested in 1970 that gene duplication allows one copy of a gene to accumulate mutations required for emergence of a new activity, while the other continues to provide the original function.[105] A problem with this hypothesis is that mutations are far more likely to destroy the function of a protein than to create a new function.

An alternative scenario proposed by Hughes[27] in 1994 suggests that new enzymes emerge by mutations that generate a novel secondary activity *prior to* gene duplication. If a secondary activity enhances fitness, selective pressure to retain a gene copy would exist immediately upon duplication; the two copies could then diverge to provide two different functions. Similarly, for an ancestral generalist enzyme, gene duplication can result in division of the properties of the original enzyme between the enzymes encoded by the two copies.[106] This hypothesis predicts that a period of gene sharing before gene duplication occurs during evolution of most, or perhaps all, novel enzymes.

If a period of gene sharing is the rule, then the evolutionary trajectory for improvement of a new activity will be constrained by a number of factors, including how mutations that enhance the new activity affect the old activity, how critical each activity is for fitness, the degree to which the substrate for each reaction inhibits the other reaction, and the capacity for the cell to adapt to enzymatic deficits by increasing enzyme expression.

The effect of mutations that enhance a new activity on the original activity is a particularly critical issue. Given that most active sites have evolved to optimally align the substrate with respect to catalytic groups in the active site, it might be expected that mutations that enhance a novel activity would impair the original activity. This supposition is supported by many examples of enhancements in promiscuous activities achieved by either site-directed mutagenesis[107–109] or *in vitro* evolution.[110] For example, a point mutation in alanine racemase increases a promiscuous aldolase activity by 2.3×10^5 but decreases the original activity by 4×10^3.[108] Surprisingly, however, this is not always the case. Khersonsky *et al.* reviewed 11 cases in which substantial increases in a promiscuous activity (10–10^6-fold) were achieved by mutations that caused only a small decrease in the original activity (<42-fold).[111] The idea that a novel activity can emerge without a significant compromise of the original activity is appealing, as it provides 'something for nothing'. Such cases are especially promising for evolution of a novel activity.

The degree to which the original and novel activities are required for growth is a second important factor. A new activity may become critical for growth because a resource has become limited and a new enzyme or pathway is needed to supply precursors for biomolecules, or because a toxin has appeared in the environment. A promiscuous activity in an enzyme that is not critical for survival would be the optimal starting place for

evolution of the newly needed enzyme. However, the necessary promiscuous activity may only be found in an enzyme that is critical for growth. In such cases, a mutation that compromises the existing activity but provides the critical new activity may be tolerated if it allows survival and/or growth. Growth of the mutant strain may be slow, but growth of competitors that cannot adapt to the new environment might be abolished.

Competition for the active site by the substrates for the two reactions is a third factor affecting the evolution of novel activities during a period of gene sharing. (This is a problem still faced by generalist enzymes today.) Each substrate would be a competitive inhibitor of the other reaction. Thus, flux through each reaction would depend not only on the k_{cat} and K_M for that substrate, but on the k_{cat} and K_M for the other substrate. In some cases, a decrease in the affinity for the original substrate might enhance fitness by diminishing what would otherwise be potent inhibition of the novel activity.

The implications of gene sharing during evolution of a novel enzyme have received little attention. A recent study addresses the early stages in the evolution of a novel activity under circumstances in which both the original and the novel activity are critical for growth.[112] ProA (glutamyl phosphate reductase) has a very low promiscuous activity with *N*-acetylglutamylphosphate, the normal substrate for ArgC (*N*-acetylglutamyl phosphate reductase). A mutation that changes Glu383 to Ala increases the promiscuous activity by 12-fold, but decreases the original activity by 2800-fold. The resulting impairment in proline and arginine synthesis results in 20-fold overexpression of E383A ProA, most likely via activation of the stringent response, which up-regulates transcription of amino acid biosynthesis genes under starvation conditions.[113,114] The combined effect of the increase in *N*-acetylglutamylphosphate reductase activity and overexpression of E383A ProA allows a strain lacking ArgC to make enough arginine and proline to grow on glucose. This study provides an example of a situation in which considerable damage to the original activity can be tolerated if the novel activity is essential for growth. It also demonstrates the importance of physiological changes due to a decrease in a normal activity caused by a mutation; overexpression can enhance the level of both the original and the novel activities, and this may be required to exceed the threshold level required to sustain growth.

8.02.5.3   Topological Effects in Metabolic Networks

Metabolic enzymes operate in the context of a metabolic network that must provide energy as well as the precursors of biomolecules. The topology of the network would be expected to influence the evolution of novel enzyme activities. Innovations that enhance a newly needed activity but compromise an original activity may not be tolerated in enzymes that produce metabolites required in many pathways. Innovations that compromise an original activity in linear pathways may be more harmful than innovations in dense parts of the metabolic network where diversion of flux through alternative pathways can allow production of critical metabolites even if a particular enzyme is temporarily compromised.

8.02.6   Enzyme Evolution in the Modern Age

After 3.8 billion years of evolution, living organisms rely on a collection of impressive enzymes that accelerate reaction rates by many orders of magnitude, provide exquisite control of the orientation and reactivity of substrates and intermediates at active sites, and respond to changes in environmental conditions. However, the process of enzyme evolution is not over. Novel enzymes are still evolving in response to anthropogenic influences. Introduction of antibiotics and pesticides in the past century has created new selective pressures for evolution of enzymes that detoxify and/or degrade such compounds.
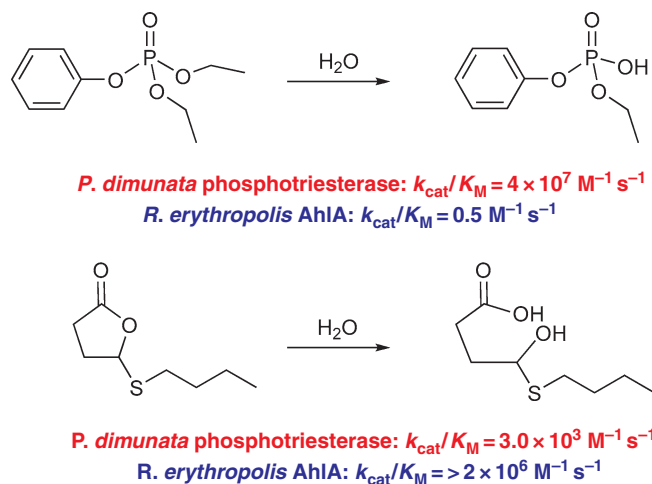
The introduction of antibiotics led to an enormous decrease in the morbidity and mortality caused by bacterial infections. Widespread use of antibiotics in the clinic, in animal feed, and in consumer products has resulted in astonishingly rapid evolution of antibiotic resistance, typically within a few decades, or even a few years, after introduction of a new antibiotic. Resistance can occur via multiple mechanisms, including detoxification by enzymes as well as target modification, decreased membrane permeability, and use of efflux pumps to expel antibiotics from cells. In many cases, antibiotic resistance genes are already present in the environment as a result of self-protection mechanisms in organisms that produce antibiotics, or of exposure of

microbes to antibiotics produced by other organisms in competitive natural environments. Horizontal gene transfer eventually results in acquisition of antibiotic resistance genes by pathogens from natural reservoirs.

A variety of enzymatic mechanisms for antibiotic resistance are known. Hydrolysis of the lactam rings of β-lactams, cephalosporins, and carbapenams destroys their ability to inhibit transpeptidases that cross-link peptidoglycan in bacterial cell walls. Modification of aminoglycoside antibiotics by acetylation, phosphorylation, or adenylation interferes with their ability to bind to the 16S subunit of the ribosome.[115] Streptogramin activity can be destroyed by acetylation or by an elimination reaction that opens the lactone ring.[115] The enzymes responsible for these detoxification reactions evolved in response to naturally occurring antibiotics, but are easily adapted to modify semisynthetic and completely synthetic antibiotics. For example, only a few point mutations are needed to enhance the ability of TEM β-lactamases to hydrolyze third-generation cephalosporins such as cefotaxime and ceftazidime.[116,117]

Detoxification generally requires only a single enzymatic step. However, microbes can also evolve pathways to degrade novel compounds by recruitment of one or more enzymes to transform a novel compound into a metabolite in a standard metabolic pathway. For example, organophosphate insecticides can be degraded to provide phosphorus, a nutrient that is often limiting in soil environments. Indeed, some bacteria can use organophosphates as a sole source of phosphorus.[118] Diverse microbes including *Pseudomonas dimunata*, *Flavobacterium* sp., and *Agrobacterium radiobacter* contain a phosphotriesterase that initiates degradation of organophosphate insecticides, and is believed to have evolved in response to the introduction of synthetic insecticides in the twentieth century.[119] Phosphotriesterases may have evolved from lactonases that cleave *N*-acyl homoserine lactones used in quorum-sensing. Although the sequence identities between phosphotriesterases and *N*-acyl homoserine lactonases are rather low (<35%), these enzymes share key active site features, including a zinc atom and most of its ligands. In addition, phosphotriesterases have low-level promiscuous activity with various lactones, and *N*-acyl homoserine lactonases have promiscuous activity with organophosphates (see **Figure 27**). Notably, the efficiency of phosphotriesterase is very high − k_{cat}/K_M for its best substrate is $>4 \times 10^7 \, \mathrm{l\,mol^{-1}\,s^{-1}}$.[120] This may mean that only a few decades have been sufficient to evolve a highly efficient enzyme, although the possibility that a naturally occurring substrate drove the evolution of this enzyme over a longer period of time cannot be ruled out.

A second example is the pathway for degradation of pentachlorophenol (PCP), a pesticide that was introduced in the 1930s. PCP is highly toxic because it uncouples oxidative phosphorylation and perturbs membrane properties. Furthermore, some of its degradation products are mutagenic. Remarkably, bacteria that can completely degrade PCP have been isolated from contaminated environments.[121,122] PCP degradation has been most thoroughly studied in the Gram-negative *S. chlorophenolicum* (see **Figure 28**). The first four steps in the pathway are catalyzed by enzymes that appear to have been recently recruited to serve these new functions,

P. dimunata phosphotriesterase: $k_{cat}/K_M = 4 \times 10^7 \, \mathrm{M^{-1}\,s^{-1}}$
R. erythropolis AhlA: $k_{cat}/K_M = 0.5 \, \mathrm{M^{-1}\,s^{-1}}$

P. dimunata phosphotriesterase: $k_{cat}/K_M = 3.0 \times 10^3 \, \mathrm{M^{-1}\,s^{-1}}$
R. erythropolis AhlA: $k_{cat}/K_M = >2 \times 10^6 \, \mathrm{M^{-1}\,s^{-1}}$

Figure 27 Comparison of the abilities of phosphotriesterase from *Pseudomonas dimunata* and a lactonase from *Rhodococcus erythropolis* (AhlA) to hydrolyze a synthetic pesticide, paraoxon, and a lactone, 4-thio-butyl butyro-lactone.
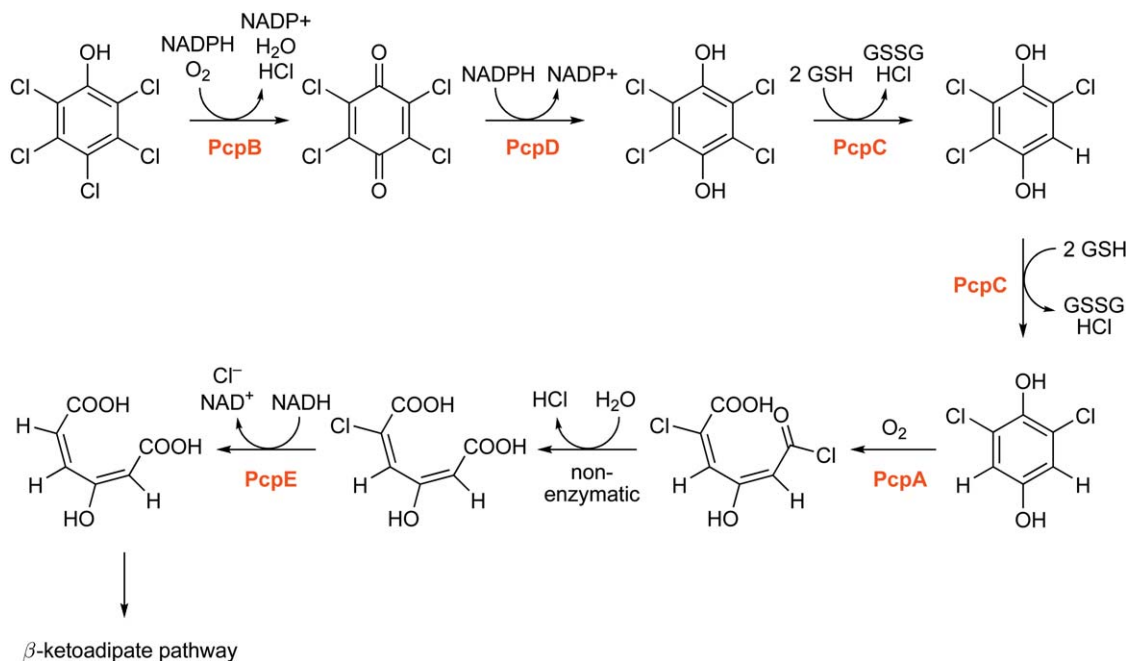
Figure 28 Pathway for degradation of pentachlorophenol in *Sphingobium chlorophenolicum*. PcpB, PCP hydroxylase; PcpD, TCBQ reductase; PcpC, TCHQ dehalogenase; PcpA, DCHQ dioxygenase; GSH, glutathione.

and all four are rather poor catalysts. The first enzyme, PCP hydroxylase, is a member of a large family of flavin monooxygenases that initiate degradation of phenolic compounds. It turns over substrate very slowly, with a k_{cat} of $0.02\,s^{-1}$ (as compared to typical k_{cat} values of $25–100\,s^{-1}$ in this family (Pietari, Behlen, and Copley, unpublished results). Furthermore, the formation of the reactive C4a-hydroperoxyflavin intermediate is poorly coupled to the hydroxylation of the substrate. Approximately 67% of the C4a-hydroperoxyflavin undergoes elimination to form oxidized flavin and H_2O_2. This reaction wastes valuable NADPH and generates a strong oxidant in the cytoplasm. The second enzyme, tetrachlorobenzoquinone reductase, is also rather sluggish, with a turnover number of only $0.7\,s^{-1}$ in the presence of $50\,\mu\,mol\,l^{-1}$ tetrachlorobenzoquinone and $1\,m\,mol\,l^{-1}$ NADPH.[123] The third enzyme, tetrachlorohydroquinone dehalogenase, is profoundly inhibited by its aromatic substrate (see **Figure 29(a)**).[124] The reaction is a ping-pong reaction (see **Figure 29(b)**); binding of the aromatic substrate (tetrachlorohydroquinone or trichlorohydroquinone) to the covalent adduct between glutathione and the enzyme prevents the thiol–disulfide exchange reaction that completes the catalytic cycle. Finally, the fourth enzyme, 2,6,-dichlorohydroquinone dioxygenase, undergoes rapid substrate-dependent inactivation during catalytic turnover (K. Hazell, Master's thesis, 2002). Extradiol dioxygenases in this family are inactivated frequently during turnover of poor substrates, but rarely during turnover of good substrates.[125] Taken together, these findings support the hypothesis that the pathway for degradation of PCP has been recently patched together. In contrast to the situation with phosphotriesterase, none of these enzymes has evolved to a high level of efficiency. This is likely due to the peculiarities of the pathway; the first step in the pathway, catalyzed by PCP hydroxylase, limits the flux through the pathway. The product formed from PCP, tetrachlorobenzoquinone, is more toxic than PCP itself. Thus, mutations that increase PCP hydroxylase activity would expose the cells to higher levels of toxic tetrachlorobenzoquinone, and actually decrease fitness. Futhermore, because the flux through the pathway is limited by the inefficiency of PCP hydroxylase, there should be no selective pressure for evolution of more effective catalysts for the subsequent steps in the pathway.
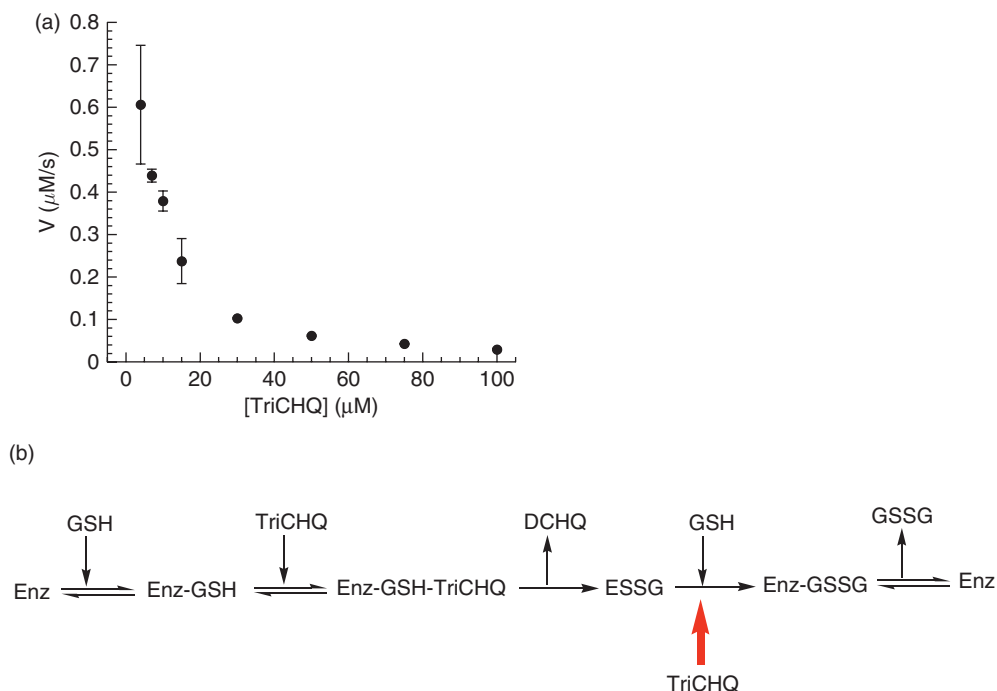
Figure 29 (a) Effect of TriCHQ on the rate of the dehalogenation reaction catalyzed by TCHQ dehalogenase. (b) Kinetic mechanism for TCHQ dehalogenase. Binding of TriCHQ to ESSG (red arrow) inhibits the thiol–disulfide exchange reaction that generates glutathione disulfide and free enzyme. GSH, glutathione; TriCHQ, trichlorohydroquinone; DCHQ, 2,6-dichlorohydroquinone; GSSG, glutathione disulfide; ESSG, covalent adduct between glutathione and Cys13 of TCHQ dehalogenase.

For 3.8 billion years, enzyme evolution has occurred primarily in microbes exposed to novel environmental conditions. However, in the last two decades, new methods have been developed for laboratory evolution of enzymes for production of chemicals, pharmaceuticals, and biofuels. Directed evolution has been widely used to improve thermostability and alter substrate specificity. Current efforts aim to improve the catalytic abilities of evolved enzymes, which are usually considerably poorer than those of naturally occurring enzymes, and to evolve novel pathways using promiscuous activities of existing enzymes. These efforts will provide new insights into the adaptation of protein scaffolds for new functions that will both help us to understand the evolutionary history of modern enzymes and provide the basis for a wide range of applications in biotechnology.

Abbreviations

ATP	adenosine triphosphate
CoA	coenzyme A
CPI	4-(4-chlorophenyl)imidazole
DHFR	dihydrofolate reductase
FAD	flavin adenine dinucleotide
FMN	flavin mononucleotide
GFP	green fluorescent protein
GMP	guanosine monophosphate
HAD	haloacid dehalogenase
HOMO	highest occupied molecular orbital
LUCA	last universal common ancestor
LUMO	lowest unoccupied molecular orbital

MIC	minimal inhibitory concentration
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NHDP	nicotinamide hypoxanthine dinucleotide phosphate
NMN	nicotinamide mononucleotide
NMR	nuclear magnetic resonance spectroscopy
OFP	orange fluorescent protein
PCP	pentachlorophenol
PDB	Protein Data Bank
PLP	pyridoxal phosphate
PRPP	5-phosphoribosyl-1-pyrophosphate
RMSD	root mean square deviation

References

1. B. G. Miller; R. Wolfenden, *Annu. Rev. Biochem.* **2002**, *71*, 847–885.
2. C. R. Woese, *The Genetic Code: The Molecular Basis for Genetic Expression.* Harper and Row: New York, 1967.
3. K. Kruger; P. J. Grabowski; A. J. Zaug; J. Sands; D. E. Gottschling; T. R. Cech, *Cell* **1982**, *31*, 147–157.
4. C. Guerrier-Takada; K. Gardiner; T. Marsh; N. Pace; S. Altman, *Cell* **1983**, *35*, 849–857.
5. R. K. Kumar; M. Yarus, *Biochemistry* **2001**, *40*, 6998–7004.
6. F. Huang; C. W. Bugg; M. Yarus, *Biochemistry* **2000**, *39*, 15548–15555.
7. M. Illangasekare; M. Yarus, *RNA* **1999**, *5*, 1482–1489.
8. S. Fusz; A. Eisenfuhr; S. G. Srivatsan; A. Heckel; M. Famulok, *Chem. Biol.* **2005**, *12*, 941–950.
9. R. A. Jensen, *Ann. Rev. Microbiol.* **1976**, *30*, 409–425.
10. C. R. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6854–6859.
11. S. Abeln; C. M. Deane, *Proteins* **2005**, *60*, 690–700.
12. D. Lee; A. Grant; R. L. Marsden; C. Orengo, *Proteins* **2005**, *59*, 603–615.
13. J. A. Ranea; A. Sillero; J. M. Thornton; C. A. Orengo, *J. Mol. Evol.* **2006**, *63*, 513–525.
14. B. G. Mirkin; T. I. Fenner; M. Y. Galperin; E. V. Koonin, *BMC Evol. Biol.* **2003**, *3*, 2.
15. C. A. Ouzounis; V. Kunin; N. Darzentas; L. Goldovsky, *Res. Microbiol.* **2006**, *157*, 57–68.
16. C. Koeberl, *Elements* **2006**, *2*, 211–216.
17. S. J. Mojzsis; G. Arrhenius; K. D. McKeegan; T. M. Harrison; A. P. Nutman; C. R. Friend, *Nature* **1996**, *384*, 55–59.
18. M. D. Toney, *Arch. Biochem. Biophys.* **2005**, *433*, 279–287.
19. P. Christen; P. K. Mehta, *Chem Rec.* **2001**, *1*, 436–447.
20. C. F. Barbas, III; A. Heine; G. Zhong; T. Hoffman; S. Gramatikova; R. Bjornestedt; B. List; J. Anderson; E. A. Stura; I. A. Wilson; R. A. Lerner, *Science* **1997**, *278*, 2085–2092.
21. Y. Zhao; M. A. White; B. K. Muralidhara; L. Sun; J. R. Halpert; C. D. Stout, *J. Biol. Chem.* **2006**, *281*, 5973–5981.
22. T. Niwa; N. Murayama; C. Emoto; H. Yamakazi, *Curr. Drug Metab.* **2008**, *9*, 20–33.
23. J. S. McCune; L. J. Risler; B. R. Phillips; K. E. Thummel; D. Blough; D. D. Shen, *Drug Metab. Dispos.* **2005**, *33*, 1074–1081.
24. T. Prueksaritanont; B. Ma; N. Yu, *Br. J. Clin. Pharmacol.* **2003**, *56*, 120–124.
25. F. Barona-Gomez; D. A. Hodgson, *EMBO Rep.* **2003**, *4*, 296–300.
26. H. Wright; L. Noda-Garcia; A. Ochoa-Leyva; D. A. Hodgson; V. Fulop; F. Barona-Gomez, *Biochem. Biophys. Res. Commun.* **2008**, *365*, 16–21.
27. A. L. Hughes, *Proc. R. Soc. Lond. B.* **1994**, *1994*, 119–124.
28. D. Lee; A. Grant; R. L. Marsden; C. Orengo, *Proteins* **2005**, *59*, 603–615.
29. C. A. Orengo; A. D. Michie; S. Jones; D. T. Jones; M. B. Swindells; J. M. Thornton, *Structure* **1997**, *5*, 1093–1108.
30. A. G. Murzin; S. E. Brenner; T. Hubbard; C. Chothia, *J Mol Biol.* **1995**, *247*, 536–540.
31. C. Bakolitsa; R. Schwarzenbacher; D. McMullan; L. S. Brinen; J. M. Canaves; X. Dai; A. M. Deacon; M. A. Elsliger; S. Eshagi; R. Floyd; A. Godzik; C. Grittini; S. K. Grzechnik; L. Jaroszewski; C. Karlak; H. E. Klock; E. Koesema; J. S. Kovarik; A. Kreusch; P. Kuhn; S. A. Lesley; T. M. McPhillips; M. D. Miller; A. Morse; K. Moy; J. Ouyang; R. Page; K. Quijano; A. Robb; G. Spraggon; R. C. Stevens; H. van den Bedem; J. Velasquez; J. Vincent; F. von Delft; X. Wang; B. West; G. Wolf; K. O. Hodgson; J. Wooley; I. A. Wilson, *Proteins* **2004**, *56*, 607–610.
32. R. A. Goldstein, *Curr. Opin. Struct. Biol.* **2008**, *18*, 170–177.
33. B. E. Shakhnovich; E. Deeds; C. Delisi; E. Shakhnovich, *Genome Res.* **2005**, *15*, 385–392.
34. Y. Zhang; I. A. Hubner; A. K. Arakaki; E. Shakhnovich; J. Skolnick, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2605–2610.
35. N. V. Grishin, *J. Struct. Biol.* **2001**, *134*, 167–185.
36. R. L. Tuinstra; F. C. Peterson; S. Kutlesa; E. S. Elgin; M. A. Kron; B. F. Volkman, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5057–5062.
37. S. V. Chittur; Y. Chen; V. J. Davisson, *Protein Expr. Purif.* **2000**, *18*, 366–377.
38. S. Beismann-Driemeyer; R. Sterner, *J. Biol. Chem.* **2001**, *276*, 20387–20396.

39. R. L. Marsden; D. Lee; M. Maibaum; C. Yeats; C. A. Orengo, *Nucleic Acids Res.* **2006**, *34*, 1066–1080.
40. G. Apic; J. Gough; S. A. Teichmann, *J. Mol. Biol.* **2001**, *310*, 311–325.
41. G. J. Bartlett; C. T. Porter; N. Borkakoti; J. M. Thornton, *J. Mol. Biol.* **2002**, *324*, 105–121.
42. R. M. de Jong; W. Brugman; G. J. Poelarends; C. P. Whitman; B. W. Dijkstra, *J. Biol. Chem.* **2004**, *279*, 11546–11552.
43. M. Tsai; J. Koo; P. Yip; R. F. Colman; M. L. Segall; P. L. Howell, *J. Mol. Biol.* **2007**, *370*, 541–554.
44. A. Weeks; L. Lund; F. M. Raushel, *Curr. Opin. Chem. Biol.* **2006**, *10*, 465–472.
45. R. S. Myers; J. R. Jensen; I. L. Deras; J. L. Smith; V. J. Davisson, *Biochemistry* **2003**, *42*, 7013–7022.
46. W. Zhang; Y. Zhou; D. F. Becker, *Biochemistry* **2004**, *43*, 13165–13174.
47. W. Zhang; M. Zhang; W. Zhu; Y. Zhou; S. Wanduragala; D. Rewinkel; J. J. Tanner; D. F. Becker, *Biochemistry* **2007**, *46*, 483–491.
48. C. M. Seibert; F. M. Raushel, *Biochemistry* **2005**, *44*, 6383–6391.
49. A. M. Burroughs; K. N. Allen; D. Dunaway-Mariano; L. Aravind, *J. Mol. Biol.* **2006**, *361*, 1003–1034.
50. S. D. Lahiri; G. Zhang; J. Dai; D. Dunaway-Mariano; K. N. Allen, *Biochemistry* **2004**, *43*, 2812–2820.
51. M. C. Morais; W. Zhang; A. S. Baker; G. Zhang; D. Dunaway-Mariano; K. N. Allen, *Biochemistry* **2000**, *39*, 10385–10396.
52. I. S. Ridder; H. J. Rozeboom; K. H. Kalk; B. W. Dijkstra, *J. Biol. Chem.* **1999**, *274*, 30672–30678.
53. J. A. Gerlt; P. C. Babbitt; I. Rayment, *Arch. Biochem. Biophys.* **2005**, *433*, 59–70.
54. D. M. Weinreich; N. F. Delaney; M. A. DePristo; D. L. Hartl, *Science* **2006**, *312*, 111–114.
55. M. A. Norrgard; Y. Ivarsson; K. Tars; B. Mannervik, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 4876–4881.
56. J. D. Bloom; S. T. Labthavikul; C. R. Otey; F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5869–5874.
57. D. R. Palmer; J. B. Garrett; V. Sharma; R. Meganathan; P. C. Babbitt; J. A. Gerlt, *Biochemistry* **1999**, *38*, 4252–4258.
58. J. D. Bloom; P. A. Romero; Z. Lu; F. H. Arnold, *Biol. Direct* **2007**, *2*, 17.
59. J. D. Bloom; Z. Lu; D. Chen; A. Raval; O. S. Venturelli; F. H. Arnold, *BMC Biol.* **2007**, *5*, 29.
60. B. Shen, *Curr. Opin. Chem. Biol.* **2003**, *7*, 285–295.
61. C. Marco-Marin; F. Gil-Ortiz; I. Perez-Arellano; J. Cervera; I. Fita; V. Rubio, *J. Mol. Biol.* **2007**, *367*, 1431–1446.
62. S. An; R. Kumar; E. D. Sheets; S. J. Benkovic, *Science* **2008**, *320*, 103–106.
63. P. Nordlund; P. Reichard, *Annu. Rev. Biochem.* **2006**, *75*, 681–706.
64. O. Herzberg; C. C. Chen; G. Kapadia; M. McGuire; L. J. Carroll; S. J. Noh; D. Dunaway-Mariano, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2652–2657.
65. D. Joseph; G. A. Petsko; M. Karplus, *Science* **1990**, *249*, 1425–1428.
66. W. S. Bennett, Jr.; T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 4848–4852.
67. S. D. Lahiri; G. Zhang; D. Dunaway-Mariano; K. N. Allen, *Bioorg. Chem.* **2006**, *34*, 394–409.
68. F. Karush, *J. Am. Chem. Soc.* **1950**, *72*, 2705–2712.
69. D. E. Koshland, *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98–104.
70. A. D. Mesecar; B. L. Stoddard; D. E. Koshland, Jr., *Science* **1997**, *277*, 202–206.
71. D. L. Gatti; B. A. Palfey; M. S. Lah; B. Entsch; V. Massey; D. P. Ballou; M. L. Ludwig, *Science* **1994**, *266*, 110–114.
72. A. Kohen; R. Cannio; S. Bartolucci; J. P. Klinman, *Nature* **1999**, *399*, 496–499.
73. L. Wang; N. M. Goodey; S. J. Benkovic; A. Kohen, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15753–15758.
74. J. L. Radkiewicz; C. L. Brooks, III, *J. Am. Chem. Soc.* **2000**, *122*, 225–231.
75. P. K. Agarwal; S. R. Billeter; P. T. Rajagopalan; S. J. Benkovic; S. Hammes-Schiffer, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2794–2799.
76. G. S. B. Williams; A. M. Hossain; S. Shang; D. E. Kranbuehl; C. K. Bagdassarian, *J. Theor. Comput. Chem.* **2003**, *2*, 323–334.
77. G. S. B. Williams; A. M. Hossain; D. E. Kranbuehl; C. K. Bagdassarian, *J. Phys. Chem. B* **2003**, *107*, 12527–12533.
78. L. J. Rothschild; R. L. Mancinelli, *Nature* **2001**, *409*, 1092–1101.
79. E. Blochl; R. Rachel; S. Burggraf; D. Hafenbradl; H. W. Jannasch; K. O. Stetter, *Extremophiles* **1997**, *1*, 14–21.
80. E. A. Galinksi; H. G. Trüper, *FEMS Microbiol. Rev.* **1994**, *15*, 95–108.
81. Y. Yonezawa; H. Tokunaga; M. Ishibashi; M. Tokunaga, *Biosci. Biotechnol. Biochem.* **2001**, *65,* 2343–2346.
82. M. Mevarech; F. Frolow; L. M. Gloss, *Biophys. Chem.* **2000**, *86*, 155–164.
83. O. Dym; M. Mevarech; J. L. Sussman, *Science* **1995**, *267*, 1344–1346.
84. A. M. Suiter; O. Banziger; A. M. Dean, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12782–12786.
85. A. M. Suiter; A. M. Dean, *J. Mol. Evol.* **2005**, *61*, 153–170.
86. D. G. Gourley; A. K. Shrive; I. Polikarpov; T. Krell; J. R. Coggins; A. R. Hawkins; N. W. Isaacs; L. Sawyer, *Nat. Struct. Biol.* **1999**, *6*, 521–525.
87. M. Yun; C. G. Park; J. Y. Kim; C. O. Rock; S. Jackowski; H. W. Park, *J. Biol. Chem.* **2000**, *275*, 28093–28099.
88. B. S. Hong; M. K. Yun; Y. M. Zhang; S. Chohnan; C. O. Rock; S. W. White,; S. Jackowski; H. W. Park; R. Leonardi, *Structure* **2006**, *14*, 1251–1261.
89. E. S. Rangarajan; J. Sivaraman; A. Matte; M. Cygler, *Proteins* **2002**, *48*, 737–740.
90. R. G. Zhang; C. E. Andersson; T. Skarina; E. Evdokimova; A. M. Edwards; A. Joachimiak; A. Savchenko; S. L. Mowbray, *J. Mol. Biol.* **2003**, *332*, 1083–1094.
91. H. Myllykallio; G. Lipowski; D. Leduc; J. Filee; P. Forterre; U. Liebl, *Science* **2002**, *297*, 105–107.
92. C. P. Tseng; C. C. Yu; H. H. Lin; C. Y. Chang; J. T. Kuo, *J. Bacteriol.* **2001**, *183*, 461–467.
93. B. L. Van Kuijk; N. D. Van Loo; A. F. Arendsen; W. R. Hagen; A. J. Stams, *Arch. Microbiol.* **1996**, *165*, 126–131.
94. N. Yumoto; M. Tokushige, *Biochem. Biophys. Res. Commun.* **1988**, *153*, 1236–1243.
95. G. Layer; K. Verfurth; E. Mahlitz; D. Jahn, *J. Biol. Chem.* **2002**, *277*, 34136–34142.
96. D. A. Peterson; D. N. Frank; N. R. Pace; J. I. Gordon, *Cell Host Microbe* **2008**, *3*, 417–427.
97. S. G. Tringe; C. von Mering; A. Kobayashi; A. A. Salamov; K. Chen; H. W. Chang; M. Podar; J. M. Short; E. J. Mathur; J. C. Detter; P. Bork; P. Hugenholtz; E. M. Rubin, *Science* **2005**, *308*, 554–557.
98. C. M. Santelli; B. N. Orcutt; E. Banning; W. Bach; C. L. Moyer; M. L. Sogin; H. Staudigel; K. J. Edwards, *Nature* **2008**, *453*, 653–656.
99. K. Anandarajah; P. M. Kiefer; S. D. Copley, *Biochemistry* **2000**, *39*, 5303–5311.

100. S. D. Copley, *Curr. Opin. Chem. Biol.* **2003**, *7*, 265–272.
101. E. A. Taylor Ringia; J. B. Garrett; J. B. Thoden; H. M. Holden; I. Rayment; J. A. Gerlt, *Biochemistry* **2004**, *43*, 224–229.
102. S. C. Wang; W. H. Johnson, Jr.; C. P. Whitman, *J. Am. Chem. Soc.* **2003**, *125*, 14282–14283.
103. P. J. O'Brien; D. Herschlag, *J. Am. Chem. Soc.* **1998**, *120*, 12369–12370.
104. P. J. O'Brien; D. Herschlag, *Biochemistry* **2001**, *40*, 5691–5699.
105. S. Ohno, *Evolution by Gene Duplication.* Springer-Verlag: New York, 1970.
106. A. Force; M. Lynch; F. B. Pickett; A. Amores; Y. L. Yan; J. Postlethwait, *Genetics* **1999**, *151*, 1531–1545.
107. H.-S. Park; S.-H. Nam; J. K. Lee; C. N. Yoon; B. Mannervik; S. J. Benkovic; H.-S. Kim, *Science* **2006**, *311*, 535–538.
108. F. P. Seebeck; D. Hilvert, *J. Am. Chem. Soc.* **2003**, *125*, 10158–10159.
109. Y. Yin; J. F. Kirsch, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 17353–17357.
110. C. Jurgens; A. Strom; D. Wegener; S. Hettwer; M. Wilmanns; R. Sterner, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9925–9930.
111. O. Khersonsky; C. Roodveldt; D. A. Tawfik, *Curr. Opin. Chem. Biol.* **2006**, *10*, 498–508.
112. S. Yu McLoughlin; S. D. Copley, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13497–13502.
113. P. L. Foster, *Crit. Rev. Biochem. Mol. Biol.* **2007**, *42*, 373–397.
114. L. U. Magnusson; A. Farewell; T. Nystrom, *Trends Microbiol.* **2005**, *13*, 236–242.
115. G. D. Wright, *Curr. Opin. Chem. Biol.* **2003**, *7*, 563–569.
116. A. Petit; L. Maveyraud; F. Lenfant; J. P. Samama; R. Labia; J. M. Masson, *Biochem. J.* **1995**, *305* (Pt 1), 33–40.
117. X. Raquet; J. Lamotte-Brasseur; E. Fonze; S. Goussard; P. Courvalin; J. M. Frere, *J. Mol. Biol.* **1994**, *244*, 625–639.
118. S. Singh; D. K. Singh, *Can. J. Microbiol.* **2003**, *49*, 101–109.
119. F. M. Raushel; H. M. Holden, *Adv. Enzymol. Relat. Areas Mol. Biol.* **2000**, *74*, 51–93.
120. C. Roodveldt; D. S. Tawfik, *Biochemistry* **2005**, *44*, 12728–12736.
121. R. L. Crawford; M. M. Ederer, *J. Ind. Microbiol. Biotechnol.* **1999**, *23*, 320–325.
122. M. Takeuchi; H. Koei; H. Akira, *Int. J. Syst. Evol. Microbiol.* **2001**, *51*, 1405–1417.
123. M. Dai; J. Bull Rogers; J. R. Warner; S. D. Copley, *J. Bacteriol.* **2003**, *185*, 302–310.
124. J. R. Warner; S. D. Copley, *Biochemistry* **2007**, *46*, 4438–4447.
125. F. H. Vaillancourt; G. Labbe; N. M. Drouin; P. D. Fortin; L. D. Eltis, *J. Biol. Chem.* **2002**, *277*, 2019–2027.

Biographical Sketch

Shelley Copley received her A.B. in Biochemistry and Ph.D. in Biophysics from Harvard University. After 3 years as a postdoc at MIT and the University of Colorado, she joined the Department of Chemistry and Biochemistry and the Cooperative Institute for Research in Environmental Sciences at the University of Colorado in 1990. She moved to the Department of Molecular, Cellular, and Developmental Biology in 2000. Her research focuses on the molecular evolution of catalysts and metabolic pathways, and particularly on the evolutionary potential of promiscuous enzyme activities.

8.03 Enzyme Promiscuity – Evolutionary and Mechanistic Aspects

Olga Khersonsky and Dan S. Tawfik, Weizmann Institute of Science, Rehovot, Israel

8.03.1   Introduction

Traditionally, enzymes are referred to as remarkably fast and specific catalysts. The fact that many enzymes are capable of catalyzing other reactions, besides the one they physiologically specialize in, or evolved for, is definitely not new. Since a long time, breaches of specificity, or promiscuity, of enzymes have been recognized. Early examples of enzyme promiscuity include pyruvate decarboxylase,[1] carbonic anhydrase,[2] pepsin,[3] chymotrypsin,[4] and L-asparaginase.[5] However, these early discussions of enzymatic versatility were scarce, and until recently, the wider implications of this 'darker' side of enzymes were largely ignored. During the last decade, protein, and especially enzyme promiscuity, received considerable attention, and their importance in various contexts was systematically studied. Reviews by O'Brien and Herschlag,[6] and later Copley,[7] were the first to highlight the mechanistic and evolutionary implications of promiscuity. Other, more recent reviews have focused on the practical implications of promiscuity in organic synthesis,[1,8,9] on promiscuity and divergence in certain enzyme families,[10–13] on mechanistic aspects of promiscuity,[1,14] and on promiscuity in the context of protein evolution[14] and design.[15]

The primary focus of this chapter is the role of promiscuity in the evolution of new enzyme functions. New enzymes have constantly emerged throughout the natural history of this planet. Over the past decades, enzymes that degrade synthetic chemicals were introduced to the biosystem,[16–20] and enzymes associated with drug resistance,[21–24] provide vivid examples of how rapid the evolution of new enzymatic functions can be. The first direct connection between protein evolution and promiscuity was made in 1976 by Jensen.[25] In his landmark review, Jensen formalized the hypothesis that the starting points for evolution were provided by broad specificity, or promiscuity, of the ancestral enzymes. Jensen proposed that unlike modern enzymes that tend to specialize in one substrate and reaction, the primordial, ancient enzymes possessed very broad specificities, and thus few enzymes could perform many functions. Divergence of specialized enzymes, through duplication, mutation, and selection, led to the current diversity of enzymes, and to increased metabolic efficiency.

Extensive research has been carried out since Jensen provided ample evidence for the idea that promiscuity is a key factor in the evolution of new protein functions. Here, we attempt to survey this accumulated knowledge. We focus on several aspects of promiscuity (and enzyme promiscuity, in particular) with an emphasis on its mechanistic aspects, and its role in enzyme evolution. In contrast to journal articles, this volume allows a very detailed discussion, including numerous examples and sidewalks to several related issues. We therefore provide a comprehensive treatise that begins with the question of generality – are promiscuous activities a rare exception, or should they be considered an accompanying trait of all enzymes (Section 8.03.2)? We continue with an attempt to define promiscuity, and its variable facets, in a more rigorous way, and quantify its levels and magnitude (Sections 8.03.3 and 8.03.4). Following that at present, promiscuity is a widely studied phenomenon, we also discuss the prospects of predicting it (Section 8.03.5). In the next sections, we discuss the structural and mechanistic aspects of promiscuity (Section 8.03.6), and the evolutionary implications (Sections 8.03.7 and 8.03.8). The latter begins with reviewing evidence for the role of promiscuity in the divergent evolution of enzyme families and superfamilies (Section 8.03.7), and continues with a broader discussion of various factors, and mechanisms, that drive the evolution of new enzymatic functions (Section 8.03.8).

8.03.2   Promiscuity – The Rule or an Exception

How wide is the phenomenon of promiscuity? The paradigm of absolute specificity: 'one enzyme = one substrate' dominates our textbooks. But this view is obviously schematic, and is mostly valid as a first approximation. Numerous examples for enzyme promiscuity are currently known, but one might argue that these are largely anecdotal and do not provide a general picture. The paragraphs that follow detail several arguments in favor of the notion that promiscuity, at different levels and magnitudes, is a wide phenomenon, and should thus be treated as a rule, rather than an exception.

Specificity bears a high cost, in substrate binding energies, and hence in k_{cat}.[26] Even the most specific enzymes, for example, enzymes involved in DNA replication or protein synthesis, exhibit measurable substrate infidelities, often at surprisingly high rates.[26] When necessary, high fidelity is achieved through proof-reading mechanisms that involve energy costs, and essentially reverse the process, and repeat it, to correct possible errors. For example, the proof-reading domain of many polymerases is an exonuclease that digests parts of the extending strand.

Specificity is shaped by natural selection. It is therefore context dependent. Cross-reactivities and promiscuous activities that are harmful were selected against. Consider aminoacyl-tRNA synthetases for example. Undoubtedly, their selectivity with respect to tRNAs and amino acids is under tight selection. Because of the close similarity of certain amino acids, proof-editing mechanisms have evolved whereby formation of a noncognate aminoacyl-tRNA is followed by its rapid hydrolysis.[27] Isoleucyl-tRNA synthetase, for example, favors the reaction with isoleucine over valine only by ~100-fold, and valine concentration *in vivo* is fivefold higher than that of isoleucine. However, the proof-editing mechanism, which involves the rapid hydrolysis and removal of mis-incorporated valine-tRNA, makes the error rate decrease to 1 in 3000.[26] Occasionally, as is the case of D-tyrosine, proofreading is performed by another enzyme that hydrolyzes the misacylated tRNA.[28] Nonetheless, unlike the case of D-tyrosine, selectivity with respect to unnatural amino acids that have never been present in living cells is very low. For example, the unnatural analogue 4-hydroxy-phenylalanine is bound in the same mode as the native phenylalanine substrate.[29]

Many enzymes perform secondary tasks,[30,31] and some of the more illuminating examples include enzymes that have been under intense selection for high specificity, such as aminoacyl-tRNA synthetases. These enzymes exhibit secondary functions, for example, the biosynthesis of signaling molecules such as Ap4A (2 adenosines linked through 4 phosphates) is mediated by lysyl-tRNA synthetases.[32,33] In addition, certain aminoacyl-tRNA synthetases bind DNA or mRNA and thus regulate transcription, splicing, and translation, or act as cofactors in RNA trafficking.[34] It is probable that these functions were recruited well after the primary function had emerged (loading of a specific tRNA with the cognate amino acid), from promiscuous, or side products, of the primary reactions (see also Section 8.03.8.6.2 on gene sharing). Once recruited, such functions remained under selection, and therefore became a native function of the enzyme (Section 8.03.3.1).

High-throughput screens for binding cross-reactivities revealed a very clear pattern, whereby the number of identified cross-reactants (e.g., small molecules or proteins), and their affinity, increase exponentially with the number of tested ligands, or binding sites (e.g., number of different antibodies).[35–37] Several theoretical models account for these observations[38–40] (for further discussion of these aspects see Griffiths and Tawfik[41] and James and Tawfik[42]). A screen using a sufficiently large diversity of substrates (and reactions that can be performed on these substrates) is probable to reveal that every enzyme exhibits a pattern of promiscuous functions. Some of these may be related to the enzyme's native function, and others might have a weak relation. Unfortunately, contrary to binding, exhaustive screens for catalytic promiscuity are technically challenging, not the least because different detection methods need to be applied for different substrates and reactions.

Few, and often none, of the promiscuous activities found *in vitro* (or those that could be found by systematic screens) bear a physiological or evolutionary meaning. But even those that might, are not necessarily relevant *in vivo*. The main reason is regulation. Of the entire enzyme diversity available to organisms, only a small fraction is accessible, and active, at a given time and cellular location. In this way, many of the undesirable outcomes of enzyme promiscuity are prevented. Regulation at the level of expression is obviously directed to prevent the spending of unnecessary resources.[34,43,44] Different regulation regimes are the key to control enzyme activity, especially with enzymes whose specificity is broad. For example, *E. coli* has 23 different members of the HAD superfamily (haloacid dehalogenase (HAD)-like hydrolases). Most of these enzymes are phosphatases that show remarkably broad and overlapping substrate spectra.[45,46] However, these paralogs that are biochemically similar operate under different regulation schemes, and specificity is achieved through regulation and not by the more familiar manner of controlling enzyme reactivity.[46]

Regulation can also occur at the protein level, such as allosteric regulation that prevents the wasteful conversion of costly metabolites. Such regulation is expected to be, and in many cases is, product-controlled. However, in many cases, the substrate also comprises of an allosteric regulator of its own enzyme – in other words, in the absence of its substrate, the enzyme active site becomes inactive. Preventing the active sites from promiscuously reacting with other, undesirable substrates could be one of the driving forces for the evolution of such mechanism.

There exists evidence indicating that despite the action of natural selection to increase enzyme selectivity by various means, ranging from shaping the active site itself to regulation of enzyme expression and activity, numerous cross-reactions and breaches of specificity occur, not just *in vitro*, but in particular within living cells. Such cross-reactivities are often unraveled by the analysis of auxotrophic knockout strains that lack a crucial enzyme. These deficiencies are often complemented by other enzymes, or even other enzyme pathways,

sometimes in an unexpected manner. For example, knockouts of the *phn* operon in *E. coli* that is known to utilize phosphite (HPO_3^{2-}) led to the identification of promiscuous phosphite-dependent hydrogenase activity in alkaline phosphatase (see **Figure 2(a)** in Section 8.03.6.1.2). This activity enables the growth of the *phn* knockout strain with phosphite as the sole phosphorous source.[47] Other examples of 'metabolic plasticity', or 'underground metabolism' were reviewed by Jensen,[10,25] and later by D'Ari and Casadesus.[48]

The fact that cross-reactivities between different metabolic pathways are highly expected was also indicated by an *in silico* experiment that attempted to dock 125 common metabolites into the active sites of 120 key metabolic enzymes. Numerous potential cross-reactions were found amongst these 15 000 potential pairs, cross-reactions that were often stronger than the cognate interactions. Although docking has obvious limitations, this study further highlights the potential for promiscuity, and underground metabolism, within numerous metabolic pathways.[49] A systematic survey of complementation of deficient *E. coli* strains, by selection from a library of *E. coli*'s own genes under overexpression, revealed a similar picture.[50] The deleted gene and its suppressor were in most cases unrelated. Complementation was achieved through the promiscuous action of another enzyme, through increased transport (and not necessarily of the deficient metabolite), and most often, through the opening of an alternative metabolic pathway and/or more global changes through the overexpression of a regulatory factor. Thus, promiscuity is not necessarily a phenomenon limited to the single enzyme level, but often whole pathways can act promiscuously, namely, outside their routine functional scope.

The above observations and considerations led to new hypotheses which suggest that genetic and metabolic pathways are inherently probabilistic and ambiguous. By these hypotheses, the well-studied linear pathways described in textbooks can be cross-wired in a variety of unexpected ways. Evolution may capitalize on these unexpected cross-wirings, in a way of adaptive plasticity, to generate new metabolic capabilities.[51] Phenomena similar to 'underground metabolism' were also observed in genetic analyses where the observed phenotypes turned out to be correlated with changes in many different genes, including genes from unrelated pathways. These studies demonstrate a remarkable flexibility in the way genomes respond to changes. As is the case with enzymes and metabolic pathways, genome flexibility is the outcome of the limited specificity, or promiscuity, of gene action and of intergenic interactions.[52,53] Thus, it appears that, beyond the linear, well-defined pathways which are extensively studied, there exist 'flexible genomes',[52] as well as 'flexible proteomes', and 'flexible metabolomes', the contribution of which to evolutionary adaptation requires further study.

8.03.3 The Definitions of Promiscuity

The term 'enzyme promiscuity' was imprinted through a review by O'Brien and Herschlag[6] that highlighted a topic that, at the time, was relatively ignored by enzymologists. However, the term promiscuity is not very well defined and has been used to describe a wide range of fundamentally different phenomena that are not necessarily related. We thus propose the following terminology, or glossary:

8.03.3.1 Native Function

The substrate, and chemical transformation, for which an enzyme has evolved, and that is relevant to physiology of the organism in which this enzyme resides. By definition, the native, or primary function is maintained under selection, and mutations that harm it may affect organismal fitness. Many enzymes have more than one physiological function,[31] and in this case, all these functions should be defined as native, even though some of these were obviously recruited at much later evolutionary stage. The 'primary function' describes the function that underlines these enzymes. For example, the primary function of aminoacyl-tRNA synthetases is activating and loading amino acids onto the cognate tRNA, whereas the generation of Ap4A is clearly 'a secondary function' executed only in certain organisms and under specific circumstances.[32] The 'original, or existing, function', are complementary terms that refer to the native, or primary, function in the context of the divergence of new functions.

8.03.3.2   Multispecificity or Broad Specificity

Many enzymes evolved to perform a certain reaction on a whole range of similar substrates, rather than on a single substrate. Their broad specificity is therefore an inherent, evolved function, and they should be regarded as multispecific enzymes, and not as promiscuous enzymes. Enzymes known for their broad substrate specificity include mammalian detoxifying enzymes, such as glutathione *S*-transferases (GSTs) and cytochrome P450s,[54–58] certain enzymes in terpenoid biosynthesis such as sesquiterpene cyclases,[59,60] and methane monooxygenase that hydroxylates over 150 various substrates.[61] Other examples include exonucleases that, contrary to restriction endonucleases, cleave double-stranded, and often also single-stranded DNA, with almost no sequence specificity. In cases of multispecificity, the reactions with the various substrates are expected to exhibit similar kinetic parameters (K_M, k_{cat}).

8.03.3.3   Substrate and Cofactor Ambiguity

This definition applies to enzymes that evolved to transform one well-defined substrate. The enzyme's activity with other substrates is purely accidental, or promiscuous. However, in cases where the structure and chemical nature of the alternative promiscuous substrates resemble the native substrate, this breach of specificity is best defined as substrate ambiguity. Common examples of substrate ambiguity include lipases,[62,63] whose natural substrates are glyceryl esters of long-chain carboxylic acids (fatty acids), and that catalyze the hydrolysis of a broad range of esters. More recent examples include asparagine synthetase that can complement a deficiency in glutamine synthetase.[50] Another type of substrate ambiguity is when an enzyme uses various nucleophiles to react with the same substrate, as in case of halohydrin dehalogenase. Normally, this enzyme uses a water molecule to open an epoxide ring and form a diol product. However, a whole range of monovalent anions, such as Br^-, Cl^-, CN^-, and N_3^-, can be promiscuously applied by the enzyme to generate a broad range of products.[64] The related category of 'cofactor ambiguity' includes enzymes that can utilize coenzymes, or cofactors, other than the one they have evolved with. For example, D-2-hydroxyacid dehydrogenase from *H. mediterranei* can use both NADH and NADPH as cofactors.[65] Another example is regarding metallo-enzymes that have been shown to catalyze promiscuous reactions when the naturally occurring active-site metal is replaced (Section 8.03.6.1.5).

8.03.3.4   Promiscuity or Catalytic Promiscuity

This category refers to enzymes that catalyze different reactions (and not just different substrates) than the one they evolved for. As is the case with substrate and coenzyme ambiguity, the enzyme's activity with these alternative substrates is purely accidental, and is under no selection, and is therefore promiscuous by definition. As suggested,[8] these cases include chemical transformations where the bonds that are broken, or formed, are different than those in the native substrate and reaction, and/or transformations that proceed through a different transition state. As discussed later, the promiscuous chemical transformations can be performed by the same catalytic side chains, and by essentially the same mechanism, as the native enzymatic function (Section 8.03.6). But there are also cases in which the enzyme utilizes different subsets of active-site residues, and somewhat different mechanisms, for the native and promiscuous functions (Section 8.03.6.1.4).

8.03.3.5   Moonlighting

In contrast to promiscuity that occurs within the same active site as the primary, native function, moonlighting relates to the utilization of protein parts outside the active site for other functions, mostly regulatory and structural,[7,66] but sometimes enzymatic ones.[67,68] Such activities can be recruited at later evolutionary stages, as indicated by the classical example of crystallins whereby metabolic enzymes were recruited later in evolution as structural components of eye lenses[30,69] (see Section 8.03.8.6.2 on gene sharing).

8.03.4   Quantifying the Degree and Magnitude of Promiscuity

A more refined definition of promiscuity should include a quantitative measure for the degree and magnitude of promiscuity. The *degree of promiscuity* refers to the level to which enzyme specificity is breached (or the degree of multispecificity for enzymes that react with a broad range of different substrates). Namely, how different are the native and promiscuous functions. The *magnitude of promiscuity* refers to the kinetic parameters for the promiscuous activity relative to the native one.

8.03.4.1   The Degree of Promiscuity

Previous work suggested that the degree of promiscuity can be qualitatively assessed along two dimensions whereby, one dimension describes differences in the type of bonds that are being formed or broken, and the other dimension describes differences in the mechanism of the catalyzed reaction.[8] A more recent work aimed at a quantitative measure dubbed 'index of promiscuity' that is calculated by mapping the substrate structure and quantifying the degree of variability between different substrates of the same enzyme.[58] However, this method does not take into account substrate chemistry, and assumes that the same chemical transformation occurs on all substrates. As such, it is more suitable for the quantitative analysis of cross-reactivity with multispecific enzymes such as GSTs (as originally demonstrated), and largely inapplicable to quantify catalytic promiscuity whereby, as defined in Section 8.03.3.4, the major differences pertain to the chemistry and not the substrate structure.

8.03.4.2   Assessing the Degree of Promiscuity with EC Numbers

Here we propose a simple, and relatively objective, way of assessing the degree of promiscuity using a comparison of Enzyme Commission numbers (EC) for the native and promiscuous activities. In enzymes exhibiting multispecificity, or substrate ambiguity, EC numbers for the various substrates should be the same, or differ only by the 4th digit that generally distinguishes between enzymes of the same class. Catalytic promiscuity should refer to cases in which the EC numbers of the various substrates and reactions catalyzed by the same enzyme differ in the 2nd, or the 3rd, digits that refer to different chemistries, and different classes of substrates, or even by the 1st digit that indicates a completely different reaction category.

Table 1 lists several examples for substrate ambiguity and catalytic promiscuity. The table indicates the EC numbers for the native and promiscuous activities, and thus categorizes their promiscuity accordingly to differences in EC numbers. Typical examples where differences in EC numbers reflect the degree of promiscuity, and cases where they might not are discussed below.

Almost all cases of substrate ambiguity and multispecificity (as defined in Section 8.03.3) are manifested in differences in the 4th digit. Examples for multispecific, or broad-specificity enzymes include sulfotransferases and GSTs (Table 1, entries 1, 2).

Examples for substrate ambiguity include enzymes such as sugar kinases, amino acid transferases, glycosidases, and methyltransferases, which can perform the same chemical transformation on substrates other than their native one (Table 1, entries 3–6).

Lipases (EC 3.1.1.3; Table 1, entry 7) comprise a clear example where EC numbers seem to reflect differences in the degree of promiscuity for a whole range of promiscuous activities. Their native substrates are triglycerides (EC 3.1.1.3), and their ability to promiscuously hydrolyze aryl esters of various carboxylic acids is manifested by a difference in the 4th digit only (3.1.1.X) and thus is correctly defined as substrate ambiguity. The promiscuous amide hydrolysis concerns the cleavage of a different bond (C–N vs C–O) and is manifested by differences in the 2nd digit (EC 3.4.X.X). Lipases were also shown to promiscuously catalyze nonhydrolytic reactions such as aldol condensations and Michael additions,[78–80] and these belong to the 4th EC category (EC 4.X.X.X). In these cases, both the mechanisms and the bonds that are being formed or broken differ, and these differences are manifested in the 1st digit.

Alkaline phosphatase (EC 3.1.3.1.; Table 1, entry 8) also possesses a wide range of promiscuous activities. Some of them differ from the native phosphate monoesters hydrolysis in the 3rd digit (sulfatase,

Table 1 Examples for classifying the degree of promiscuity based on differences in EC numbers

	Enzyme	Native activity (EC number)	Promiscuous activity (EC number)	Promiscuity type	Reference(s)
1	Human cytosolic sulfo-transferases (hSULTs)	Broad-specificity enzymes, sulfonate group transfer from PAPS to the various substrates EC 2.8.2.X.	Sulfonate group transfer from PAPS to the various substrates EC 2.8.2.X.	Multispecificity	70
2	Glutathione transferases	Broad substrate specificity GST of A-class 2.5.1.18.	Glutathione coupling with various ligands EC 2.5.1.18.	Multispecificity	57
3	*N*-acetyl-D-mannosamine kinase (NanK)	Phosphorylation of *N*-acetyl-D-mannosamine EC 2.7.1.60	Phosphorylation of glucose EC 2.7.1.2.	Substrate ambiguity	71, 72
	Fructose kinase (YajK)	Phosphorylation of fructose EC 2.7.1.4	Phosphorylation of glucose EC 2.7.1.2.		
	Allose kinase (AlsK)	Phosphorylation of allose EC 2.7.1.55	Phosphorylation of glucose EC 2.7.1.2.		
4	Aspartate amino-transferase (AATase)	Transamination of dicarboxylic substrates EC 2.6.1.1.	Transamination of tyrosine and phenylalanine EC 2.6.1.X.	Substrate ambiguity)	73
5	Beta-glucuronidase (GUS)	Hydrolysis of beta-glucuronides EC 3.2.1.31	Hydrolysis of pNP-galactoside EC 3.2.1.23	Substrate ambiguity	74, 75
6	*HaeIII* methyltransferase	Methylation of GGCC sites EC 2.1.1.X.	Methylation of AGCC sites EC 2.1.1.X.	Substrate ambiguity	76
7	Lipases	Triglyceride hydrolysis EC 3.1.1.3.	Hydrolysis of aryl esters of various carboxylic acids 3.1.1.X.	Substrate ambiguity	62, 63
			Amide bond hydrolysis EC 3.4.X.X.	Catalytic promiscuity	77
			Aldol reaction (C–C bond formation) EC 4.1.X.X.	Catalytic promiscuity	78
			Michael-type additions EC 2.5.1.18. (as in GST) EC 4.4.X.X. (as in lyases)	Catalytic promiscuity	79, 80
			Oligomerization of siloxanes Unnatural reaction for which no native enzyme in known	Catalytic promiscuity	81
8	Alkaline phosphatase	Hydrolysis of monophosphate esters EC 3.1.3.1.	Phosphodiesterase EC 3.1.4.X.	Catalytic promiscuity	82
			Sulfate esters hydrolysis EC 3.1.6.X.	Catalytic promiscuity	83, 84
			Phosphite-dependent dehydrogenase EC 1.1/2.X.X	Catalytic promiscuity	47
9	Muconate lactonizing enzyme (MLE)	Cycloisomeri-zation EC 5.5.1.1.	OSBS (β-elimination) EC 4.2.1.113	Catalytic promiscuity	85

(Continued)

Table 1 (Continued)

	Enzyme	Native activity (EC number)	Promiscuous activity (EC number)	Promiscuity type	Reference(s)
10	Phosphotri-esterase from *P. diminuta* (PTE)	Phosphotriester hydrolysis EC 3.1.8.1.	Phosphodiesterase EC 3.1.4.X. Esterase EC 3.1.1.X. Lactonase EC 3.1.1.X.	Catalytic promiscuity	19, 86, 87
11	PLLs (PTE-like lactonases)	Hydrolysis of quorum sensing lactones EC 3.1.1.X.	Phosphotriester hydrolysis EC 3.1.8.1.	Catalytic promiscuity	88
12	*AiiA* from *B. thuringiensis*	Hydrolysis of quorum sensing lactones EC 3.1.1.X.	Phosphotriester hydrolysis EC 3.1.8.1	Catalytic promiscuity	88, 89; H.-S. Kim, personal communication
13	PONs (serum paraoxonases)	Mammalian lactonases EC 3.1.1.X.	Phosphotriesterase EC 3.1.8.X.	Catalytic promiscuity	90–92
			Aryl esterase EC 3.1.1.X.	Substrate ambiguity	
14	Serum albumins	Nonenzymatic proteins	Esterase EC 3.1.1.X.	Catalytic promiscuity	93
			Carbamate hydrolysis EC 3.1.1.X.	Catalytic promiscuity	94
			Kemp elimination (unnatural reaction, for which no native enzyme in known).	Catalytic promiscuity	95
15	Carbonic anhydrase	Hydration of CO_2 EC 4.2.1.1.	Esterase EC 3.1.1.X.	Substrate ambiguity	2, 96
			Epoxide synthase (styrene epoxidation, by metal exchange) EC 1.14.X.X.	Catalytic promiscuity, or cofactor ambiguity	97

phosphodiesterase), but others (phosphite dehydrogenation) differ in the 1st digit, and represent a higher degree of promiscuity.

Other cases in which the native and promiscuous activities differ in the 1st digit include muconate lactonizing enzyme (MLE; **Table 1**, entry 9), whose native activity is cycloisomerization (EC 5.5.1.1) and possesses a promiscuous OSBS (β-elimination) activity (EC 4.2.1.113).

Examples for a high degree of promiscuity obviously include nonenzymatic proteins such as serum albumins that exhibit promiscuous catalytic activities[93–95] (**Table 1**, entry 14). Other cases may include catalysis of unnatural reactions, meaning reactions for which, to our knowledge, no natural enzyme has evolved (e.g., the Kemp elimination performed by serum albumin (**Table 1**, entry 14, or siloxane hydrolysis by lipase (**Table 1**, entry 7)).

Another example where EC numbers seem to reflect differences in the degree of promiscuity concerns lactonases. These enzymes that belong to three different superfamilies have all been shown to exhibit promiscuous phosphotriesterase (PTE) activity (**Table 1**, entries 11–13). Lactonase activity involves the hydrolytic cleavage of a C–O bond and is described as EC 3.1.1.X (where X refers to a specific lactone substrate). The hydrolytic cleavage of the P–O bond of phosphotriesters is described by EC number 3.1.8.X. The difference in the 3rd digit reflects the difference in the bond that is being broken (C–O versus P–O), while applying essentially the same mechanism. For comparison, many of these lactonases (but not all of them) also exhibit esterase activity with aryl esters in particular.[88] This activity is described as EC 3.1.1.X, as is the case with lactonase activity. Because both reactions involve cleavage of C–O bonds, and the lactonase–esterase differences are manifested in the 4th digit, this case is better described as substrate ambiguity.

It should be noted, however, that EC numbers can also be misleading. There are notable cases where, despite considerable similarity in the chemistry of catalysis, the EC numbers differ, and even by the 1st digit. This is primarily because the EC definitions relate not only to the chemistry, but also to the physiological context. A clear example is carbonic anhydrase (**Table 1**, entry 15; EC 4.2.1.1) that exhibits promiscuous aryl esterase activity (3.1.1.X). The EC numbers suggest a totally different chemistry. But although the substrates differ significantly, primarily in size, both reactions involve the attack of a hydroxide ion on a carbonyl (**Figure 1**). The phosphite-dependent hydrogenase activity (EC 1.1/2.X.X) of alkaline phosphatase (**Table 1**, entry 8; EC 3.1.3.1) also represents a borderline case. The very different native and promiscuous activities (as manifested in the different EC categories) actually utilize a similar mechanism (see Section 8.03.6.1.2).
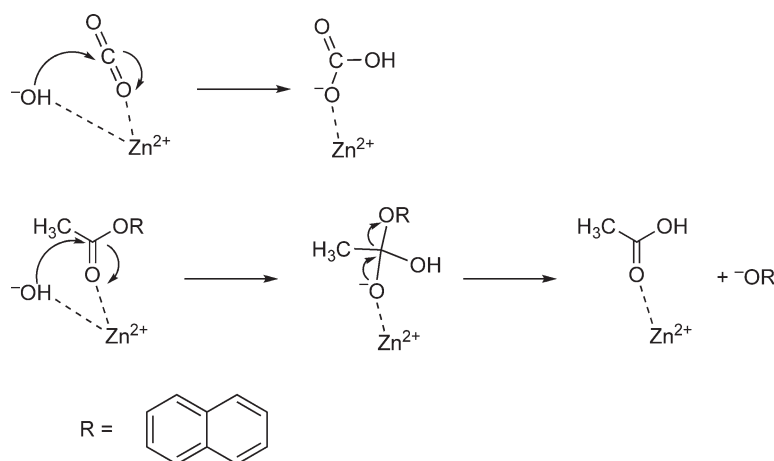
Figure 1 The native reaction of carbonic anhydrase (CO_2 hydration, *top*) and its promiscuous aryl esterase reaction, exemplified with naphthyl acetate (*bottom*). Both reactions proceed by the same mechanism of hydroxide ion attack on a carbonyl, followed by the stabilization of an oxyanion intermediate by the active-site Zn^{2+}. Despite this obvious similarity, the EC numbers of these reactions differ in the first digit (**Table 1**, entry 15).

8.03.4.3   The Magnitude of Promiscuity

Another quantitative measure addresses the magnitude of promiscuous activities, that is, how do the kinetic parameters for the promiscuous reaction/substrate compare with those for the native substrate. Whereas most enzymes exhibit k_{cat}/K_M values in the order of 10^5–$10^8 \, mol^{-1} \, l \, s^{-1}$ for their native substrates,[26] the magnitude of promiscuous activities varies over more orders-of-magnitude, both in absolute terms, and relative to the native activity. In many cases, the promiscuous activities are relatively high and fall within just an order-of-magnitude, or two, from the native function (e.g., **Table 2**, entries 1, 2, 3, 5, 6, 9). Such activities can obviously provide a distinct and immediate selective advantage. These cases may resemble 'generalist' intermediates (see Section 8.03.8.3) and their divergence may proceed through a 'gene sharing' mechanism, as described in Sections 8.03.8.6.2 and 8.03.8.6.3. In many cases, however, the promiscuous activities are very low, or even barely detectable. Weak promiscuous functions can still provide a selective advantage, certainly under high expression levels.[47,98,99] Gene duplication is another way by which enzyme levels could be increased, and thus endow a weak promiscuous activity a selective advantage (Section 8.03.8.6.4).

Furthermore, a useful way of assessing the magnitude of promiscuous activities is the rate acceleration (k_{cat}/k_{uncat}) or catalytic proficiency ($k_{cat}/K_M/k_{uncat}$). These parameters are indicative because they take into account the inherent reactivity of the substrate.[6] In many cases, promiscuous activities occur, or are measured, with highly reactive substrates. Such activities are in a way expected.[100] However, there are many cases in which promiscuous activities take place with substrates that are less activated than the native one. Examples include, the amidase activity of esterases such as lipases (**Table 1**, entry 7), the phosphodiesterase activities of *P. diminuta* PTE and alkaline phosphatase (**Table 1**, entries 10 and 8), and the PTE activities or various lactonases (**Table 1**, entries 11–13; and the notable fact that some of these lactonases do not hydrolyze the more activated aryl esters). In such cases, the chemical challenge posed by a less activated substrate is reflected in the more favorable comparisons of rate accelerations, or catalytic proficiencies, for the native versus the promiscuous substrates.

8.03.5   Predicting Promiscuity

The growing interest in promiscuity, from both a fundamental point of view, and an applicative one (applications of enzymes in organic synthesis), provides an incentive for the development of computational and bioinformatic tools for its prediction. By default, promiscuity is a phenomenon which is unpredictable – and the more interesting and unrelated is the promiscuous function, the harder would be its prediction. The difficulty of prediction is further augmented by the fact that in many cases, promiscuity involves some degree of structural plasticity, namely, when the native and promiscuous functions are mediated by different active-site conformations (see Section 8.03.6.1.1).

Early attempts to assess the frequency and potential for promiscuous enzyme–substrate encounters were made by docking a set of substrates into a set of enzymes and examining the number and distribution of the resulting virtual substrate–enzyme matches.[49] This study, however, aimed at evaluating the potential for cross-reactivities and had intention, or computational accuracy, only to identify individual cases of promiscuity. Significant improvements in structural predictions and ligand docking, and in particular, successful attempts to dock transition states rather than substrates,[101] may enable such studies to be performed with much higher accuracy and scope.

The hope that promiscuity is predictable is also supported by the identification of systematic patterns of promiscuity. For example, lactonases, and in particular lactonases that favor hydrophobic lactones, show a consistent tendency to promiscuously catalyze the hydrolysis of phosphotriesters. This pattern has now been seen in lactonases from three different superfamilies:[76] PLLs (TIM-barrels from the amidohydrolase super-family; **Table 1**, entry 11); PONs, or serum paraoxonases (calcium-dependent six-bladded β-propellers; **Table 1**, entry 13); and *AiiA* (a lactonase from the metallo-β-lactamase superfamily; **Table 1**, entry 12). That very different scaffolds and active-sites configurations share the same promiscuity pattern suggests that these reactions share a key feature, probably in the geometry of their transition states. This feature must be distinct, also because many of these lactonases do not hydrolyze esters that are much closer to lactones than phosphotriesters, and should thus be amenable to structural analysis and prediction.

The ongoing enrichment of structure and sequence databases, and the development of novel computational and bioinformatic tools, should also facilitate the prediction of promiscuous functions. In particular, as discussed in Section 8.03.7, the observation that within highly diverse superfamilies, the native function of one enzyme family often comprises of a promiscuous function of another family, and vice versa, also suggests that certain promiscuous functions are predictable.

8.03.6   Mechanistic Aspects of Promiscuity

8.03.6.1   How Do Specificity and Promiscuity Coincide within the Same Active Site?

A frequently asked question is how the very same active-site and catalytic machinery can show exquisite specificity with respect to the native substrate (and thus avoid catalyzing other, closely related substrates), but still catalyze other, and often completely unrelated, functions in a promiscuous manner. The answer to this question is somewhat complex, because different scenarios, or mechanisms, seem to account for the coexistence of specificity and promiscuity within the very same active site. Several scenarios are outlined in the next section, and typical examples are described.

8.03.6.1.1   *Conformational diversity*

Despite the rigid, lock-and-key images of active sites obtained by X-ray crystallography, active sites, and even their protein scaffolds, are remarkably flexible. The role of structural plasticity in facilitating enzyme action, and evolution, has been discussed in several reviews.[102–104] In many cases promiscuity is linked to conformational diversity, whereby the native and the promiscuous functions are mediated by different active-site configurations.

A notable example of the role of conformational plasiticity comes from α-lytic protease, where a single amino acid substitution increased the activity toward promiscuous substrates by a factor of 10^5, whereas the native activity was reduced by only twofold.[105] This large shift in the selectivity of this enzyme and its other family members is allowed by the structural flexibility of the substrate binding loops.[106,107] The mobility of active-site loops was also demonstrated to a play a key role in mediating promiscuity in isopropylmalate isomerase, an enzyme with dual substrate specificity, where a loop structure is dependent on the substrate presence.[108] Other examples include human sulfotransferase SULT1A1, where conformational changes enable the same enzyme to accommodate different substrates.[102] The importance of conformational plasticity in mediating functional promiscuity was also demonstrated in a study about glutathione-S-transferase isoforms GSTA1-1 and GSTA4-4.[57] In an evolved aminoacyl-tRNA synthetase, the disruption of an α-helix, introduced structural plasticity to the enzyme's active site and thus enabled it to accept a relatively broad range of unnatural amino acid substrates.[109]

It should be noted that cases of different active-site configurations can include substantial backbone rearrangements, in some cases, and only different side-chain rotamers in others. In principle, however, in all these cases promiscuity makes use of a structurally modified active site. That alternative active-site conformations can mediate alternative promiscuous functions also has interesting implications regarding the evolvability of promiscuous functions.[103,104] (see Section 8.03.8.1).

8.03.6.1.2   *Accommodating alternative substrates*

In many cases, promiscuous activities share the main active-site features (and often the same active-site configuration) with the native activity, and besides differences such as substrate positioning, their mechanism is largely the same. For example, in the family of guanidine-transferring enzymes, three mutually inter-promiscuous enzymes PaADI, PaAgDI, and PaDDAH utilize the same catalytic triad (Cys-His-Asp), in their action on various derivatives of arginine.[110] However, in this case of substrate ambiguity, the active-site residues that bind the $C\alpha$-carboxyl and the guanidino-NH_2 of these different substrates are different.

Other examples include cases in which the enzyme applies nucleophilic catalysis, and the same active-site nucleophile is utilized in both the promiscuous and the native function. For example, alkaline phosphatase is a highly proficient ($k_{cat}/K_M > 10^7 \, mol^{-1} \, 1 \, s^{-1}$) phosphate monoesterase that promiscuously hydrolyzes phosphodiesters, phosphoamides, and sulfate esters,[82–84] as well as phosphite (while actually reducing water to release hydrogen; **Figure 2(a)**).[47] The catalytic mechanism is presumed to be similar for all these reactions, and involves nucleophilic attack by Ser102, and stabilization of the negatively charged intermediate by the
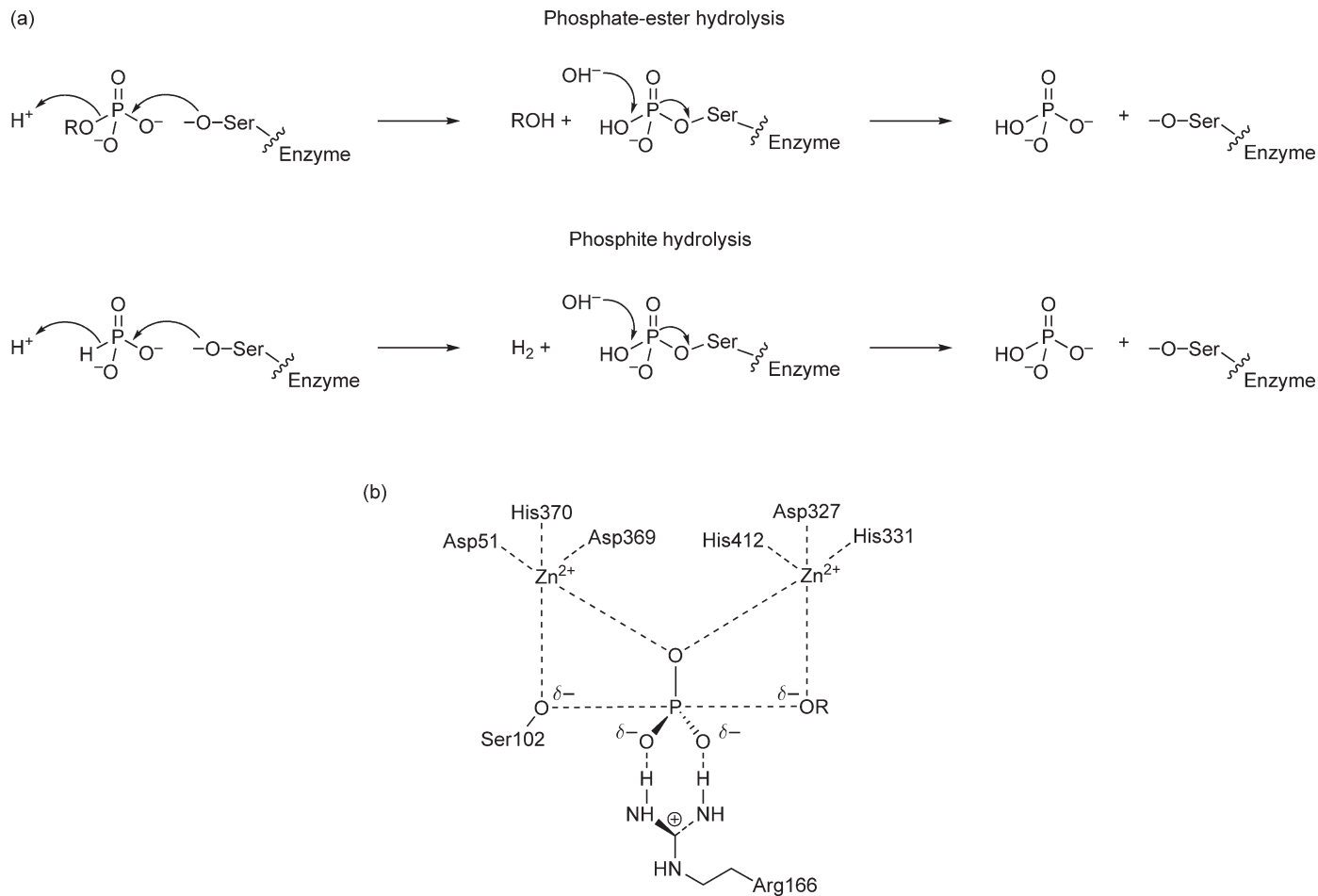
Figure 2 (a) The native monoester phosphatase activity, and the promiscuous phosphite oxidation reactions catalyzed by alkaline phosphatase. Adapted from K. Yang; W. W. Metcalf, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7919–7924. (b) The active site arrangement of alkaline phosphatase with a bound transition state model. Adapted from I. Catrina; P. J. O'Brien; J. Purcell; I. Nikolic-Hughes; J. G. Zalatan; A. C. Hengge; D. Herschlag, *J. Am. Chem. Soc.* **2007**, *129*, 5760–5765.

active-site Zn^{2+} ions and Arg166[47,82] (**Figure 2(b)**). Comparison between the phosphate monoesterase, phosphodiesterase, and sulfatase activities of alkaline phosphatase revealed that, whereas these substrates all bind in a similar mode, the interactions with both Zn^{2+} ions and Arg166 are much more favorable for the native phosphate monoester substrates than for other, promiscuous substrates.[82,83] This difference accounts for the orders-of-magnitude higher rates and catalytic proficiencies of the native substrates versus the promiscuous ones.

There are also cases where the network of hydrogen bonds is the main feature that differs the native reaction from the promiscuous one. D-2-keto-3-deoxy-gluconate aldolase (KDGA) from hyperthermophilic *Sulfolobus solfataricus* reacts with both gluconate and galactonate substrates with similar rates.[111] The mechanism with both substrates involves Schiff base formation by Lys155, and subsequent hydration and cleavage. The differences between gluconate and galactonate are in the hydrogen bonds formed with KDGA's active site, and in particular in the manner by which the 5′ and 6′ hydroxyl groups are bound (**Figure 3**).[112]

It therefore seems that the very same active site can offer numerous modes of interactions, and some of these might be utilized by promiscuous substrates. It should be noted, however, that most of the above describes cases analyzed by kinetics and site-directed mutagenesis. Very few structures of the enzyme–substrate, or enzyme transition-state complexes, exist for both the native and promiscuous substrates. And thus, small, or even significant, changes in active-site configuration cannot be excluded in the described cases.

8.03.6.1.3   *Different protonation states*

There are cases in which the same active-site residue acts in two different protonation states in the native compared to the promiscuous function. In the tautomerase superfamily, various enzymes share the catalytic Pro residue at the enzyme N-terminus, but the mechanism of catalysis depends on its pK_a. In 4-oxalocrotonate tautomerase (4-OT) that catalyzes the 1,5-keto–enol tautomerization of 2-hydroxymuconate to 2-oxo-3-hexendioate, the pK_a of Pro1 is $\sim$6.4, and it acts as a general base. In another tautomerase family member, CaaD, which catalyzes the hydrolytic halogenation of chloro- and bromoacrylates, Pro1 is protonated ($pK_a \sim 9.2$) and serves as a general acid.[113–115] Since in 4-OT little proportion of Pro1 is present in the correct protonation state for general acid catalysis, its promiscuous hydratase activity is quite low ($2.6 \times 10^{-2} \, mol^{-1} l \, s^{-1}$). The knowledge of 4-OT and CaaD catalytic mechanisms was used to elucidate the catalytic mechanism of yet another member of tautomerase family, malonate semialdehyde decarboxylase (MSAD). MSAD has a substantial hydratase activity (with even higher turnover number than CaaD, $5.8 \, s^{-1}$ vs. $0.7 \, s^{-1}$), and it was proposed that its Pro1 is protonated and serves as a general acid also in the mechanism of the native MSAD activity.[116,117]
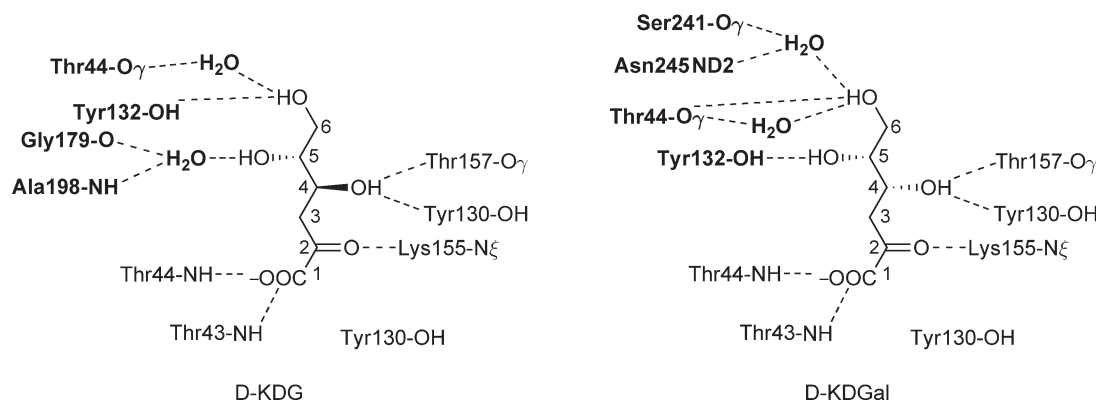
Figure 3 Schematic summary of the different interactions made in the active site of D-2-keto-3-deoxy-gluconate aldolase (KDGA) with its two substrates: D-2-keto-3-deoxy-gluconate (D-KDG) and D-2-keto-3-deoxy-galactonate (D-KDGal). Adapted from A. Theodossis; H. Walden; E. J. Westwick; H. Connaris; H. J. Lamble; D. W. Hough; M. J. Danson; G. L. Taylor, *J. Biol. Chem.* **2004**, *279*, 43886–43892.

8.03.6.1.4   Different subsites within the same active site

In probably fewer cases, although both the original and promiscuous activities reside in the same active site, and rely on its major feature (e.g., an oxyanion hole), other parts of the catalytic machinery may differ significantly. One such example is serum paraoxonase (PON1), a mammalian lactonase with promiscuous esterase and PTE activities. All these activities depend on a calcium ion that serves as PON1's 'oxyanion hole', but the general base, which activates a water molecule, differs[118] (**Figure 4(a)**). The hydrolysis of lactones and esters is mediated by a His115–His134 dyad, that deprotonates a water molecule to generate the attacking hydroxide. The promiscuous PTE activity appears to be mediated by another set of residues. Indeed, mutations of both histidine residues increase the promiscuous phosphotriesterase activity, and even shift selectivity is some cases. For example, the His115Trp mutation dramatically shifts the selectivity in favor of P–S versus P–O bonds.[120,121] Coordination of the phosphoryl oxygen to the active-site calcium is one feature that is shared with the lactonase mechanism. However, the His115–His134 has no role as a base in the hydrolysis of PTEs. It has been suggested that the PTE mechanism of PON1 is analogous to the mechanism of the squid diisopropyl fluorophosphatase (DFPase),[119] and involves a nucleophilic attack of Asp269 that comprises one of the Ca^{2+}-ligating residues.[119,122] However, this mechanism remains to be established by manifestations of a phosphoryl-enzyme intermediate, such as burst kinetics (that have not been observed with any of PON1's substrates), and the isolation of a phosphorylated enzyme species. Nonetheless, the mutagenesis data clearly indicate that key features of the lactonase and PTE mechanisms differ.

An analogous example is *Candida antarctica* lipase B (CALB) whose native activity (lipids hydrolysis) is mediated by a Ser105-His224-Asp187 catalytic triad. Using its oxyanion hole, formed by Gln106 and Thr40, CALB also catalyzes various carbon–carbon bond formation reactions, such as Michael additions and aldol condensations, with various ketone and aldehyde substrates.[78–80] However, in these reactions, the nucleophilic serine – the key part of the catalytic triad, plays no role, and the acid–base transfer is thought to be mediated by His224 in conjunction with Asp187 (**Figure 4(b)**). Indeed, as in PON1, the Ser105Ala mutant exhibits higher promiscuous activities than wild-type (WT) CalB.
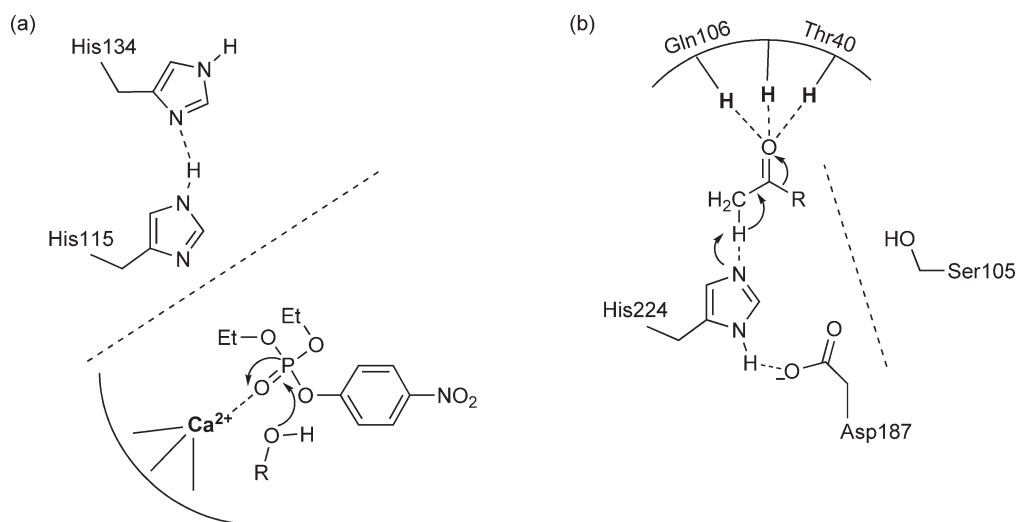
Figure 4   Different subsites within the same active site. (a) The main active site feature of the serum paraoxonase PON1 is the catalytic calcium ion, which lies at the bottom of a deep and hydrophobic active site, and is thought to act as the 'oxyanion hole' of PONs. The native function, hydrolysis of lactones, is mediated by a His115–His134 dyad, that deprotonates a water molecule to generate the attacking hydroxide. Although the same dyad appears to mediate the promiscuous arylesterase activity of PON1, the promiscuous phosphotriesterase activity (shown here for paraoxon) appears to be independent, and mediated by other residues that act as a base, or nucleophile.[118,119] Indeed, mutations of both His residues may increase the promiscuous phosphotriesterase activity by >300-fold with certain organophosphate substrates.[120,121] (b) A similar scenario has been described for the lipase CalB. Its native activity (lipid hydrolysis) is mediated by the Ser105–His224–Asp187 triad and the negative charge of the transition states, and the acyl-enzyme intermediate is stabilized by its 'oxyanion hole'. CalB also catalyzes promiscuous C–C bond formation reactions. In these promiscuous activities, the oxyanion hole is also utilized for negative charge stabilization (shown here). However, the catalytic serine takes no part, and acid–base transfer is thought to be mediated by His224 in conjunction with Asp187.[78–80]

8.03.6.1.5   Promiscuity due to alternative cofactors

Changes in chemical selectivity can also be induced by metal substitution. Such changes can be defined as *cofactor ambiguity*. Following the pioneering work by Kaiser,[123] the introduction of copper ions (Cu^{2+}) has been shown to introduce promiscuous oxidase activities in several hydrolytic enzymes.[124,125] In carbonic anhydrase, substitution of the Zn^{2+} by Mn^{2+} conferred the enzyme with enantioselective epoxidation of styrene.[97] Incorporating selenocysteine into the active sites of subtilisin,[126] glyceraldehyde-3-phosphate dehydrogenase GADPH,[127] and GST,[128] provided these enzymes with novel peroxidase activities.

8.03.6.1.6   Water-assisted promiscuity

Although the native substrate may interact directly with active-site residues, it is anticipated that water molecules probably play an important role in the indirect mediation of promiscuous interactions. Water molecules can bridge between a promiscuous substrate and active-site groups thus forming weak promiscuous hydrogen bonds. They can partially neutralize opposing dipoles and charges between the substrate and active-site residues, or act as acid, base, or nucleophile, in catalyzing the promiscuous reaction. Indeed, spatially defined active-site water molecules have catalytic power that is not fundamentally inferior to amino acid residues. Such water molecules may have played a key role in primordial enzymatic active sites,[26] and therefore probably participate in promiscuous activities. The current evidence for water-mediated promiscuity is slim, primarily because of the few structures of enzymes complexed with promiscuous substrates. However, the *Bacillus subtilis* esterase, which is also capable of amide hydrolysis, comprises an example whereby water molecules may assist the catalysis of a promiscuous substrate. A molecular dynamics study suggested that amide hydrolysis is affected by a network of hydrogen bonds consisting of water molecules. The esterase reaction is not influenced by these hydrogen bonds, due to the fact that esters lack the N-H amide group.[129]

8.03.6.2   Deciphering Enzyme Mechanisms by Studying Promiscuous Functions

Enzymologists have discovered that a systematic research of the 'hidden skills' of enzymes can provide valuable insights regarding their catalytic mechanisms (see also Section 8.03.6.1.3). For example, the promiscuous hydrolysis of phosphonate diester by *Tetrahymena thermophila* ribozyme provided key insights regarding the relative importance of transition state geometry versus charge.[130] The native activity of this ribozyme is the phosphodiesters hydrolysis, which differs from the promiscuous aminoacyl esterase activity by both the geometry and the charge of the transition state. Phosphonate diester was hydrolyzed by the ribozyme with similar turnover and analogous mechanism to the phosphodiesters hydrolysis, thus demonstrating that the geometry of the transition state plays a more important role than its charge.

In another study, the promiscuous chorismate mutase activity of isochorismate pyruvate-lyase (PchB) was used to derive mechanistic insights into its native activity (isochorismate pyruvate lyase).[131] Presumed key active-site residues were randomized, and the resulting variants of PchB were selected for the promiscuous chorismate mutase activity. Consequently, a common mechanism was proposed for both functions of PchB, with the rare [1,5]-sigmatropic rearrangement for the lyase activity, being distinct from other pyruvate lyases.

8.03.6.3   Mechanistic Origins of Differences in the Catalytic Parameters for Native Versus Promiscuous Functions

As noted above, the magnitude of promiscuous activities varies over many orders-of-magnitude, both in absolute terms, and relative to the native activity (Section 8.03.4.3). The differences in reactivity between the native and promiscuous substrates can be manifested in differences in both k_{cat} and K_M. The schematic view is that the energetics of substrate binding are reflected in the K_M, and catalysis by k_{cat}. It is therefore expected that promiscuous substrates that bind poorly, due to steric hindrance, for example, will exhibit high K_M values. However, many promiscuous substrates are characterized by low k_{cat} values. For example, a systematic analysis of >50 substrates for the enzyme PON1, the primary function of which is lipophilic lactonase, indicated that the promiscuous aryl ester, and phosphotriester, substrates all exhibit K_M values in the millimolar range $(0.8–5 \text{ mmol l}^{-1})$.[92] This is despite the fact that their k_{cat}/K_M values vary over three orders of magnitude. The differences in reactivity are therefore primarily due to k_{cat} values that vary by >1000-fold. The probable

reason is that for the promiscuous aryl ester and phosphotriester substrates, substrate binding is driven primarily by nonspecific hydrophobic forces with the deep and hydrophobic active site of PON1. It appears that, in many cases, promiscuous substrates are inadequately positioned relative to the catalytic machinery, and therefore exhibit very low k_{cat} values. Interestingly, for the lactones that comprise that native substrate of this enzyme, K_M values vary by $\sim$200-fold (from about 0.1 up to 20 mmol l^{-1}), whereas the variations in k_{cat} values are orders of magnitude lower ($\sim$10–200 s^{-1}).

It is therefore anticipated, and often observed, that the mode of binding of the native substrate – that is typically mediated by several independent, enthalpy-driven interactions – is fundamentally different from that of the promiscuous substrates where hydrophobic and other entropy-driven interactions play a key role.

8.03.7   Promiscuity and the Divergence of Enzyme Superfamilies

Enzyme superfamilies include numerous enzymes that although distant in sequence, share the same fold and the same catalytic mechanism. Members of such diverse superfamilies catalyze different chemical transformations of many different substrates, but share a common motive of catalysis.[11] Analysis of enzyme families and superfamilies provides the most solid and convincing body of evidence for the role of promiscuity in the evolution of new functions. Specifically, the identification of promiscuous activities, or cross-reactivities, between different members of the same enzyme family or superfamily, and the directed evolution of these activities, provide important hints regarding evolutionary, structural, and mechanistic relationships within enzyme families (**Figure 5**).

Several examples of the promiscuous catalytic activities within enzyme families and superfamilies are listed in **Table 2**. Key observations based on these data are summarized below:
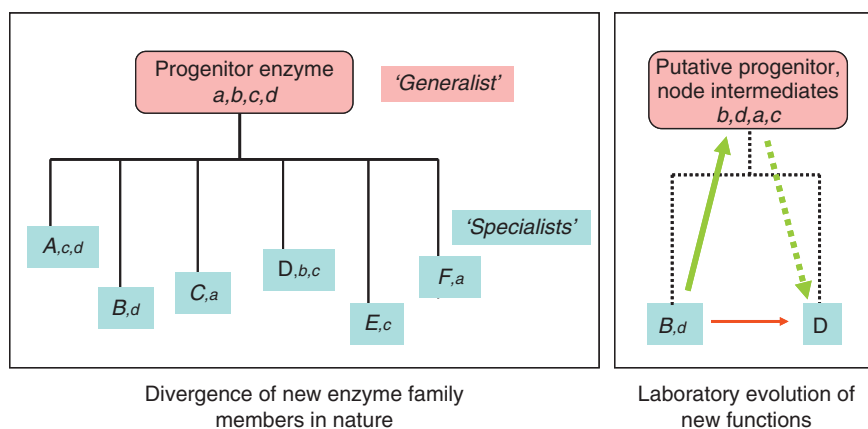
Figure 5   Experimental evidence favors the model of divergence of a 'generalist' progenitor enzyme to a family of 'specialist' enzymes. *Left panel*: Jensen's hypothesis[25] surmises that, in nature, an ancestor protein displaying a low level of a range of activities (denoted as *a*, *b*, *c*, *d*) had been subjected to selection pressures for those activities, thus duplicating and diverging into a family of potent and highly specialized enzymes of the kind seen today (denoted A, B, etc.). Today's 'specialists' may still retain some of the functions of the common ancestor (denoted in lower case), as low levels of promiscuous activities. Indeed, several reports indicate a low level of shared activities within a family, and in particular that the native activity of one member is the promiscuous activity of another, and vice versa (Table 2). *Right panel*: Additional support to the above model comes from the results of many directed evolution experiments. Direct switches of specificity, for example, from B to D (red arrow) are rare, and are typically seen following a parallel selection for an increase in the target activity and elimination of the original one. Upon mutation and selection for an increase of a promiscuous activity (green arrow), the resulting variants usually show significant increases in the target activity, and a smaller decrease in the original one, thus yielding, in effect, a 'generalist' intermediate exhibiting both *d* and *b* at relatively high levels (the 'weak negative trade-off' line in **Figure 6**). Such intermediates are often observed in the laboratory; some even gain other activities, never selected for (denoted *a*, *c*), and may therefore resemble the progenitor of this enzyme family, or node intermediates along past routes of its divergence.

Table 2 Examples for promiscuous activities within enzyme families and superfamilies

	Family/ Superfamily	Enzymes	Native activity; substrate (k_{cat}/K_M in $mol^{-1} l s^{-1}$)	Promiscuous activity; substrate (k_{cat}/K_M in $mol^{-1} l s^{-1}$)	Reference(s)
1	Mammalian Paraoxonases (PONs)	PON1 (serum paraoxonase) PON2 PON3	Lipo-lactonase – aliphatic 5-, 6-membered ring lactones with lipophylic side chains (γ-dodecanoic lactone, 1.2×10^5)	PON1: Aryl esterase (phenyl acetate, $\sim 6 \times 10^5$) Phosphotriesterase (paraoxon, 6×10^3) PON2: barely detectable aryl esterase; no phosphotriesterase PON3: low aryl esterase; barely detectable phosphotriesterase	91, 92
2	Tautomerase superfamily	Malonate semialdehyde decarboxylase (MSAD)	Decarboxylation of malonate semialdehyde (2.2×10^7)	Hydration of 2-oxo-3-pentynoate (6×10^2)	113–117
		4-oxalocrotonate tautomerase (4-OT)	Tautomerization of 2-hydroxymuconate to 2-oxo-3-hexenoate ($>10^5$)	Hydration of 3E-chloroacrylate (CaaD activity) (2.6×10^{-2})	
		YwhB tautomerase (4-OT analogue)	Tautomerization of 2-hydroxymuconate to 2-oxo-3-hexenoate (~ 105) (1.3×10^5)	Hydration of 3E-chloroacrylate (CaaD activity) (4.4×10^{-2})	
		trans-3-chloroacrylic acid dehalogenase (CaaD)	Hydrolytic dehalogenation (hydratase) of 3E-haloacrylates (1.2×10^5, 3E-chloroacrylate)	Hydration of 2-oxo-3-pentynoate (6.4×10^3)	
3	ROK family (repressor, open reading frame, kinase)	NanK	N-acetyl-D-mannosamine kinase (2.7×10^5)	Glucose kinase (5.1×10^2)	71, 72
		YajF	Fructose kinase (1.1×10^4)	Glucose kinase (2×10^2)	
		YcfX	Unknown	Glucose kinase (2.4×10^3)	
		AlsK	Allose kinase, (6.5×10^4)	Glucose kinase (15)	
4	Enolase superfamily: MLE (muconate lactonizing enzyme) subgroup	o-succinylbenzoate synthase (OSBS)	Dehydration of SHCHC (2-succinyl-6R-hydroxy-2,4-cyclohexadiene-1R-carboxylate) (2.5×10^5)	N-acylaminoacid racemase (NAAAR) reaction with N-acetyl methionine isomers ($4.9-5.9 \times 10^2$)	13, 132

(*Continued*)

Table 2 (Continued)

	Family/ Superfamily	Enzymes	Native activity; substrate (k_{cat}/K_M in mol^{-1}l s^{-1})	Promiscuous activity; substrate (k_{cat}/K_M in mol^{-1}l s^{-1})	Reference(s)
5	Amidohydrolase superfamily	Phosphotriesterase from P. diminuta (PTE)	Phosphotriesterase (paraoxon, 4×10^7)	Aryl esterase (2-naphthyl acetate, 500); lactonase (dihydrocoumarin, 6.5×10^5)	88, 87
		Phosphotriesterase homology protein (PHP)	Unknown	Aryl esterase (2-naphthyl acetate, 70)	
		Dihydroorotase (DHO)	Dihydroorotic acid hydrolysis (1.2×10^6)	Phosphotriesterase (paraoxon, 2.8)	
		AhlA; a member of the PLL family (PTE-like lactonases)	Lactonase; N-3-oxooctanoyl L-homoserine lactone (0.7×10^6)	Phosphotriesterase (paraoxon, 0.5)	
		PPH; a member of the PLL family (PTE-like lactonases)	Lactonase; N-3-oxooctanoyl L-homoserine lactone (0.55×10^5)	Phosphotriesterase (paraoxon, 8.6)	
		PTE-like lactonase SsoPox	Lactonase; N-3-oxooctanoyl L-homoserine lactone ($>10^6$)	Aryl esterase (naphthyl acetate, 400); phosphotriesterase (paraoxon, 4000)	
6	Orotidine 5′ monophosphate decarboxylase suprafamily (OMPDC)	3′ keto L-gluconate 6-phosphate decarboxylase (KGPDS)	Decarboxylation of 3′ keto L-gluconate 6-phosphate (7.7×10^4)	Aldol condensation of D-ribulose 5-phosphate and formaldehyde (HPS activity, 8.2×10^{-2})	133
		D-arabino-hex-3-ulose 6-phosphate synthase (HPS)	Aldol condensation of D-ribulose 5-phosphate and formaldehyde (1.6×10^4)	Decarboxylation of 3′ keto L-gluconate 6-phosphate (KGPDS activity, 2.3×10^3)	
7	N-acetyl-neuraminate lyase (NAL) family, pyruvate-dependent aldolases	N-acetyl-neuraminate lyase (NAL)	Cleavage of N-acetyl-neuraminate (3.1×10^3)	Aldol condensation of pyruvate and L-aspartate – β-semialdehyde (DHDPS activity) (20)	134
		Dihydrodi-picolinate synthase (DHDPS)	Aldol condensation of pyruvate and L-aspartate – β-semialdehyde		
8	Alkaline phosphatase superfamily	Alkaline phosphatase	Phosphomonoesters hydrolysis (p-nitrophenyl phosphate, 3.3×10^7)	Phosphodiesters hydrolysis (bis-p-nitrophenyl phosphate, 5×10^{-2}) Sulfate ester hydrolysis (p-nitrophenyl sulfate, 1×10^{-2})	82–84
		Nucleotide pyrophosphatase/ Phosphodiesterase (NPP)	Phosphodiesters hydrolysis (thymidine 5′-monophosphate 4-nitrophenyl ester, 1.6×10^6)	Phosphomonoesters hydrolysis (p-nitrophenyl phosphate, 1.1)	

9	Guanidino-modifying enzyme superfamily (GMSF), hydrolase branch	Arginine deiminase (PaADI)	Arginine hydrolysis (4.5×10^4)	N^w,N^w-dimethylarginine hydrolysis (PaDDAH activity) (1.8×10^3)	110
		Agmantine deiminase (PaAgDI)	Agmantine hydrolysis (7×10^3)		
		N^w,N^w-dimethyl-arginine dimethyl-aminohydrolase (PaDDAH)	N^w,N^w-dimethylarginine hydrolysis (1.8×10^3)	Arginine hydrolysis (PaDDAH activity) (1.8)	
10	C–C hydrolase family (branch of α/β hydrolase superfamily)	C–C hydrolase MhpC from E. coli	C–C bond cleavage (2-hydroxy 6-keto-nona-2,4-dienoic acid, 28 units)	Esterase (mono-ethyl adipate, 0.0027 units)	135–137
				Thioesterase (thioethyl adipate, 0.46 units)	
				Hydroxamic acid formation (mono-ethyl adipate + NH_4OH, 0.013 units)	
		Haloperoxidase/Esterase ThcF from Rhodococcus erythropolis	Haloperoxidase (monochlorodimedon, $V_{max} = 0.45$ nmol min^{-1})	Esterase (p-nitrophenyl acetate, $V_{max} = 2.58$ nmol min^{-1})	
		Lactonase from Acinetobacter calcoaceticus	Lactonase (3,4-dihydrocoumarin, $V_{max} = 4760$ units)	Haloperoxidase (monochlorodimedon, $V_{max} = 199$ units)	

(i) The same promiscuous activity is often shared between more than one family member (**Table 2**, entries 2, 3, 5, 10). For example, several members of the tautomerase family share a promiscuous hydratase activity (although its efficiency varies greatly),[116,117] and all known kinases of the ROK family can utilize glucose as a promiscuous substrate.[71,72] That the same promiscuous activity is shared by more than one family member can hint toward the existence of yet unidentified family members in which this promiscuous activity comprises the native activity, and also provide new starting points for directed evolution.[87,88]

(ii) Promiscuous functions can also appear in one family member but not in the others (**Table 2**, entries 1, 5, 9). For example, in the guanidino-modifying enzyme superfamily (GMSF), no promiscuity was observed in agmantine deiminase (PaAgDI), whereas the other two family members exhibit promiscuous activities.[110] In the mammalian paraoxonases family, the promiscuous PTE activity is quite efficient in PON1 ($k_{cat}/K_M \sim 10^4 \, mol^{-1} l \, s^{-1}$), barely detectable in PON3, and nonexistent in PON2. Indeed, the consistency of the lactonase function in all PON family members, and the haphazardness of the others activities (i.e., the paraoxonase and aryl esterase observed only in some family members; **Table 2**, entry 1), enabled the identification of the lactonase as the native function of PONs.[91,92] This trend is also seen in a recently identified family of bacterial lactonases within the amidohydrolase superfamily. The magnitude (in terms of k_{cat}/K_M) of promiscuous phosphotriesterase activity varies from $0.5 \, mol^{-1} l \, s^{-1}$ in one member, to $4000 \, mol^{-1} l \, s^{-1}$ in another.[88] The above patterns are consistent with promiscuous activities being under no selection, and also with the observation that promiscuous activities show large increases and decreases in response to one or few mutations which are neutral with respect to the primary function.[90]

(iii) The primary, or native, function of one family member is often identified as a promiscuous activity in other family members (**Table 2**, entries 2, 5–9). In the tautomerase superfamily, the promiscuous activity of 4-OT, and its homologue YwhB tautomerase, is the primary native activity of *trans*-3-chloroacrylic acid dehalogenase (CaaD).[113] A similar picture is observed in the orotidine 5′ monophosphate decarboxylase (OMPDC) family, the amidohydrolase superfamily, the *N*-acetyl-neuraminate lyase (NAL) family, the guanidino-modifying enzyme family, and in the alkaline phosphatase superfamily.[133]

(iv) Following the above, promiscuous activities may comprise a vestige of the progenitor of an enzyme, or enzyme family (**Figure 5**). This principle was recently demonstrated in an attempt to trace the origins of a bacterial PTE (from *P. diminuta*) – an enzyme thought to have evolved toward the degradation of the synthetic insecticide paraoxon that has only been introduced in the twentieth century. It was found that PTE possesses a promiscuous lactonase activity,[87] and assumed that this activity could comprise a vestige of its progenitor. Indeed, three homologues from the same superfamily (amidohydrolase) turned out to be representatives of a new group of microbial lactonases dubbed PTE-like lactonases (PLLs). These enzymes proficiently hydrolyze lactones, and in particular *N*-acyl homoserine quorum-sensing lactones, and exhibit weaker promiscuous PTE activities. PLLs share key sequence and active-site features with PTE, and differ primarily by an insertion in one active-site loop. Following their biochemical and biological function, PLLs are probable to have existed for many millions of years. We therefore suggested that PTE could have evolved from a member of the PLL family while utilizing its latent promiscuous paraoxonase activity as an essential starting point.[88]

(v) Laboratory evolution of one promiscuous activity often leads, indirectly, to the appearance of other promiscuous activities (e.g., **Table 3**, entries 1 and 3) thus yielding 'generalist' intermediates.[74] Some of the latter might appear in other family members, as either their native, or promiscuous function.[73,87]

8.03.8   Evolutionary Aspects of Promiscuity

As mentioned in Section 8.03.7, the studies of divergent evolution within enzyme families and superfamilies provide support for the hypothesis that, throughout evolution, promiscuous activities served as the starting points for the divergence of new functions, and for Jensen's hypothesis that broad-specificity enzymes served as progenitors for the whole families and superfamilies of today's specialized enzymes.[25] At present, however, this evidence is largely circumstantial, and provides little insight as to the actual mechanisms, and mutational paths, that underline the processes of divergence. Discussed below are several issues that relate to the evolution of new enzymatic functions, and various models that describe the divergent evolution of new genes carrying new enzymatic functions.

8.03.8.1 The Evolvability of Promiscuous Enzyme Functions

Accumulating experience in the laboratory indicates that there are very few cases in which 'something could be evolved out of nothing', namely that a completely novel activity was evolved, or engineered, in the laboratory. For example, the emergence of an enzymatic function in a noncatalytic fold demanded the exploration of vast libraries, the genetic diversity of which exceeded natural genetic diversities by many orders of magnitude.[138] In another case, although the starting point was an enzyme from the same super-family, and the key active-site catalytic features (a bi-metallo catalytic center in this case) were maintained, the incorporation of a novel function demanded major sequence alterations such as simultaneous deletion and insertion of few active-site loops, a series of engineered point mutations, and the parallel exploration of random mutations.[139] Computational design has also been applied toward the generation of novel enzymes, but the introduction of novel enzymatic functions also involved a large number (>8) simultaneous amino acid changes.[140,141] Most notably, these cases, in which a novel activity was introduced, all involve starting points, and/or intermediates, that possess no activity whatsoever. Evolution, however, is a gradual and smooth process. It involves discrete steps of one mutation at a time (be it a point mutation, an insertion, or a deletion) that yields a folded and functional protein. Thus, all intermediates must be functional, at least to some degree.[142]

8.03.8.1.1 Promiscuity as a starting point – the three basic assumptions

Taking for granted the demand for smooth transitions, it is probable that natural evolution routinely takes advantage of promiscuous activities as starting points for the divergence of new enzymes. However, for promiscuity to benefit, let alone lead, the divergence of new enzyme functions, three basic prerequisites must be assumed.

(i) Once a promiscuous function becomes available, it can be easily improved through one or just few mutations. Indeed, almost all laboratory evolution projects (or at least the successful ones that are reported) aim at further evolving a promiscuous activity, typically for a substrate, or a reaction, that bears some resemblance to the original function. The conclusion from hundreds of such works is that promiscuous functions exhibit high 'plasticity' – few mutations can readily increase a promiscuous activity, typically by $10–10^3$-fold, and $10^4–10^6$-fold improvements in response to a single mutation were also reported.[85,143] More examples, and other aspects related to the evolvability of promiscuous functions, are listed in **Table 3**, and discussed in Sections 8.03.8.3–8.03.8.5.
(ii) Weak promiscuous activities can provide an immediate advantage, and thus become under selection.
(iii) The divergence path can be completed to give a newly specialized enzyme for which, the promiscuous activity became the native one.

These points are discussed in detail in Sections 8.03.8.2–8.03.8.5.

8.03.8.2 Promiscuous Functions Can Provide an Immediate Advantage

Several reports indicate that, when necessary, weak promiscuous activities can provide an immediate selective advantage to an organism. This has been often seen in the emergence of promiscuous functions following a deficiency created by genetic manipulation in the laboratory. Several examples, including a systematic study conducted by Patrick et al.,[50] are discussed in Section 8.03.2. In another study, performed with an E. coli strain deficient of glucokinase activity, several sugar kinases were found that promiscuously phosphorylate glucose.[71] Some of these promiscuous activities are notably weak (**Table 2**); for example, the k_{cat}/K_M values of the promiscuous sugar kinase *YajF* is the range of $10^2 \, mol^{-1} \, l \, s^{-1}$, and is $\sim 10^4$ lower than that of the primary E. coli glucokinase (*Glk*). In all these cases, the promiscuous function complemented a deficiency in a native enzyme by overexpression of the promiscuous enzyme. Clearly, low catalytic efficiency can be compensated by higher enzyme levels.[154] However, the levels of overexpression from a multiple-copy plasmid, and a powerful promoter, are usually not comparable with expression levels from chromosomal copies, and weak promoters, under which most natural enzymes are expressed.

Table 3 Examples for directed evolution of promiscuous enzyme functions and their trade-offs with the native function[a]

	Enzyme	Native activity (k_{cat}/K_M of wild type, $mol^{-1} l s^{-1}$)	Promiscuous activity under selection (k_{cat}/K_M of wild type, $mol^{-1} l s^{-1}$)	Mutations in selected variants	Effect on native activity ($k_{cat}/K_M^{variant}/k_{cat}/K_M^{wt}$)	Effect on the evolved promiscuous activity ($k_{cat}/K_M^{variant}/k_{cat}/K_M^{wt}$)	Comments	Reference(s)
1	Aspartate aminotransferase (AATase) from *E. coli*	Transamination of dicarboxylic substrates (9.1)	Transamination of tyrosine (0.055) and phenylalanine (0.012) (TATase activity)	Pro13Thr Asn69Ser Gly72Asp Arg129Gly Thr167Ala Ala293Val Asn297Ser Asn339Ser Ala381Val Asn396Asp Ala398Val	1.2-fold higher	130- and 270-fold higher, respectively	This work provides a clear example of a 'generalist' intermediate. The *in vitro* evolved enzyme exhibits wild-type-like AATase activity, and TATase activity that is >10% that of wild-type TATase.	73
2	Muconate lactonizing enzyme (MLE II) from *Pseudomonas* sp. P51	Cycloisomerization (2×10^4)	β-Elimination (*o*-succinylbenzoate synthase, OSBS activity). No detectable promiscuous activity (nondetectable) ($<1.5 \times 10^{-3}$)	Glu323Gly	15-fold lower	>1.2 million-fold higher	The corresponding mutation when engineered in a homologous enzyme (Asp297Gly, in AEE) decreased the native function far more significantly (see Vick *et al*.[144]).	85
3	Galactokinase (GalK) from *E. coli*	Phosphorylation of D-galactose to produce α-D-galactose-1-phosphate (860)	Phosphorylation of C5- or C6-substituted sugars (9.8 for D-fucose, and nondetectable for the other substrates)	Tyr371His	1.3-fold lower	21-fold higher for D-fucose, and higher improvements for the other target substrates	This variant expanded the spectrum of substrates to substrates that were not used in the screen. Although the Y317H mutation retains the stringent requirement for the C-4 galactose architecture, it exhibits enhanced substrate flexibility at all other positions.	145
4	β-Glucuronidase (GUS) from *E. coli*	Hydrolysis of β-glucuronides (8.3×10^5)	Hydrolysis of pNP-galactoside (2.3)	Ile12Val Phe365Ser Trp529Leu Ser557Pro Ile560Val	8.3-fold lower	16-fold higher	Larger increases in the evolving promiscuous galactosidase function of *E. coli* GUS, with smaller changes of the native function, and acquisition of specificities not selected for were previously described.[74]	75

#	Enzyme	Native activity	New activity	Mutations	Native activity change	New activity change	Comments	Ref.
5	SinI DNA-methyltransferase from bacteriophage	Methylation of the internal cytosine of the GG(A/T)CC sequence (2.9×10^5)	Relaxation of sequence specificity toward GG(N)CC (2×10^3)	Leu214Ser Tyr229His	4.5-fold lower	18.5-fold higher for the GG(G/C)CC sequence	Similar trends of specificity broadening were observed with *HaeIII* methyltransferase.[76]	146
6	Phosphotriesterase from *Pseudomonas diminuta* (PTE)	Phosphotriesterase (e.g., paraoxon, 4×10^7)	Ester hydrolysis (e.g., 2-naphthyl acetate, 480)	His254Arg Phe306Cys Pro342Ala	Threefold lower	13-fold higher	Up to 150-fold higher activity was observed with esters not selected for.	86, 87
7	Human carbonic anhydrase (hCAII)	Bicarbonate dehydration (3×10^7)	Esterase (e.g., *p*-nitrophenyl acetate, 2×10^3)	Ala65Val, Asp110Asn Thr200Ala	Twofold lower	40-fold higher	Mutations in conserved regions of the protein did not affect the highly proficient native activity despite the absence of a purifying selection for bicarbonate dehydration.	86, 96
8	Mammalian serum Paraoxonase (PON1)	Lipo-lactonase[b] (e.g., δ-valerolactone, 1.3×10^5; and γ-heptanolide, 2×10^4)	Thiolactonase (e.g., γ-butyryl thiolactone,[94])	Ile291Leu Thr332Ala	Approx. no change[b]	80-fold higher	The selected mutations are all located on surface loops that comprise the substrate-binding pocket.	86, 147
			Esterase (e.g., 2-naphthyl octanoate, 1.5×10^3)	Phe292Val Tyr293Asp	Approx. no change	31-fold higher		
			Esterase (e.g., 7-acetoxy coumarin, 1.2×10^5)	Phe292Ser Val346Met	∼22-fold lower	62-fold higher		
			Phosphotri-esterase (7-diethyl-phosphoro 4-cyano-7-hydroxycoumarin, 9×10^3)	Leu69Val Ser138Leu Ser193Pro Asn287Asp	2.6-fold lower	155-fold higher		

Table 3 (Continued)

Enzyme	Native activity (k_{cat}/K_M of wild type, $mol^{-1}ls^{-1}$)	Promiscuous activity under selection (k_{cat}/K_M of wild type, $mol^{-1}ls^{-1}$)	Mutations in selected variants	Effect on native activity ($K_M^{variant}/k_{cat}/K_M^{wt}$)	Effect on the evolved promiscuous activity ($k_{cat}/K_M^{variant}/k_{cat}/K_M^{wt}$)	Comments	Reference(s)
9 Deacetoxy cephalosporin C synthase (DAOCS) from *Streptomyces clavoligerus*	Ring expansion of penicillin N into deacetoxy cephalosporin C (2.2×10^4)	Ring expansion of penicillin G into phenylacetyl-7-aminodeacetoxy-cephalosporanic acid[18]	Val275Ile Ile305Met	1.1-fold higher	32-fold higher		148
			Cys155Tyr Tyr184His Val275Ile Cys281Tyr	42-fold lower	41-fold higher		
11 β-Lactamase TEM-1	Ampicillin hydrolysis (4.18×10^7)	Cefotaxime hydrolysis (2.07×10^3)	Gly238Ser	6.2-fold lower	86-fold higher	These mutants evolved resistance in the clinic and were later reproduced in the laboratory	23
		Ceftazidime hydrolysis (32.1)	Gly238Ser		19-fold higher		
		Cefotaxime hydrolysis (2.07×10^3)	Gly238Ser E104K	29-fold lower	806-fold higher		
		Ceftazidime hydrolysis (32.1)	Gly238Ser Glu104Lys		284-fold higher		
12 Extended-spectrum β-lactamase CTX-M	Hydrolysis of cephalothin and cefotaxime ($4 \times 10^6 – 2 \times 10^7$)	Hydrolysis of ceftazidime (3.3×10^3)	Gln87Leu His112Tyr Thr230Ile, Ala231Val Asp240Gly Arg276His	1.4-fold higher, and 1.4-fold lower, for cephalothin and cefotaxime, respectively	24-fold higher	These mutants evolved resistance in the clinic and were later reproduced in the laboratory	149
13 *NotI* from *Nocardia otitis-caviarum*	Recognition and cleavage of GCGGCCGC DNA sequence (5×10^5 U mg^{-1} enzyme)	Recognition and cleavage of altered 8-bp sequence (no detectable star activity)	Met91Val Glu156Gly	23-fold lower	>32-fold higher than the Glu156Gly intermediate with GCTGCCGC sequence	Although a considerable reduction in the rate of cleavage of the original sequence is reported, the cleavage specificity of M91V/E156 appears to be relaxed toward a whole set of 8 bp sequence targets, with a distinct preference for the original target.	150

14	D-allose kinase (AlsK),	AlsK − phosphorylation of D-allose (2.5×10^5)	AlsK – phosphorylation of D-glucose (3.4×10^2)	Ala73Gly	1.25-fold lower	62-fold higher		151
				Phe145Leu	1.28-fold higher	11.4-fold higher		
	N-acetyl D-mannosamine kinase (NanK)	NanK – phosphorylation of N-acetyl D-mannosamine (1.5×10^5)	NanK – phosphorylation of D-glucose (3.4×10^3)	Leu84Pro	Twofold lower	11.8-fold higher		
				Val138Met	1.25-fold lower	6.4-fold higher		
15	ProFAR isomerase (HisA)	Isomerization of N′-[(5′-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (1.2×10^6)	Isomerization of phosphoribosyl-lanthranilate = TrpF activity (ND)	Asp127Val	$\sim 10^4$-fold lower	The wild-type activity is below detection limits	Almost all the original HisA activity was lost	152

^{a} Shown are examples from the last few years for which kinetic parameters are available for both the promiscuous activity under selection, and the original activity. For more examples see Supplementary Table 8 in Aharoni et al.[86] Because the above analysis aims at providing insights on the evolution of new enzyme functions in nature, the examples selected involve selection for only one parameter – increase in a promiscuous activity, and make use of gene libraries prepared by mutagenesis in a completely random manner (point mutations or shuffling) and throughout the genes.

^{b} Since the publication of this work Aharoni et al.,[86] it has been established that serum paraoxonase (PON1) is a lipo-lactonase, and its preferred substrates are five- and six-membered ring lactones, typically with aliphatic side chains.[94,95,153] In the original article,[90] data for trade-offs with the native activity were presented with both the aromatic lactone dihydrocoumarin, and aliphatic lactones. However, more recent works indicated that dihydrocoumarin does not bind PON1's active site in the same mode as aliphatic lactones.[95,118] Thus, the trade-offs presented here are the average values of two aliphatic lactones (δ-valerolactone and and γ-heptanolide).

A notable exception is the case of alkaline phosphatase, the promiscuous phosphite oxidation of which can complement the deficiency of *E. coli* knockout strains (see Section 8.03.6.1.2). Complementation is this case, and the ability to grow on phosphite as the sole source of inorganic phosphorous, occurred through the chromosomal copy of alkaline phosphatase, owing to the extremely high expression levels of the native alkaline phosphatase under phosphate starvation.[47] In other cases, changes in regulation leading to higher expression were observed. Indeed, few reports describe changes in regulation in cases where a promiscuous enzyme activity came under selection. For example, Hall's classical experiment of the emergence of an alternative *β*-galactosidase (*egb*) was performed on the *E. coli* chromosome, and not through complementing plasmids. Besides mutations that increased a weak promiscuous *β*-galactosidase activity in the *egb* glycosylase (whose native function remains unknown), the first mutation that occurred dramatically increased the expression of *egb* by removing its repressor.[155] Miller and Raines also observed a promoter mutation in a complementing plasmid that led to ∼100-fold increase in expression level of the promiscuous glucokinase *YajF* mentioned above.[72] As described in Section 8.03.8.6 , gene duplication is another relatively abundant event that can lead to an increase in enzyme levels. Thus, if and when a new activity becomes necessary, the combination of a weak promiscuous activity with an increase in enzyme levels can provide the organism an immediate advantage.

8.03.8.3   Negative Trade-offs and the Evolvability of Promiscuous Functions

As discussed above, the first assumption regards the evolvability of promiscuous functions, namely, once a promiscuous function becomes available, it can be easily improved through one or just few mutations (Section 8.03.8.1). A key related issue regards the notion that mutations leading to improvements in promiscuous functions need not induce parallel decreases in the native function.[86] For the reasons explained below, we dubbed this feature as 'the evolvability of promiscuous functions'. Negative trade-offs between the evolving function and existing function are a dominant factor in evolution.[156] It is therefore of crucial importance that in many cases, promiscuous protein functions evolve with negative trade-offs that are weaker than generally assumed. Indeed, the weak trade-off hypothesis directly inflicts on our understanding of how new genes carrying new functions emerge.

Evolvability, or evolutionary adaptability, is the capacity of biological systems, be they organisms, cells, or proteins, to evolve. Evolvability comprises of two elements:[31,157] the first one is the induction of novel phenotypic traits by a relatively low number of mutations (this feature is often dubbed 'plasticity'). As mentioned in Section 8.03.8.1, this property of promiscuous enzyme functions has been established by numerous directed evolution experiments. Not only single proteins, but whole metabolic pathways were found to be plastic and evolvable. Moreover, it seems that the more plastic is a metabolic pathway, the more evolvable are the proteins comprising it.[158] However, plasticity is in conflict with the fact that most mutations are deleterious.[159–161] Since organisms must constantly endure a significant number of mutations while maintaining their fitness, and the structure and function of their proteins, they have evolved a certain level of resistance to the effects of mutations ('robustness').[31] It appears that proteins exhibit both traits, namely plasticity and robustness, and the two need not be mutually exclusive.[31,162] The promiscuous, accidental functions of the protein are highly plastic. They can be reshaped through few mutations that significantly increase or decrease them. However, these mutations need not have a large effect on the protein's native activity. Indeed, the results of many directed evolution experiments indicate that, in clear contrast to the dramatic shifts observed with the promiscuous substrates, the native activities, that take place in the very same active site, often show comparatively small changes. The robustness of the native function is observed despite the fact that the only selection criterion applied in these experiments was an increase in one of the promiscuous activities of the target enzyme.

We have initially described this trend in three different enzymes subjected to a selection for an increase in six different promiscuous activities.[86] The same trend was identified in other laboratory experiments aimed at increasing promiscuous enzymatic and binding activities of various proteins (see Supplementary Table 8 in Aharoni *et al.*[86]). Averaging eighteen cases in which data was provided for the effect of the selected mutations on both the evolving promiscuous activity and the original, native function (for both binding and enzymatic functions) indicated that, 1–4 mutations increased the promiscuous activity that was under selection by >1000-fold, on average, whereas the original activity of these proteins decreased, by 3.2-fold. More recent examples

are listed in **Table 3**. They show a similar trend: 1–11 mutations increased the promiscuous activity under selection by $10-10^6$-fold, whereas the original activity of these proteins decreased, by 0.8–42-fold. In the majority of cases, the ratio of increase in the selected promiscuous function to decrease in the original one is >10. In fewer cases, the change in the evolving promiscuous function, and in the original function, is comparable. Besides, other variants from the same selection show weaker trade-offs (e.g., **Table 3**, entry 9). There are also cases in which large negative trade-offs were observed, and these are addressed in detail in Section 8.03.8.4.

Although out of the scope of this review, it is notable that similar trends can be clearly seen in various receptors, where the acquisition of specificity for a new effector exploits the promiscuity of existing receptors.[163,164] As also demonstrated with bacterial transcription factors, new effector specificities can then be acquired by natural evolution, or laboratory rounds of mutagenesis and selection, often with weak negative trade-offs with respect to the original effector.[165]

There is little doubt that ultimately, the acquisition of a highly proficient new enzyme comes at the expense of the old function. However, the *relative rates* by which a new function is gained, and the old one is lost, matters. The model depicted in **Figure 6** assumes that trade-offs can be determined and quantified, in particular with enzymatic activities. It suggests that in those cases where the negative trade-off is weak (red line), the divergence of a new function can proceed through a 'generalist' intermediate that exhibits broad specificity. Gene duplication may then follow this process, rather than initiate it, and lead to divergence of a new 'specialist'. The results of several directed evolution experiments also convincingly demonstrate that, the concave route – 'strong negative trade-off', is also applicable (**Figure 6**, blue line), in particular when a *dual* selection pressure applies, namely, when a parallel selection for an increase of a promiscuous activity, and decrease in the native activity, is applied.[143,166–168] Thus, in the face of selection for specialization, proteins can evolve to yield 'specialists', sometimes with a surprisingly abrupt shift in selectivity.[143]

Although the convex route denoted in **Figure 6** might be common, in other cases, a single amino acid exchange can completely switch the specificity of an enzyme (see also Section 8.03.8.4). For example, the His89Phe mutation
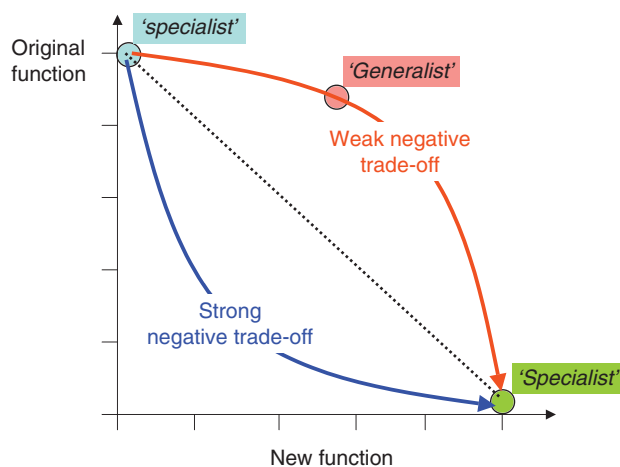
Figure 6 Possible routes to the divergence of a new function. Under selection, a weak, promiscuous activity of a protein with an existing function (blue circle) gradually evolves. By the end of this process, that typically requires many generations of mutation and selection, the 'new' function has traded off with the original one (green circle). However, the dynamics of this process may vary. The gain–loss of the new versus old function, and the conversion of one 'specialist' protein into another, may trade-off linearly (dashed line), or follow either concave, or convex, routes. Results of numerous directed evolution experiments indicate that, the convex route ('weak negative trade-offs') is the most probable one – large increases in the promiscuous function under selection ('new function') are accompanied by significantly smaller decreases in the original function (**Table 3**). By virtue of gaining a 'new' function without losing the original one (and often gaining other new functions not selected for), the intermediates of these routes are 'generalists', and their evolution can therefore proceed *prior* to gene duplication. In contrast, the concave route implies that gene duplication is a necessary prerequisite, because acquisition of even low levels of the 'new' function is accompanied by large losses of the original one. This route is also observed in the laboratory, in particular under a dual selection, for gain of a new function and loss of the old one.

in the active site of tyrosine ammonia lyase (TAL) switched its substrate selectivity from tyrosine to phenylalanine, thereby converting it into an enzyme whose kinetic parameters and selectivity are comparable to native PAL (phenylalanine ammonia lyase).[169] Indeed, in a living cell, the toll of a 'generalist' on fitness might be too high, and the driving force for specialization might be stronger than observed under *in vitro* selection.[166]

The different effects of mutations on the native versus the promiscuous functions are particularly striking in view of the fact that many of these mutations are found within the active site, typically at the wall and perimeter. Structural and thermodynamic insights into the effects of these 'generalist' mutations are needed before any definite statements could be made. Yet it seems probable that the plasticity of these residues lies in the fact that they are not part of the protein's scaffold, or of the catalytic machinery of the enzyme. The mutated residues are typically located on surface loops that exhibit high conformational flexibility and comprise the substrate binding part of the active site.[86,102,103,170,151] As discussed in Section 8.03.6.3, there exist fundamental differences between the mode of binding of the native substrate – that is typically mediated by several independent, enthalpy-driven interactions such as hydrogen bonds – versus the promiscuous substrates where hydrophobic, and other entropy-driven interactions, play an important role.[42,171] It is therefore probable that the same mutation could affect the native and the promiscuous substrates in a very different way. Although, an in-depth understanding of the effect of mutations awaits a sufficient number of cases (or even one case at this stage) in which structures become available for both the wild-type protein and its evolved mutants, in complex with analogues of both the native and promiscuous substrates.

Altogether, the above observations support the hypothesis of evolutionary progenitors and intermediates being of broad specificity, or high promiscuity,[25] and that, a frequent (but not exclusive) evolutionary route leads from a 'specialist' to a 'generalist', and, in turn, to a new 'specialist' (**Figures 5** and **6**). The reconstruction of evolutionary ancestors of both enzymes, and receptors, also supports the idea of 'generalist' progenitors.[172,173] The implications of negative trade-offs, and of 'generalists' and 'specialists', are further discussed in Sections 8.03.8.4–8.03.8.6.

8.03.8.4   Exceptions to Weak Negative Trade-offs

Although the above strongly argues in favor of 'weak negative trade-offs', this generalization has notable exceptions.

8.03.8.4.1   Size and charge considerations

The magnitude of trade-offs is obviously dependent on the structural differences such as size and charge between the original and promiscuous substrates.[96,144] For example, most reported studies involve promiscuous substrates that are larger than the native one, and cases in which both the native and the promiscuous substrates are hydrophobic. In these cases it is easy to see how a mutation that makes the active site larger, and thus increases the activity toward the promiscuous substrate, would not have a drastic effect on the native substrate. However, what about promiscuous substrates that are smaller than the native one? In this case, mutations that reshape the active site to minimize its volume and increase contacts with the smaller substrate might significantly reduce the activity with the larger native substrate. Other cases in which the native and promiscuous activities might trade-off involve differences in charge. Imagine a charged native substrate, and a neutral (let alone hydrophobic) promiscuous substrate. Mutations that favor the charged form are probable to restrict binding of the hydrophobic one, and vice versa. It is probable therefore that as the number of enzymes studied and mutations increase, more refined trade-off rules will be established.

8.03.8.4.2   Stability trade-offs

An important facet of the trade-off concerns the effect of mutations on stability. Most mutations destabilize, and mutations that affect function often exhibit even higher destabilizing effects. Destabilization may result in reduced enzyme levels, due to misfolding and aggregation, proteolytic digestion, or clearance. Thus, some of the mutations that show little effect on specific activity as measured with purified proteins *in vitro* (k_{cat}, or K_M), may lower the enzyme concentration ($[E]_0$), and hence decrease the levels of enzymatic activity *in vivo*. This phenomenon was first highlighted by Wang *et al.*,[23] who studied various mutations found in clinical isolates of TEM-1 β-lactamase that evolved to degrade third-generation cephalosporin antibiotics such as cefotaxime.

The mutations that alter TEM-1's active site are all destabilizing. The key cefotaxime resistance mutation that appeared in the clinic (Gly238Ser) increases k_{cat}/K_M by 86-fold (Wang et al.[23], **Table 3**, entry 11), and only reduces k_{cat}/K_M for ampicillin (a native substrate of TEM-1) by sixfold. However, this mutation has a significant detrimental effect on bacterial growth under ampicillin due to destabilization of the enzyme by $\sim$2 kcal mol^{-1}, and a significant reduction in the levels of soluble, active TEM-1. At later stages, the loss of stability was compensated by Met182Thr, which increases TEM-1 stability by 2.7 kcal mol^{-1}.[23]

Following the TEM-1 analysis, Wang et al.[23] suggested that, in general, mutations that endow a new or improved function, trade-off with protein stability. Thus, for the evolutionary process to continue (in nature or in the laboratory) this loss of stability must be compensated. A more recent study was based on computational predictions of the stability changes for >500 mutations that arose from the directed evolution of 22 different enzymes.[174] The stability effects of function-altering mutations ($\Delta\Delta G$ values) were compared to stability changes arising from any random point mutation in the same enzymes. It was found that, as was the case with TEM-1, mutations that modulate enzymatic functions are mostly destabilizing (average $\Delta\Delta G = +1.1$ kcal mol^{-1}). Although the effects of function-altering mutations are actually not more destabilizing than the 'average' mutation in these enzymes ($+1.3$ kcal mol^{-1}), these mutations weaken stability. The analysis also indicated that many mutations that appear in directed evolution variants with no obvious role in the new function exert stabilizing effects that may compensate for the destabilizing effects of the crucial function-altering mutations. Thus, despite a lack of a specific trade-off between function and stability, the evolution of new enzymatic activities, both in nature and in the laboratory, is dependent on the compensatory, stabilizing effect of apparently 'silent' mutations in regions of the protein that are irrelevant to its function. Indeed, other works showed that limited protein stability constrains the acquisition of new function,[175] and highlighted the role of stability compensators such as Met182Thr of TEM-1 in expediting the evolution of new functions.[176]

Thus, reductions in activity in terms of k_{cat} and K_M (**Tables 3 and 4**) may not reflect the full impact of these mutations. Loss of stability is also part of the trade-off that may accompany a gain in the evolving promiscuous activity.

8.03.8.4.3   *Targeted versus random mutagenesis*

It is also the case that mutations incorporated through rational design show larger trade-offs relative to mutations obtained by selection from random repertoires (**Table 4**, in comparison with **Table 3**). This is not surprising given that rational design usually aims at the replacement of key active-site residues. Exchanges in such key residues yield drastic changes, including dramatic enhancements of a promiscuous function at the expense of the native one. However, mutations isolated in directed evolution rarely occur in key active-site residues, and they typically exhibit more subtle effects. It is probable that key active-site mutations that yield more drastic changes occur at later stages, typically at the last stages of divergence when relatively small improvements in the evolving function trade-off with larger drops in the old one (**Figure 6**, convex route). Nevertheless, some of the engineered mutations (e.g., **Table 4**, entries 2, 6, 7, 10, 11) clearly reflect changes that may lead, or may have actually led, to the divergence of a new function through the strong negative trade-off route (**Figure 6**, concave route).

8.03.8.5   On- and Off-Pathway Evolutionary Intermediates

A notable case of large negative trade-offs was described for the directed evolution of HisA to yield TrpF activity (**Table 3**, entry 15; **Figure 7**). The gene was randomly mutated, and the Asp127Val mutant of *Thermotoga maritima* HisA isolated from the selection exhibited measurable TrpF activity ($k_{cat}/K_M = 120$ mol^{-1} l s^{-1}) that was sufficient to complement the *E. coli* TrpF knockout strain used for the selection. The newly evolving TrpF activity led to a dramatic drop in the original HisA activity ($\sim$10^4-fold). Perhaps this effect is not so surprising given that the starting point had no measurable TrpF activity, and that the mutation occurred in a key active-site residue: Asp127 is the putative acid catalyst in the Amadori rearrangement catalyzed by HisA.[185]

Interestingly, few years later after this directed evolution experiment, a bi-functional enzyme dubbed *PriA* was discovered that performs both reactions with high efficiency, and within the very same active site [186, 187]. In

Table 4 Examples for trade-offs in enzymatic functions following targeted mutations

	Enzyme	Approach	Native activity (k_{cat}/K_M of wild type, mol^{-1} l s^{-1})	Promiscuous/target activity (k_{cat}/K_M of wild type, mol^{-1} l s^{-1})	Incorporated mutations	Effect on native activity ($k_{cat}/K_M^{variant}$ / k_{cat}/K_M^{wt})	Effect on the evolved promiscuous activity ($k_{cat}/K_M^{variant}$ / k_{cat}/K_M^{wt})	Comments	Reference(s)
1	*HaeIII* methyl-transferase	Semi-rational (saturation mutagenesis of 8 positions in the target recognition domain, based on crystal structure)	Methylation of GGCC sites (26 000 mol^{-1} l s^{-1})	Methylation of AGCC sites (1180 mol^{-1} l s^{-1})	Arg225Ala, Asn260Leu Leu261Met Asn262Trp	Ninefold higher	670-fold higher	With the exception of Arg225, the mutated positions do not seem to make direct DNA contact.	76
2	tHisA and tHisF from *Termatoga maritima*	Semi-rational (saturation mutagenesis of Asp127 in HisA and Asp130 in HisF, based on previous results)	HisA: ProFAR isomerase (Amadori re-arrangement). HisF: imidazole glycerol phosphate synthase.	Isomerization of phosphoribosyl-anthranilate (TrpF activity) No detectable promiscuous activity	HisA: Asp127Val Thr164His HisF: Asp130Val/ Thr/ Pro	Not tested	The wild-type activity is below detection limits	A Val residue at position 127 in HisA, and Val, Thr, or Pro residues at position 130 in HisF, are not compatible with HisA or HisF native activities, which require a negatively charged residue at that position	177
3	Aspartate aminotransferase from *E. coli* (eAATase)	Semi-rational: combination of seven of the most frequent mutations when selecting for aromatic specificity (Rothman *et al.*,[179]), plus Ala293Asp	Transamination of dicarboxylic substrates (9.1 mol^{-1} l s^{-1})	Transamination of phenylalanine (eTATase activity, 0.012 mol^{-1} l s^{-1})	Ala12Thr Pro13Thr Asn34Asp Thr109Ser Gly261Ala Ser285Gly Ala293Asp Asn297Ser	40-fold lower	280-fold higher	The selected mutations not only alter the direct contact with the ligand, but also promote flexibility of large regions of the active site.	178

#	Enzyme	Design method	Native reaction	New reaction	Mutation	Native activity change	New activity change	Comments	Ref.
4	Leucine amino-peptidase from *Streptomyces streptatus*	Semi-rational (saturation mutagenesis at position 221, based on predicted structure)	Hydrolysis of L-Leu peptides (or L-Leu-pNA derivative, $1.8 \times 10^5\,\mathrm{mol^{-1}\,l\,s^{-1}}$)	Hydrolysis of L-Phe peptides (or L-Phe-pNA derivative, $5 \times 10^4\,\mathrm{mol^{-1}\,l\,s^{-1}}$)	Phe221Ala	3.7-fold lower	10.2-fold higher	Residue at position 221 was predicted to interact with the side chain of the substrate.	180
5	L-Ala-D/L-Glu epimerase (AEE)	Rational design based on structural alignment with *o*-succinylbenzoate synthase (OSBS, 30% sequence identity)	1,1-Proton transfer ($7.7 \times 10^4\,\mathrm{mol^{-1}\,l\,s^{-1}}$)	β-Elimination (OSBS, $<5 \times 10^{-3}\,\mathrm{mol^{-1}\,l\,s^{-1}}$)	Asp297Gly	7800-fold lower	>2400-fold higher	Large effects of single mutations	85, 144, 181
					Asp297Gly Ile19Phe	>790 000-fold lower	>17 000-fold higher		
					Asp297Gly Ile19Phe Arg24Trp	>10^6-fold lower	>404 000-fold higher		
6	Alanine racemase from *G. stearothermophilus*	Rational design (based on structure and proposed mechanism)	Isomerase	Aldolase ($5 \times 10^{-5}\,\mathrm{mol^{-1}\,l\,s^{-1}}$)	Tyr265Ala	3000-fold lower	2.3×10^5-fold higher	A single point mutation can change both substrate specificity and reaction profile simultaneously	182
7	Lipase B from *C. antarctica*	Rational design (molecular dynamic simulations and automated docking), based on a previously described mutant (Branneby *et al.*[78]).	Hydrolysis of triglyceride esters	Michael-type addition of thiols to α,β-unsaturated carbonyl compounds	Ser105Ala	Not tested, but probable to be >10^3-fold lower	6–1600-fold higher (k_{cat}) relative to wild-type, for the addition of thiols to methyl acrylate.		79

(Continued)

Table 4 (Continued)

	Enzyme	Approach	Native activity (k_{cat}/K_M of wild type, $mol^{-1}\,l\,s^{-1}$)	Promiscuous/target activity (k_{cat}/K_M of wild type, $mol^{-1}\,l\,s^{-1}$)	Incorporated mutations	Effect on native activity ($k_{cat}/K_M^{variant}/k_{cat}/K_M^{wt}$)	Effect on the evolved promiscuous activity ($k_{cat}/K_M^{variant}/k_{cat}/K_M^{wt}$)	Comments	Reference(s)
8	3-Keto-L-gulonate 6-phosphate decarboxylase (KGPDC) from *E. coli*.	Rational design by sequence alignment with D-arabino-hex-3-ulose 6-phosphate synthase (HPS, 30% identity)	Decarboxylation of 3-keto-L-gulonate 6-phosphate ($7.7 \times 10^4\,mol^{-1}\,l\,s^{-1}$)	Aldol condensation of D-ribulose 5-phosphate with formaldehyde ($0.082\,mol^{-1}\,l\,s^{-1}$)	Glu112Asp Arg139Val Thr169Ala	30-fold lower	260-fold higher		133
9	Human GST A2-2	Rational design (based on homology with GST A3-3)	Glutathione peroxidase (2.9×10^4 with cumene peroxide)	Steroid double bond isomerase (1×10^3 with Δ^5-androstene-3,17- dione)	Ser10Phe Ile12Gly Phe111Leu Met208Ala Ser216Ala	Twofold lower	3500-fold higher	Restricted to substrate-binding residues	183
10	Rat liver 3 α-hydroxy-steroid dehydro-genase (HSD) – AKR1C9	Rational design (based on homology with steroid 5β-reductase)	Position and stereo-specific interconversion of steroid ketones and alcohols (1.6×10^5)	Steroid 5β-reductase (reduction of C–C double bonds), no detectable promiscuous activity	His117Glu	>600-fold lower	The wild type activity is below detection limits	Single-point mutation of catalytic residue	184
11	Tyrosine ammonia lyase (TAL) from *Rba. sphaeroides*	Rational design by homology with phenylalanine ammonia lyase (PAL)	Deamination of tyrosine to 4-coumaric acid and ammonia (1.1×10^5)	Deamination of phenylalanine to cinnamic acid and ammonia (403)	His89Phe	>19 000-fold lower	220-fold higher	Single-point mutation of catalytic residue that completely switches specificity	169
12	*N*-acetyl -neuraminate lyase (NAL)	Rational design (by homology with DHDPS)	Cleavage of *N*-acetyl-neuraminate to produce pyruvate and *N*-acetyl mannosamine (3.1×10^3)	Aldol condensation of pyruvate and L-aspartate – β-semialdehyde (DHDPS activity)	Leu142Arg Tyr190Asp Glu192Ala	62-fold lower	Sixfold higher	Despite very modest improvement, a single mutation Leu142Arg is enough for *in vivo* complementation	134
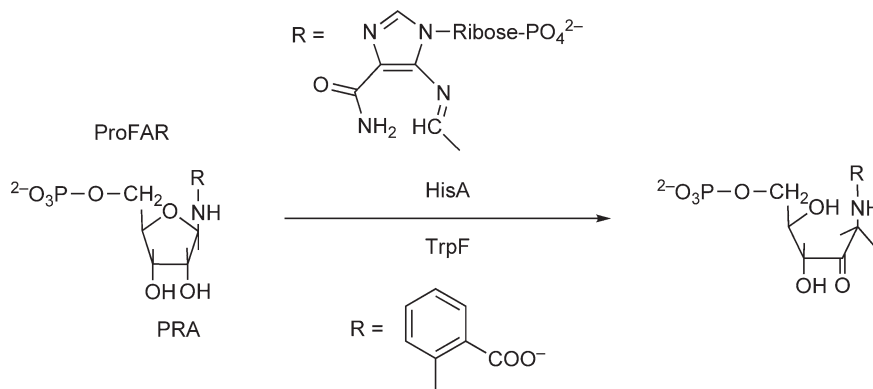
Figure 7   The reactions catalyzed by HisA (isomerization of N'-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide) and TrpF (isomerization of phosphoribosyl-anthranilate).

contrast to the laboratory selection for TrpF activity, in nature, a generalist enzyme evolved under selection to maintain both the HisA and TrpF functions. Thus, the 'generalist' intermediate originally proposed by Jurgens *et al.*[152] exists in certain bacteria. However, it seems that the Asp127Val mutant isolated in the laboratory evolution experiment does not reflect the sequence and structural features of such a generalist intermediate.

In a recent directed evolution experiment, a HisA Asp127Val mutant was further evolved for TrpF activity[188]. However, it led to a complete change in the reaction mechanism, and the evolved variants completely lost the original HisA activity. Thus, although this mutant provides a clear example of how TrpF activity could emerge in HisA, it seems to comprise an 'off-pathway' intermediate – namely, an intermediate that provides a temporary advantage but cannot lead to the eventual divergence of a proficient bifunctional TrpF-HisA enzyme. Off-pathway intermediates might be observed in nature (for example, see the E3 esterase example discussed in Section 8.03.8.6.3). They are as likely, if not more likely, to appear in laboratory evolution experiments, and even more so in rationally designed enzyme variants. It could may well be that some of the 'generalist' intermediates observed in laboratory evolution experiments are also 'off-pathway'. Indeed, the ultimate proof for 'on-pathway' evolutionary intermediates lies in the ability to complete the divergence process and generate a new 'specialist' enzyme with native-like kinetic parameters. Completing the process, however, involves many rounds of mutation and selection, and numerous mutations. This is very rarely pursued, and when it is, the aim is the final enzyme product, and not the pathway leading to it[168,189,190]. Nonetheless, these applicative engineering projects have the potential to provide interesting insights into the pathway, the role and order of individual mutations, and the nature of the intermediates.

8.03.8.6   Promiscuity and the Mechanisms for the Divergence of New Gene Functions

8.03.8.6.1   Ohno's model

The mechanisms governing the divergence of new gene/protein functions comprise a central part of evolutionary theory. Early models that have become known as Ohno's model,[191] were later expanded by Kimura and Ohta,[192] and are currently the textbook paradigm. This model assumes that duplication is a frequent event, which is largely neutral – that is, duplication provides no fitness advantage, or disadvantage, and is therefore not under selection. The redundant duplicated copy is therefore free to accumulate mutations, including deleterious ones. If and when the need arises, some of these mutations that endow a new function come under positive, adaptive selection, thus leading to the divergence of the new gene and function (neo-functionalization).

The fact that duplication and relief from selection is a prerequisite stems from the negative trade-off assumption, namely, maintaining selection for the existing function is expected to purge mutations with adaptive potential. However, as noted in Section 8.03.8.3, in many cases, a promiscuous function can further evolve with little effect on the original function. Indeed, the underlining assumptions of Ohno's model differ fundamentally from what is described in the previous sections. Contrary to the notion that promiscuous

activities can provide an immediate advantage (Section 8.03.8.2), in Ohno's model, the original gene plays a completely passive role in the emergence of a new function.

Indeed, the part of Ohno's model surmises that gene duplication and the subsequent mutational drift occur under no selection is being intensively questioned. First, in contrast to Ohno's model, most duplicated genes found in existing genomes appear to drift under functional selection that purges deleterious mutations.[193–195] Second, expression of redundant copies carries substantial energetic costs,[34,44] and there exists a strong selection pressure to inactivate their expression.[43,196] Third, as discussed in Section 8.03.8.6.4, many cases have been recorded in which gene duplication is not a neutral event, but is rather positively selected under demands for higher protein doses.[197,198] Finally, about a third of random mutations in a given protein are deleterious,[159–161] whereas beneficial mutations that can promote new functions are very rare. Thus, when drifting in the absence of any selection, loss of all functions (nonfunctionalization) primary due to mutations that diminish gene expression, or the ability to form a stable fold,[161,199] is orders of magnitude more probable than neo-functionalization. In the sections that follow we describe alternative mechanisms of divergence that are based on promiscuity and its evolutionary features.

8.03.8.6.2   *Gene sharing*

The first evidence indicating lack of trade-off, and thus the emergence of a new protein function prior to, or even without, gene duplication, led to the hypothesis of 'gene sharing', by which, a gene with a given function is recruited for a very different function without significant changes in the coding region.[200] Gene sharing is a feasible event, as indicated by the classical example of crystallins whereby metabolic enzymes (e.g., argininosuccinate lyase) were recruited later in evolution as structural proteins in eye lenses, with no sequence changes and while retaining their enzymatic activity.[69,200] This is certainly not a singular example. Several other examples for the recruitment of a protein, for example, an enzyme, for completely different tasks, and under different regulation regimes are currently known.[7] Secondary roles of aminoacyl-tRNA synthetases are mentioned in Section 8.03.2. Another notable example is thymidine phosphorylase that also acts as an endothelial growth factor. Indeed, inhibitors of this enzyme also inhibit cell growth.[67,68]

Piatigorsky's findings opened the door for a reconsideration of Ohno's model, but at the same time, it is clear that in most cases, a new function must eventually trade-off with the existing one (**Figure 6**). Hence, the 'gene sharing' model was extended to include cases in which divergence is dependent on duplication.

These alternative models are detailed in the sections that follow. They all include duplication. However, they fundamentally differ from Ohno's model in that duplication is not a neutral event, but it rather provides an immediate fitness advantage. Consequently, both gene copies (the original and the newly duplicated gene) are maintained under selection throughout the process.

8.03.8.6.3   *Divergence prior to duplication*

This model, which comprises an extension of 'gene sharing', gathers growing levels of support.[90,156,193] The 'divergence prior to duplication' model, or IAD (innovation–amplification–divergence)[197] model, assumes that the very first step toward divergence is the selection of a mutant enzyme (an allele) with higher secondary, promiscuous activity. Using this model, the mutant appeared in the population well before the need for a new function had appeared, primarily because the mutation had little, or no, effect on the existing function of this enzyme. Thus, a mutation that initially accumulated as neutral becomes adaptive if and when a change is required. The 'generalist' nature of these apparently neutral mutants is such that duplication may become necessary only at the later stages of further divergence and the emergence of a new specialist. If mutations with adaptive potential accumulate prior to duplication, as neutral (namely with no, or little effect on fitness under a current state), adaptive events become more frequent.

The 'divergence prior to duplication' model is strongly supported by the notion of neutral networks. Theoretical, computational, and experimental works indicate that a drift under selection to maintain the existing function and structure (a 'neutral drift') can increase the potential for adaptation (see Means and Bender[93], Smith,[142] Nei,[201] and Wroe *et al.*[202] and references therein). The existence of this latent pleiotropy – a range of promiscuous functions that were neither selected for, nor against (and conformational isomers that can mediate these promiscuous activities) – facilitates evolution by providing ample starting points for new functions while retaining the primary activity.

An intriguing example for the feasibility of 'latent adaptations' came from the recent adaptation of the sheep blowfly *Lucilia cuprina* to organophosphate (OP) pesticides.[16,18] Early in the twentieth century, OP resistance in blowflies evolved through two separate pathways. Initially, the introduction of OPs resulted in the rapid enrichment of the Trp251Leu mutation in carboxyesterase E3 that endowed this enzyme with weak OP hydrolase activity, and thus led to pesticide resistance. Sequencing of blowfly specimens preserved from the period before the introduction of OPs revealed that the Trp251Leu mutation was already present in the population at high frequency. Indeed, this mutation allowed retention of the native esterase activity while improving the promiscuous OP-hydrolase activity. Interestingly, after several years, a second mutation arose (Gly137Asp) that rapidly superseded Trp251Leu in OP-treated populations. In contrast to the Trp251Leu mutation, Gly137Asp leads to a very significant reduction in the esterase activity, and is associated with developmental defects that were later relieved by a suppressor mutation elsewhere in the blowfly genome. Future experiments may reveal whether these two mutations can be combined to yield even higher levels of resistance, or perhaps the neutral Trp251Leu mutation comprises an 'off-pathway' mutation that cannot lead to an enzyme variant with much higher rates of OP hydrolysis (see Section 8.03.8.5).

Recent laboratory experiments followed the notion of a 'neutral drift' by placing an enzyme under mutation and selection to maintain its native function. The data provide empirical evidence in support of the hypothesis that neutrality enables the formation of latent changes, or 'latent adaptation'. It was found[93,203] that latent evolutionary potentials are indeed very frequent within a neutral set of related enzyme mutants, and that these potentials are most often seen as changes in specificity for one or more promiscuous substrates.

The initial manifestation of such 'latent adaptation' might be in providing an immediate selective advantage by expanding the range of activities of existing enzymes. As demonstrated by the E3 esterase case, once a latent promiscuous function has become advantageous due to a change in the environment, neutral mutants in which this function is higher can be rapidly selected. Duplication may follow, and enable further divergence toward a completely new function. The acquisition of neutral mutations therefore shortens the adaptation and enables facile transitions between one function to another (**Figure 8**).

8.03.8.6.4   Duplication is positively selected

Contrary to the above-described model (Section 8.03.8.6.3), in certain cases, duplication may precede divergence. However, contrary to Ohno's model which assumes that duplication is a neutral event, duplication can be under positive selection. Specifically, when divergence capitalizes on a minor or promiscuous activity in an existing protein, immediate selective advantage can be provided by increasing protein doses.[50] Duplication is known to occur under such selection pressures, and under this scenario, both duplicates will be maintained under selection to maintain both the primary function and the minor one.[197]

8.03.8.6.5   Subfunctionalization

By virtue of increasing enzyme doses, duplication can also have a key role in enabling a wider variety of function altering mutations to accumulate. Despite the generally weak trade-offs, at the end of the day, mutations that endow new enzymatic functions have a measurable effect on the existing enzymatic function, and on the enzyme's stability. In particular, mutations that alter enzymatic functions tend to be destabilizing,[23,174] and can thus reduce the levels of soluble active enzyme (Section 8.03.8.6.4). Indeed, in the neutral drift experiment described in Section 8.03.8.6.3 and **Figure 8**, most neutral variants tested *in vitro* exhibited specificity changes (namely, higher ratios of the promiscuous relative to the native activity), and not necessarily improvements in the absolute levels of promiscuous activity. This indicates that the mutations that improved these activities in terms of k_{cat}, or K_M, values for the promiscuous function, may have also reduced expression ($[E]_0$) and hence the total activity (v_0) remained largely unchanged. Even if the effects on enzyme stability, expression levels, and catalytic efficiency in relation with the native activity, are minor relative to the potential innovation benefit, the acquisition of beneficial mutations can only continue as long as the existing function is reduced to an extent that does not compromise organismal fitness. Following that not all genes contribute to organismal fitness at all times, long periods of relaxation in the selection pressure is a conceivable scenario. During such periods, mutations exhibiting weak trade-offs can accumulate. An alternative solution to this problem is gene duplication, provided that, contrary to Ohno's model, the two genes remain under selection. By virtue of the two genes carrying the same level of function, duplication can offer a margin that allows a wider variety of potentially beneficial mutations to accumulate.
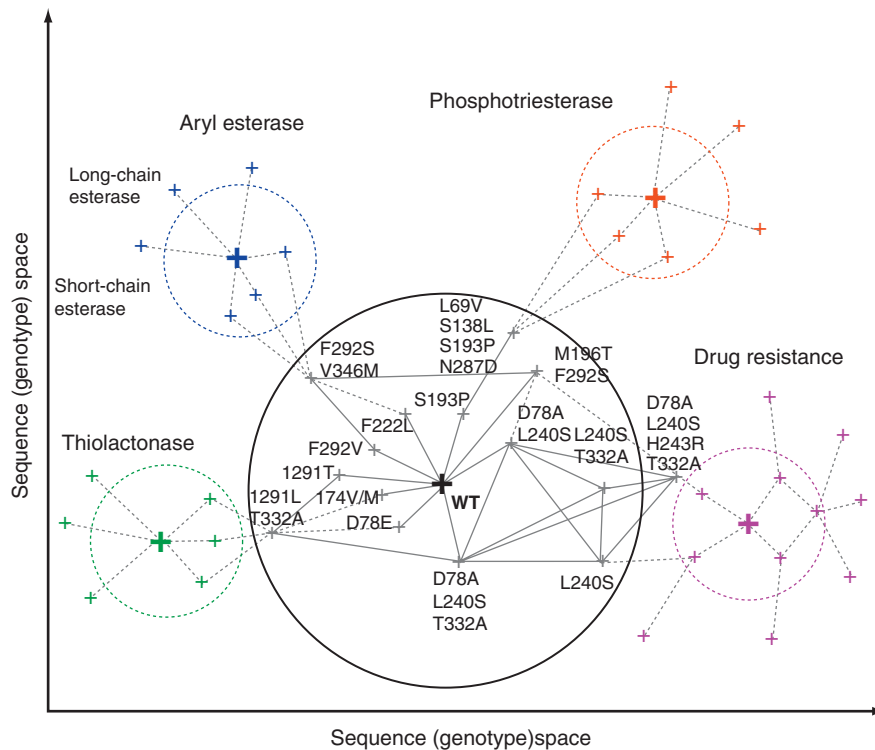
Figure 8   Schematic representation of a putative neutral network of an enzyme. The scheme is based on the results of a 'neutral drift' of serum paraoxonase (PON1) performed in the laboratory, and PON1 mutants identified in various directed evolution experiments. The mutations of neutral variants with significantly altered phenotypes are noted. The large circle denotes the hypothetical boundaries of the neutral network of PON1's native phenotype (lipo-lactonase). The length of the edges corresponds to the degree of neutrality – variants that are more distant from the center (WT) vary in their phenotype to a larger degree. Connecting edges illustrate common mutations between two or more nodes. The other neutral regions relate to new phenotypes that can stem from PON1 (e.g., esterase, phosphotriesterase, and a 'drug-resistant PON1'). The dashed nodes and edges are hypothetical, whereas the continuous ones relate to the apparently neutral variants that were characterized in the experiment. Adapted from G. Amitai; R. Devi-Gupta; D. S. Tawfik, *HFSP J.* **2007**, *1*, 67–78.

Indeed, as previously observed, once duplication has occurred, both copies are probable to be maintained under purifying selection. This may occur in the subfunctionalization (or DDC; duplication–degeneration–complementation) model.[204,205] Following duplication, one, or even both copies may acquire loss-of-function mutations, such that both genes are now required to maintain the level of function provided by the single ancestral gene. Although this model was devised for complex gene functions and regulatory elements, it can be readily extended to simple enzymatic functions. Indeed, a study of yeast genome duplications indicated that the relaxation in selection pressure afforded by redundant gene copies led to the increased accumulation of activity-reducing mutations (ARMs).[195] The appearance of these mutations enforces both copies to be maintained viable as in the DDC model. However, whereas the DDC model describes loss-of-function mutations that can be very difficult to regain (such as mutations that knockout an entire protein domain), the ARMs are dose-dependent mutations. These mutations decrease enzyme dose and activity, and are easily compensated by the enzymatic activity of the second gene copy. As discussed here, most mutations that increase an existing, latent promiscuous activity, and may therefore possess adaptive potential, belong to the category of ARMs in the sense that they reduce in one way or another, but do not abolish, the enzyme's primary activity. Duplication, followed by the fixation of both copies under purifying selection, can provide the key to rapid divergence owing to the fact that duplication allows, and is sustained, by such ARMs.

Acknowledgments

Abbreviations

4-OT	4-oxalocrotonate tautomerase
AEE	L-Ala-D/L-Glu epimerase
Ap4A	two adenosines linked through four phosphates
ARM	activity-reducing mutation
CaaD	*trans*-3-chloroacrylic acid dehalogenase
CALB	*Candida antarctica* lipase B
DAOCS	deacetoxycephalosporin C synthase
DDC	duplication–degeneration–complementation model
DFPase	diisopropyl fluorophosphatase
DHDPS	dihydrodi-picolinate synthase
DHO	dihydroorotase
eAATase	aspartate aminotransferase
GADPH	glyceraldehyde-3-phosphate dehydrogenase
GMSF	guanidino-modifying enzyme superfamily
GST	glutathione *S*-transferase
GUS	β-glucuronidase
HAD	superfamily haloacid dehalogenase-like hydrolases
HisA	ProFAR isomerase
HisF	imidazole glycerol phosphate synthase
HPS	D-*arabino*-hex-3-ulose 6-phosphate synthase
HSD	3 α-hydroxy-steroid dehydrogenase
KDGA	D-2-keto-3-deoxy-gluconate aldolase
KGPDS	3′ keto L-gluconate 6-phosphate decarboxylase
MLE	muconate lactonizing enzyme
MSAD	malonate semialdehyde decarboxylase
NAL	*N*-acetyl-neuraminate lyase
NPP	nucleotide pyrophosphatase
OMPDC	orotidine 5′ monophosphate decarboxylase suprafamily
OP	organophosphate
OSBS	*o*-succinylbenzoate synthase
PaADI	arginine deiminase
PaAgDI	agmantine deiminase
PaDDAH	N^w,N^w-dimethyl-arginine dimethyl-aminohydrolase
PAL	phenylalanine ammonia lyase
PchB	isochorismate pyruvate-lyase
PHP	phosphotriesterase homology protein
PLL	PTE-like lactonase
PON1	serum paraoxonase
PTE	phosphotriesterase
TAL	tyrosine ammonia lyase
TrpF	phosphoribosyl-anthranilate isomerase
WT	wild type

References

1. K. Hult; P. Berglund, *Trends Biotechnol.* **2007**, *25*, 231–238.
2. Y. Pocker; J. T. Stone, *J. Am. Chem. Soc.* **1965**, *87*, 5497–5498.
3. T. W. Reid; D. Fahrney, *J. Am. Chem. Soc.* **1967**, *89*, 3941–3943.
4. Y. Nakagawa; M. L. Bender, *J. Am. Chem. Soc.* **1969**, *91*, 1566–1567.
5. R. C. Jackson; R. E. Handschumacher, *Biochemistry* **1970**, *9*, 3585–3590.
6. P. J. O'Brien; D. Herschlag, *Chem. Biol.* **1999**, *6*, R91–R105.
7. S. D. Copley, *Curr. Opin. Chem. Biol.* **2003**, *7*, 265–272.
8. U. T. Bornscheuer; R. J. Kazlauskas, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 6032–6040.
9. R. J. Kazlauskas, *Curr. Opin. Chem. Biol.* **2005**, *9*, 195–201.
10. C. A. Bonner; T. Disz; K. Hwang; J. Song; V. Vonstein; R. Overbeek; R. A. Jensen, *Microbiol. Mol. Biol. Rev.* **2008**, *72*, 13–53, table of contents.
11. M. E. Glasner; J. A. Gerlt; P. C. Babbitt, *Curr. Opin. Chem. Biol.* **2006**, *10*, 492–497.
12. P. J. O'Brien, *Chem. Rev.* **2006**, *106*, 720–752.
13. D. R. Palmer; J. B. Garrett; V. Sharma; R. Meganathan; P. C. Babbitt; J. A. Gerlt, *Biochemistry* **1999**, *38*, 4252–4258.
14. O. Khersonsky; C. Roodveldt; D. S. Tawfik, *Curr. Opin. Chem. Biol.* **2006**, *10*, 498–508.
15. M. D. Toscano; K. J. Woycechowsky; D. Hilvert, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 3212–3236.
16. C. J. Hartley; R. D. Newcomb; R. J. Russell; C. G. Yong; J. R. Stevens; D. K. Yeates; J. La Salle; J. G. Oakeshott, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8757–8762.
17. D. B. Janssen; I. J. Dinkla; G. J. Poelarends; P. Terpstra, *Environ. Microbiol.* **2005**, *7*, 1868–1882.
18. R. D. Newcomb; P. M. Campbell; D. L. Ollis; E. Cheah; R. J. Russell; J. G. Oakeshott, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7464–7468.
19. F. M. Raushel; H. M. Holden, *Adv. Enzymol. Relat. Areas Mol. Biol.* **2000**, *74*, 51–93.
20. L. P. Wackett, *J. Biol. Chem.* **2004**, *279*, 41259–41262.
21. M. Barlow; B. G. Hall, *J. Mol. Evol.* **2002**, *55*, 314–321.
22. B. G. Hall, *Nat. Rev. Microbiol.* **2004**, *2*, 430–435.
23. X. Wang; G. Minasov; B. K. Shoichet, *J. Mol. Biol.* **2002**, *320*, 85–95.
24. D. M. Weinreich; N. F. Delaney; M. A. Depristo; D. L. Hartl, *Science* **2006**, *312*, 111–114.
25. R. A. Jensen, *Annu. Rev. Microbiol.* **1974**, *30*, 409–425.
26. A. Fersht, *Structure and Mechanism in Protein Science*; W. H. Freeman and Company: New York, 1999.
27. O. Kotik-Kogan; N. Moor; D. Tworowski; M. Safro, *Structure* **2005**, *13*, 1799–1807.
28. H. Yang; G. Zheng; X. Peng; B. Qiang; J. Yuan, *FEBS Lett.* **2003**, *552*, 95–98.
29. L. Reshetnikova; N. Moor; O. Lavrik; D. G. Vassylyev, *J. Mol. Biol.* **1999**, *287*, 555–568.
30. J. Piatigorsky, *Gene Sharing and Evolution: The Diversity of Protein Functions*; Harvard University Press: Cambridge, MA; London, 2007.
31. A. Wagner, *FEBS Lett.* **2005**, *579*, 1772–1778.
32. A. Brevet; P. Plateau; B. Cirakoglu; J. P. Pailliez; S. Blanquet, *J. Biol. Chem.* **1982**, *257*, 14613–14615.
33. Y. N. Lee; H. Nechushtan; N. Figov; E. Razin, *Immunity* **2004**, *20*, 145–151.
34. A. Wagner, *Mol. Biol. Evol.* **2005**, *22*, 1365–1374.
35. G. A. Michaud; M. Salcius; F. Zhou; R. Bangham; J. Bonin; H. Guo; M. Snyder; P. F. Predki; B. I. Schweitzer, *Nat. Biotechnol.* **2003**, *21*, 1509–1512.
36. J. M. Varga; G. Kalchschmid; G. F. Klein; P. Fritsch, *Mol. Immunol.* **1991**, *28*, 655–659.
37. J. M. Varga; G. Kalchschmid; G. F. Klein; P. Fritsch, *Mol. Immunol.* **1991**, *28*, 641–654.
38. J. Inman, The Antibody Combining Region: Speculation on the Hypothesis of General Multispecificity. In *Theoretical Immunology*; G. I. Bell, A. S. Perelson; G. H. Pimbley, Jr., Eds.; Marcel Dekker, Inc.: New York, 1978.
39. D. Lancet; E. Sadovsky; E. Seidemann, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 3715–3719.
40. A. S. Perelson; G. F. Oster, *J. Theor. Biol.* **1979**, *81*, 645–670.
41. A. D. Griffiths; D. S. Tawfik, *Curr. Opin. Biotechnol.* **2000**, *11*, 338–353.
42. L. C. James; D. S. Tawfik, *Protein Sci.* **2003**, *12*, 2183–2193.
43. E. Dekel; U. Alon, *Nature* **2005**, *436*, 588–592.
44. D. M. Stoebel; A. M. Dean; D. E. Dykhuizen, *Genetics* **2008**, *178*, 1653–1660.
45. E. Kuznetsova; M. Proudfoot; C. F. Gonzalez; G. Brown; M. V. Omelchenko; I. Borozan; L. Carmel; Y. I. Wolf; H. Mori; A. V. Savchenko; C. H. Arrowsmith; E. V. Koonin; A. M. Edwards; A. F. Yakunin, *J. Biol. Chem.* **2006**, *281*, 36149–36161.
46. L. W. Tremblay; D. Dunaway-Mariano; K. N. Allen, *Biochemistry* **2006**, *45*, 1183–1193.
47. K. Yang; W. W. Metcalf, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7919–7924.
48. R. D'Ari; J. Casadesus, *Bioessays* **1998**, *20*, 181–186.
49. A. Macchiarulo; I. Nobeli; J. M. Thornton, *Nat. Biotechnol.* **2004**, *22*, 1039–1045.
50. W. M. Patrick; E. M. Quandt; D. B. Swartzlander; I. Matsumura, *Mol. Biol. Evol.* **2007**, *24*, 2716–2722.
51. A. Kurakin, *J. Mol. Recog.* **2007**, *20*, 205–214.
52. R. J. Greenspan, *Nat. Rev. Genet.* **2001**, *2*, 383–387.
53. B. Van Swinderen; R. J. Greenspan, *Genetics* **2005**, *169*, 2151–2163.
54. R. Bernhardt, *J. Biotechnol.* **2006**, *124*, 128–145.
55. M. Ekroos; T. Sjogren, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13682–13687.
56. K. E. Griswold; N. S. Aiyappan; B. L. Iverson; G. Georgiou, *J. Mol. Biol.* **2006**, *364*, 400–410.
57. L. Hou; M. T. Honaker; L. M. Shireman; L. M. Balogh; A. G. Roberts; K. C. Ng; A. Nath; W. M. Atkins, *J. Biol. Chem.* **2007**, *282*, 23264–23274.
58. A. Nath; W. M. Atkins, *Biochemistry* **2008**, *47*, 157–166.

59. S. Lodeiro; Q. Xiong; W. K. Wilson; M. D. Kolesnikova; C. S. Onak; S. P. Matsuda, *J. Am. Chem. Soc.* **2007**, *129*, 11213–11222.
60. C. L. Steele; J. Crock; J. Bohlmann; R. Croteau, *J. Biol. Chem.* **1998**, *273*, 2078–2089.
61. S. Pilkington; H. Dalton, Soluble Methane Monooxygenase from Methylococcus Capsulatus Bath. In *Methods in Enzymology*; M. Lindstrom, Ed.; Academic Press: San Diego, 1990; Vol. 188, pp 181–190.
62. M. Konarzycka-Bessler; U. T. Bornscheuer, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 1418–1420.
63. M. T. Reetz, *Curr. Opin. Chem. Biol.* **2002**, *6*, 145–150.
64. G. Hasnaoui-Dijoux; M. Majeric Elenkov; J. H. Lutje Spelberg; B. Hauer; D. B. Janssen, *ChemBioChem* **2008**, *9*, 1048–1051.
65. J. Domenech; J. Ferrer, *Biochim. Biophys. Acta* **2006**, *1760*, 1667–1674.
66. C. Jeffery, *Trends Biochem. Sci.* **1999**, *24*, 8–11.
67. M. Haraguchi; K. Miyadera; K. Uemura; T. Sumizawa; T. Furukawa; K. Yamada; S. Akiyama; Y. Yamada, *Nature* **1994**, *368*, 198.
68. S. Liekens; A. Bronckaers; M. J. Perez-Perez; J. Balzarini, *Biochem. Pharmacol.* **2007**, *74*, 1555–1567.
69. J. Piatigorsky; W. E. O'Brien; B. L. Norman; K. Kalumuck; G. J. Wistow; T. Borras; J. M. Nickerson; E. F. Wawrousek, *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 3479–3483.
70. A. Allali-Hassani; P. W. Pan; L. Dombrovski; R. Najmanovich; W. Tempel; A. Dong; P. Loppnau; F. Martin; J. Thornton; A. M. Edwards; A. Bochkarev; A. N. Plotnikov; M. Vedadi; C. H. Arrowsmith, *PLoS Biol.* **2007**, *5*, e97.
71. B. G. Miller; R. T. Raines, *Biochemistry* **2004**, *43*, 6387–6392.
72. B. G. Miller; R. T. Raines, *Biochemistry* **2005**, *44*, 10776–10783.
73. S. C. Rothman; J. F. Kirsch, *J. Mol. Biol.* **2003**, *327*, 593–608.
74. I. Matsumura; A. D. Ellington, *J. Mol. Biol.* **2001**, *305*, 331–339.
75. L. A. Rowe; M. L. Geddie; O. B. Alexander; I. Matsumura, *J. Mol. Biol.* **2003**, *332*, 851–860.
76. H. M. Cohen; D. S. Tawfik; A. D. Griffiths, *Protein Eng. Des. Sel.* **2004**, *17*, 3–11.
77. E. Henke; U. T. Bornscheuer, *Anal Chem.* **2003**, *75*, 255–260.
78. C. Branneby; P. Carlqvist; A. Magnusson; K. Hult; T. Brinck; P. Berglund, *J. Am. Chem. Soc.* **2003**, *125*, 874–875.
79. P. Carlqvist; M. Svedendahl; C. Branneby; K. Hult; T. Brinck; P. Berglund, *ChemBioChem* **2005**, *6*, 331–336.
80. O. Torre; I. Alfonso; V. Gotor, *Chem. Commun. (Camb.)* **2004**, *15*, 1724–1725.
81. A. R. Bassindale; K. F. Brandstadt; T. H. Lane; P. G. Taylor, *J. Inorg. Biochem.* **2003**, *96*, 401–406.
82. P. J. O'Brien; D. Herschlag, *Biochemistry* **2001**, *40*, 5691–5699.
83. I. Catrina; P. J. O'Brien; J. Purcell; I. Nikolic-Hughes; J. G. Zalatan; A. C. Hengge; D. Herschlag, *J. Am. Chem. Soc.* **2007**, *129*, 5760–5765.
84. P. J. O'Brien; D. Herschlag, *J. Am. Chem. Soc.* **1998**, *120*, 12369–12370.
85. D. M. Schmidt; E. C. Mundorff; M. Dojka; E. Bermudez; J. E. Ness; S. Govindarajan; P. C. Babbitt; J. Minshull; J. A. Gerlt, *Biochemistry* **2003**, *42*, 8387–8393.
86. A. Aharoni; L. Gaidukov; O. Khersonsky; Q. G. S. Mc; C. Roodveldt; D. S. Tawfik, *Nat. Genet.* **2005**, *37*, 73–76.
87. C. Roodveldt; D. S. Tawfik, *Biochemistry* **2005**, *44*, 12728–12736.
88. L. Afriat; C. Roodveldt; G. Manco; D. S. Tawfik, *Biochemistry* **2006**, *45*, 13677–13686.
89. D. Liu; B. W. Lepore; G. A. Petsko; P. W. Thomas; E. M. Stone; W. Fast; D. Ringe, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 11882–11887.
90. G. Amitai; R. Devi-Gupta; D. S. Tawfik, *HFSP J.* **2007**, *1*, 67–78.
91. D. I. Draganov; J. F. Teiber; A. Speelman; Y. Osawa; R. Sunahara; B. N. La Du, *J. Lipid Res.* **2005**, *46*, 1239–1247.
92. O. Khersonsky; D. S. Tawfik, *Biochemistry* **2005**, *44*, 6371–6382.
93. G. E. Means; M. L. Bender, *Biochemistry* **1975**, *14*, 4989–4994.
94. M. A. Sogorb; V. Carrera; M. Benabent; E. Vilanova, *Chem. Res. Toxicol.* **2002**, *15*, 520–526.
95. F. Hollfelder; A. J. Kirby; D. S. Tawfik, *Nature* **1996**, *383*, 60–62.
96. S. M. Gould; D. S. Tawfik, *Biochemistry* **2005**, *44*, 5444–5452.
97. A. Fernandez-Gacio; A. Codina; J. Fastrez; O. Riant; P. Soumillion, *ChemBioChem* **2006**, *7*, 1013–1016.
98. E. Morett; G. Saab-Rincon; L. Olvera; M. Olvera; H. Flores; R. Grande, *J. Mol. Biol.* **2008**, *376*, 839–853.
99. W. M. Patrick; I. Matsumura, *J. Mol. Biol.* **2008**, *377*, 323–336.
100. F. M. Menger; M. Ladika, *J. Am. Chem. Soc.* **1978**, *109*, 3145–3146.
101. J. C. Hermann; R. Marti-Arbona; A. A. Fedorov; E. Fedorov; S. C. Almo; B. K. Shoichet; F. M. Raushel, *Nature* **2007**, *448*, 775–779.
102. N. U. Gamage; S. Tsvetanov; R. G. Duggleby; M. E. McManus; J. L. Martin, *J. Biol. Chem.* **2005**, *280*, 41482–41486.
103. L. C. James; D. S. Tawfik, *Trends Biochem. Sci.* **2003**, *28*, 361–368.
104. S. Meier; S. Ozbek, *Bioessays* **2007**, *29*, 1095–1104.
105. R. Bone; J. L. Silen; D. A. Agard, *Nature* **1989**, *339*, 191–195.
106. R. Bone; D. Frank; C. A. Kettner; D. A. Agard, *Biochemistry* **1989**, *28*, 7600–7609.
107. J. J. Perona; A. M. Martin, *J. Mol. Biol.* **1997**, *273*, 207–225.
108. Y. Yasutake; M. Yao; N. Sakai; T. Kirita; I. Tanaka, *J. Mol. Biol.* **2004**, *344*, 325–333.
109. J. M. Turner; J. Graziano; G. Spraggon; P. G. Schultz, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6483–6488.
110. X. Lu; L. Li; R. Wu; X. Feng; Z. Li; H. Yang; C. Wang; H. Guo; A. Galkin; O. Herzberg; P. S. Mariano; B. M. Martin; D. Dunaway-Mariano, *Biochemistry* **2006**, *45*, 1162–1172.
111. H. J. Lamble; N. I. Heyer; S. D. Bull; D. W. Hough; M. J. Danson, *J. Biol. Chem.* **2003**, *278*, 34066–34072.
112. A. Theodossis; H. Walden; E. J. Westwick; H. Connaris; H. J. Lamble; D. W. Hough; M. J. Danson; G. L. Taylor, *J. Biol. Chem.* **2004**, *279*, 43886–43892.
113. S. C. Wang; W. H. Johnson; C. P. Whitman, Jr., *J. Am. Chem. Soc.* **2003**, *125*, 14282–14283.
114. S. C. Wang; M. D. Person; W. H. Johnson; C. P. Whitman, Jr., *Biochemistry* **2003**, *42*, 8762–8773.
115. G. J. Poelarends; V. P. Veetil; C. P. Whitman, *Cell Mol. Life Sci.* **2008**, *65*, 3606–3618.
116. G. J. Poelarends; H. Serrano; W. H. Johnson; D. W. Hoffman, Jr.; C. P. Whitman, *J. Am. Chem. Soc.* **2004**, *126*, 15658–15659.
117. G. J. Poelarends; H. Serrano; W. H. Johnson; C. P. Whitman, Jr., *Biochemistry* **2005**, *44*, 9375–9381.
118. O. Khersonsky; D. S. Tawfik, *J. Biol. Chem.* **2006**, *281*, 7649–7656.

119. M. M. Blum; F. Lohr; A. Richardt; H. Ruterjans; J. C. Chen, *J. Am. Chem. Soc.* **2006**, *128*, 12750–12757.
120. G. Amitai; L. Gaidukov; R. Adani; S. Yishay; G. Yacov; M. Kushnir; S. Teitlboim; M. Lindenbaum; P. Bel; O. Khersonsky; D. S. Tawfik; H. Meshulam, *FEBS J.* **2006**, *273*, 1906–1919.
121. D. T. Yeung; D. E. Lenz; D. M. Cerasoli, *FEBS J.* **2005**, *272*, 2225–2230.
122. M. M. Blum; C. M. Timperley; G. R. Williams; H. Thiermann; F. Worek, *Biochemistry* **2008**, *47*, 5216–5224.
123. E. T. Kaiser; D. S. Lawrence, *Science* **1984**, *226*, 505–511.
124. M. Bakker; F. Van Rantwijk; R. A. Sheldon, *Can. J. Chem./Rev. Can. Chim.* **2002**, *80*, 622–625.
125. G. F. Da Silva; L. J. Ming, *J. Am. Chem. Soc.* **2005**, *127*, 16380–16381.
126. Z. P. Wu; D. Hilvert, *J. Am. Chem. Soc.* **1990**, *112*, 5647–5648.
127. S. Boschi-Muller; S. Muller; A. Van Dorsselaer; A. Bock; G. Branlant, *FEBS Lett.* **1998**, *439*, 241–245.
128. H. J. Yu; J. Q. Liu; A. Bock; J. Li; G. M. Luo; J. C. Shen, *J. Biol. Chem.* **2005**, *280*, 11930–11935.
129. R. Kourist; S. Bartsch; L. Fransson; K. Hult; U. T. Bornscheuer, *ChemBioChem* **2008**, *9*, 67–69.
130. M. Forconi; D. Herschlag, *J. Am. Chem. Soc.* **2005**, *127*, 6160–6161.
131. D. E. Kunzler; S. Sasso; M. Gamper; D. Hilvert; P. Kast, *J. Biol. Chem.* **2005**, *280*, 32827–32834.
132. E. A. Taylor Ringia; J. B. Garrett; J. B. Thoden; H. M. Holden; I. Rayment; J. A. Gerlt, *Biochemistry* **2004**, *43*, 224–229.
133. W. S. Yew; J. Akana; E. L. Wise; I. Rayment; J. A. Gerlt, *Biochemistry* **2005**, *44*, 1807–1815.
134. A. C. Joerger; S. Mayer; A. R. Fersht, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5694–5699.
135. R. De Mot; A. De Schrijver; G. Schoofs; A. H. A. Parret, *FEMS Microbiol. Lett.* **2003**, *224*, 197–203.
136. M. Kataoka; K. Honda; S. Shimizu, *Eur. J. Biochem.* **2000**, *267*, 3–10.
137. C. Li; M. Hassler; T. D. H. Bugg, *ChemBioChem* **2008**, *9*, 71–76.
138. B. Seelig; J. W. Szostak, *Nature* **2007**, *448*, 828–831.
139. H. S. Park; S. H. Nam; J. K. Lee; C. N. Yoon; B. Mannervik; S. J. Benkovic; H. S. Kim, *Science* **2006**, *311*, 535–538.
140. L. Jiang; E. A. Althoff; F. R. Clemente; L. Doyle; D. Rothlisberger; A. Zanghellini; J. L. Betker; F. Tanaka; C. F. Barbas; D. Hilvert; K. N. Houk; B. L. Stoddard; D. Baker, *Science* **2008**, *319*, 1387–1391.
141. D. Rothlisberger; O. Khersonsky; A. M. Wollacott; L. Jiang; J. DeChancie; J. Betker; J. L. Gallaher; E. A. Althoff; A. Zanghellini; O. Dym; S. Albeck; K. N. Houk; D. S. Tawfik; D. Baker, *Nature* **2008**, *453*, 190–195.
142. J. M. Smith, *Nature* **1970**, *225*, 563–564.
143. N. Varadarajan; J. Gam; M. J. Olsen; G. Georgiou; B. L. Iverson, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6855–6860.
144. J. E. Vick; D. M. Schmidt; J. A. Gerlt, *Biochemistry* **2005**, *44*, 11722–11729.
145. D. Hoffmeister; J. Yang; L. Liu; J. S. Thorson, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13184–13189.
146. E. Timar; G. Groma; A. Kiss; P. Venetianer, *Nucleic Acids Res.* **2004**, *32*, 3898–3903.
147. A. Aharoni; L. Gaidukov; S. Yagur; L. Toker; I. Silman; D. S. Tawfik, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 482–487.
148. C. L. Wei; Y. B. Yang; C. H. Deng; W. C. Liu; J. S. Hsu; Y. C. Lin; S. H. Liaw; Y. C. Tsai, *Appl. Environ. Microbiol.* **2005**, *71*, 8873–8880.
149. J. Delmas; F. Robin; F. Carvalho; C. Mongaret; R. Bonnet, *Antimicrob. Agents Chemother.* **2006**, *50*, 731–738.
150. J. C. Samuelson; R. D. Morgan; J. S. Benner; T. E. Claus; S. L. Packard; S. Y. Xu, *Nucleic Acids Res.* **2006**, *34*, 796–805.
151. M. Larion; L. B. Moore; S. M. Thompson; B. G. Miller, *Biochemistry* **2007**, *46*, 13564–13572.
152. C. Jurgens; A. Strom; D. Wegener; S. Hettwer; M. Wilmanns; R. Sterner, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9925–9930.
153. L. Gaidukov; D. S. Tawfik, *Biochemsitry* **2005**, *44*, 11843–11854.
154. L. C. James; D. S. Tawfik, *Pro. Sci.* **2001**, *10*, 2600–2607.
155. B. G. Hall, *FEMS Microbiol. Lett.* **1999**, *174*, 1–8.
156. F. A. Kondrashov, *Nat. Genet.* **2005**, *37*, 9–10.
157. M. Kirschner; J. Gerhart, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8420–8427.
158. D. Umeno; A. V. Tobias; F. H. Arnold, *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 51–78.
159. S. Bershtein; M. Segal; R. Bekerman; N. Tokuriki; D. S. Tawfik, *Nature* **2006**, *444*, 929–932.
160. M. Camps; A. Herman; E. Loh; L. A. Loeb, *Crit. Rev. Biochem. Mol. Biol.* **2007**, *42*, 313–326.
161. N. Tokuriki; F. Stricher; J. Schymkowitz; L. Serrano; D. S. Tawfik, *J. Mol. Biol.* **2007**, *369*, 1318–1332.
162. A. Wagner, *Proc. Biol. Sci.* **2008**, *275*, 91–100.
163. C. Adami, *Science* **2006**, *312*, 61–63.
164. J. T. Bridgham; S. M. Carroll; J. W. Thornton, *Science* **2006**, *312*, 97–101.
165. T. C. Galvao; V. de Lorenzo, *Curr. Opin. Biotechnol.* **2006**, *17*, 34–42.
166. T. L. O'Loughlin; D. N. Greene; I. Matsumura, *Mol. Biol. Evol.* **2006**, *23*, 764–772.
167. N. Ran; K. M. Draths; J. W. Frost, *J. Am. Chem. Soc.* **2004**, *126*, 6856–7685.
168. N. Varadarajan; S. Rodriguez; B. Y. Hwang; G. Georgiou; B. L. Iverson, *Nat. Chem. Biol.* **2008**, *4*, 290–294.
169. K. T. Watts; B. N. Mijts; P. C. Lee; A. J. Manning; C. Schmidt-Dannert, *Chem. Biol.* **2006**, *13*, 1317–1326.
170. I. Bertini; V. Calderone; M. Cosenza; M. Fragai; Y. M. Lee; C. Luchinat; S. Mangani; B. Terni; P. Turano, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5334–5339.
171. L. C. James; D. S. Tawfik, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12730–12735.
172. J. W. Thornton; E. Need; D. Crews, *Science* **2003**, *301*, 1714–1717.
173. M. A. Wouters; K. Liu; P. Riek; A. Husain, *Mol. Cell.* **2003**, *12*, 343–354.
174. N. Tokuriki; F. Stritcher; L. Serrano; D. S. Tawfik, *PloS Comput. Biol.* **2008**, *4*, e1000002.
175. J. D. Bloom; J. J. Silberg; C. O. Wilke; D. A. Drummond; C. Adami; F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 606–611.
176. S. Bershtein; K. Goldin; D. S. Tawfik, *J. Mol. Biol.* **2008**, *379*, 1029–1044.
177. S. Leopoldseder; J. Claren; C. Jurgens; R. Sterner, *J. Mol. Biol.* **2004**, *337*, 871–879.
178. M. A. Chow; K. E. McElroy; K. D. Corbett; J. M. Berger; J. F. Kirsch, *Biochemistry* **2004**, *43*, 12780–12787.
179. S. Rothman; M. Voorhies; J. Kirsch, *Protein Science* **2004**, *13* (3), 763.
180. J. Arima; Y. Uesugi; M. Iwabuchi; T. Hatanaka, *Appl. Environ. Microbiol.* **2005**, *71*, 7229–7235.
181. J. E. Vick; J. A. Gerlt, *Biochemistry* **2007**, *46*, 14589–14597.

182. F. P. Seebeck; D. Hilvert, *J. Am. Chem. Soc.* **2003**, *125*, 10158–10159.
183. P. L. Pettersson; A. S. Johansson; B. Mannervik, *J. Biol. Chem.* **2002**, *277*, 30019–30022.
184. J. M. Jez; T. M. Penning, *Biochemistry* **1998**, *37*, 9695–9703.
185. M. Henn-Sax; R. Thoma; S. Schmidt; M. Hennig; K. Kirschner; R. Sterner, *Biochemistry* **2002**, *41*, 12032–12042.
186. F. Barona-Gomez; D. A. Hodgson, *EMBO Rep.* **2003**, *4*, 296–300.
187. H. Wright; L. Noda-Garcia; A. Ochoa-Leyva; D. A. Hodgson; V. Fulop; F. Barona-Gomez, *Biochem. Biophys. Res. Commun.* **2008**, *365*, 16–21.
188. J. Claren; C. Malisi; B. Höcker; R. Sterner, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 3704–3709.
189. L. A. Castle; D. L. Siehl; R. Gorton; P. A. Patten; Y. H. Chen; S. Bertain; H. J. Cho; N. Duck; J. Wong; D. L. Liu; M. W. Lassner, *Science* **2004**, *304*, 1151–1154.
190. I. Sarkar; I. Hauber; J. Hauber; F. Buchholz, *Science* **2007**, *316*, 1912–1915.
191. S. Ohno, *Evolution by Gene Duplication*; Allen & Unwin, Springer-Verlag: London, New York, 1970.
192. M. Kimura; T. Ota, *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 2848–2852.
193. A. L. Hughes, *Trends Genet.* **2002**, *18*, 433–434.
194. M. Lynch; V. Katju, *Trends Genet.* **2004**, *20*, 544–549.
195. D. R. Scannell; K. H. Wolfe, *Genome Res.* **2008**, *18*, 137–147.
196. V. S. Cooper; R. E. Lenski, *Nature* **2000**, *407*, 736–739.
197. U. Bergthorsson; D. I. Andersson; J. R. Roth, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 17004–17009.
198. S. Y. McLoughlin; D. L. Ollis, *Chem. Biol.* **2004**, *11*, 735–737.
199. P. Yue; Z. Li; J. Moult, *J. Mol. Biol.* **2005**, *353*, 459–473.
200. J. Piatigorsky, *Gene Sharing and Evolution*; Harvard University Press: Boston, 2007.
201. M. Nei, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12235–12242.
202. R. Wroe; H. S. Chan; E. Bornberg-Bauer, *HFSP J.* **2007**, *1*, 79–87.
203. J. D. Bloom; Z. Lu; D. Chen; A. Raval; O. S. Venturelli; F. H. Arnold, *BMC Biol.* **2007**, *5*, 29.
204. A. Force; M. Lynch; F. B. Pickett; A. Amores; Y. L. Yan; J. Postlethwait, *Genetics* **1999**, *151*, 1531–1545.
205. M. Lynch; A. Force, *Genetics* **2000**, *154*, 459–473.

Biographical Sketches

Olga Khersonsky is a Ph.D. student in the group of Professor Dan S. Tawfik at the Department of Biological Chemistry in the Weizmann Institute of Science. She received her B.Sc. in chemistry from the Hebrew University of Jerusalem. Olga Khersonsky is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities.

Dan S. Tawfik is at the Department of Biological Chemistry in the Weizmann Institute of Science. He received his B.Sc. and M.Sc. degrees in chemistry and biochemistry from the Hebrew University of Jerusalem, and his Ph.D. from the Weizmann Institute of Science. He was a postdoctoral fellow and group leader at Cambridge University, and the MRC Centre for Protein Engineering, under the directorship of Professor Sir Alan Fersht. His research interests include protein evolution, mechanistic enzymology, and enzyme engineering.

8.04 Mechanistic and Structural Studies of Microbial Dehalogenases: How Nature Cleaves a Carbon–Halogen Bond

Gerrit J. Poelarends, University of Groningen, Groningen, The Netherlands

Christian P. Whitman, The University of Texas, Austin, TX, USA

8.04.1 Introduction

Halogenated organic compounds are ubiquitous in the environment and result from a combination of natural processes and man-made efforts. In nature, organohalogens are produced in the course of various abiotic events (e.g., biomass fires, volcanoes, and other geothermal occurrences) and by many living organisms. For example, terrestrial plants, fungi, lichen, bacteria, insects, some higher animals, and even humans account for a diverse collection of organohalogens.[1] The more recent explorations of deeper levels of the oceans have yielded a large number of new and unusual organohalogens, which are produced by marine plants, animals, and bacteria.[2] It is estimated that more than 3800 organohalogen compounds, largely containing chlorine or bromine but a few with iodine and fluorine, are produced by living organisms or by natural abiotic processes.[1,2]

Halogenated compounds are not, however, limited to those of natural origin. The more insidious ones have been introduced into the environment by their rampant use in industry and agriculture over the past two centuries.[3] These synthetic chemicals generally persist and accumulate in the environment and can be highly toxic, mutagenic, carcinogenic, or otherwise harmful and destructive. Notorious examples include polychlorinated dioxins, polychlorinated biphenyls (PCBs), dichlorodiphenyltrichloroethanes (DDTs), chlorofluorocarbons (CFCs), polychlorinated organic solvents such as the tri- and tetrachloroethenes, 1,2-dichloroethane, 1,2-dibromoethane, and vinyl chloride. The undesirable properties (and much negative publicity) have caused great public alarm about any continued use of these synthetic compounds so that many of them have now been banned and replaced with environmentally less harmful chemicals.[3]

Bacteria have enormous nutritional versatility and can seemingly use the most recalcitrant of compounds (including synthetic organohalogens) as carbon sources. Interest in co-opting the catabolic potential of bacteria and channelling it into a means to rid the environment of organohalogen compounds (i.e., bioremediation) has fueled research into the genes and enzymes responsible for the degradation of these compounds. Several

synthetic organohalogens such as chloroform, 1,2-dichloropropane, and 1,2,3-trichloropropane have limited water solubility (ranging from 1.7 to 8.2 g l^{-1}) but nonetheless are bioavailable and could serve as potential growth substrates. However, no organisms have been reported that oxidatively degrade and use these compounds as carbon sources. This recalcitrance is mainly due to a lack of enzymes that can carry out critical steps in a catabolic pathway, rendering these organohalogens xenobiotic.[4] For several other synthetic organohalogens, however, degradative organisms have been isolated, in most cases from contaminated soil, water, sediment, or sewage sludge samples by selective enrichment for growth on the pollutant as sole source of carbon and energy. Apparently, these organisms have assembled functional catabolic pathways, most likely by the recruitment and adaptation of enzymes from pathways for naturally occurring (halogenated) compounds.[4–6]

Thus far, bioremediation has met with limited success, but the efforts to delineate the molecular basis for the nutritional versatility of bacteria towards synthetic organohalogens have uncovered a treasure trove of mechanistic, structural, and evolutionary questions. These questions revolve around two fundamental issues: how do these enzymes cleave a carbon–halogen bond with such efficiency and how did they evolve in such a short period of time. Breaking a carbon–halogen bond is an inherently difficult reaction due to the high bond dissociation energies and nature has provided us with different examples of strategies used for the enzymatic cleavage of aliphatic, alkene, and aromatic halogen bonds.[4,6] Moreover, because many of these compounds have only been recently synthesized and introduced into the environment, the enzymes catalyzing their degradation and the associated pathways are not always fully optimized and can provide snapshots of evolution in action.[4–7]

In this review we focus on the different catalytic strategies that are used by bacterial dehalogenases to cleave the carbon–halogen bond. Until recently, only three bacterial dehalogenases had been studied by X-ray crystallography. These three enzymes, haloalkane dehalogenase, haloacid dehalogenase (HAD), and 4-chlorobenzoyl-coenzyme A (4-CBA-CoA) dehalogenase, cleave the carbon–halogen bond by making use of nucleophilic substitution mechanisms that proceed via a covalent ester intermediate. In the last few years, the three-dimensional structures of three other bacterial dehalogenases, haloalcohol dehalogenase, *trans*-3-chloroacrylic acid dehalogenase (CaaD) and *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD), have revealed the details of two other elegant catalytic strategies. These strategies exploit fundamentally different dehalogenation mechanisms that do not involve the formation of a covalent ester intermediate.

8.04.2   Dehalogenation Via Covalent Catalysis

8.04.2.1   Haloalkane Dehalogenases

Haloalkane dehalogenases catalyze the hydrolytic cleavage of the carbon–halogen bond in halogenated aliphatic hydrocarbons, yielding the corresponding alcohols and hydrogen halides. The first haloalkane dehalogenase was described in the early 1980s. It was discovered in *Xanthobacter autotrophicus* GJ10, a 1,2-dichloroethane-degrading bacterium isolated in the Netherlands (**Figure 1(a)**).[8–10] The same enzyme (DhlA) has now been detected in at least 12 other bacterial strains that were grown in media enriched with either 1,2-dichloroethane or 2-chloroethylvinyl ether.[11,12] In fact, DhlA is the only known haloalkane dehalogenase that functions in 1,2-dichloroethane degrading bacteria, no variants have been described, and the enzyme has been exclusively detected in Gram-negative bacteria.

Several related haloalkane dehalogenases have been cloned and characterized, both from Gram-positive and Gram-negative haloalkane degraders. The two best studied examples are DhaA from *Pseudomonas pavonaceae* 170 and LinB from *Sphingomonas paucimobilis* UT26, which play key roles in catabolic pathways for the pesticides 1,3-dichloropropene and γ-hexachlorocyclohexane (γ-HCH), respectively (**Figures 1(b) and 1(c)**).[13–15] DhaA or close sequence variants have also been found as part of catabolic pathways for 1-chlorobutane, 1-chlorohexane, 1,6-dichlorohexane, and 1,2-dibromoethane in *Rhodococcus erythropolis* and *Mycobacterium* strains.[16–18] Together, the haloalkane dehalogenases degrade a broad range of substrates, which includes chlorinated, brominated, and some iodinated haloalkanes, with a preference for relatively short substrates with primary carbon–halogen bonds.[19–22] Several haloalcohols, halogenated amides and haloethers can also be processed. There is no evidence for the conversion of fluorinated substrates. The most unusual substrates are two chlorinated cyclohexadienes formed during the catabolism of γ-HCH by *S. paucimobilis* UT26 (**Figure 1(c)**).[15]
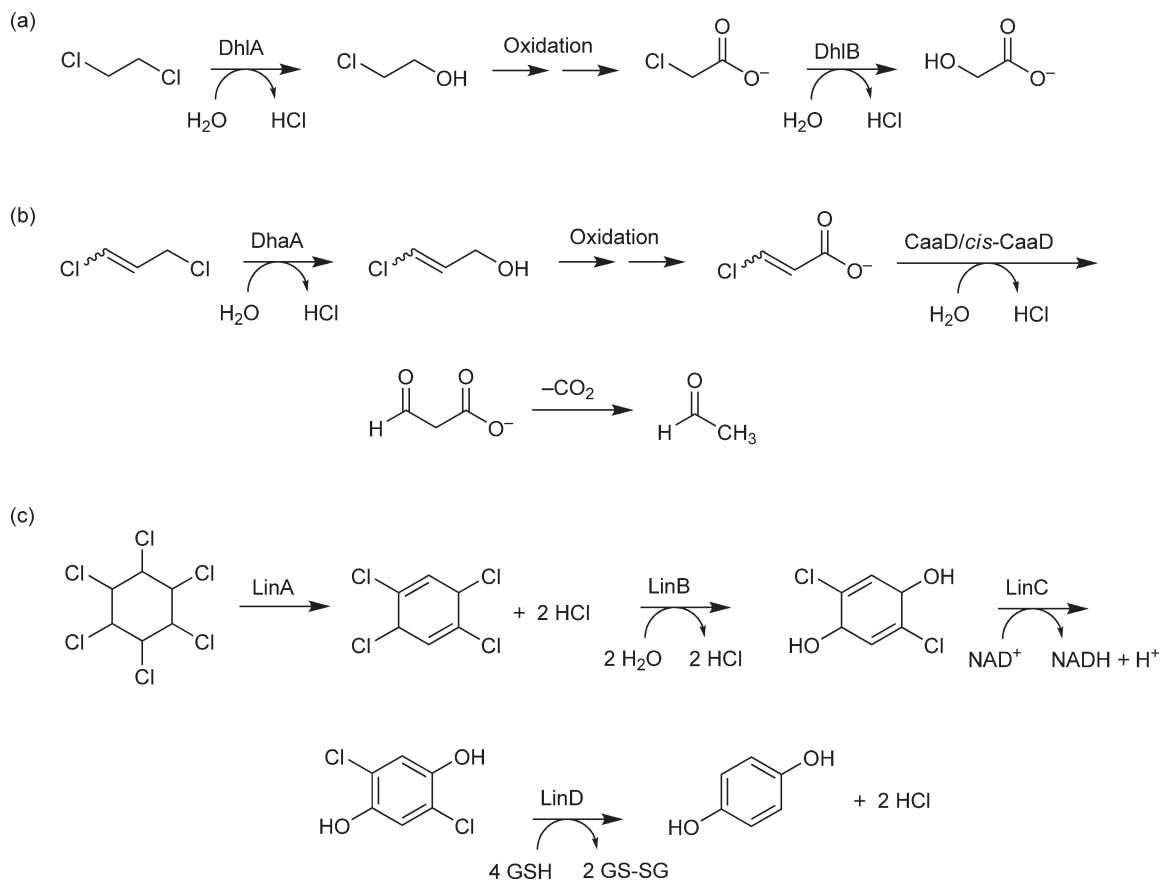
Figure 1   Three pathways for degradation of halogenated compounds. (a) The 1,2-dichloroethane catabolic pathway in *Xanthobacter autotrophicus* GJ10. (b) The 1,3-dichloropropene catabolic pathway in *Pseudomonas pavonaceae* 170. (c) The γ-HCH catabolic pathway in *Sphingomonas paucimobilis* UT26.

For quite some time, haloalkane dehalogenases were thought to be present only in soil bacteria that colonize contaminated environments. However, microbial genome and metagenome sequencing projects have revealed that there is a large diversity of protein sequences that based on sequence similarities to known haloalkane dehalogenases might be classified as putative haloalkane dehalogenases. Janssen *et al.*[4] compared the protein sequences of the three well-characterized haloalkane dehalogenases DhlA, DhaA, and LinB with the whole NCBI microbial databases and the environmental Sargasso Sea proteins. They found roughly 50 homologues in microbial genomes and more than 130 homologues in the Sargasso Sea database. Pairwise identities between DhlA, DhaA, or LinB and these identified haloalkane dehalogenase homologues range from 21 to 69%. Experimental confirmation of the dehalogenase activity of these putative haloalkane dehalogenases has not been reported, but they may have novel substrate and reaction specificities that could be valuable for future biotransformations. Recently, it was shown that putative haloalkane dehalogenase genes in the genomes of *Mycobacterium avium* N85, *Mycobacterium bovis* 5033/66, *Mesorhizobium loti* MAFF303099, and *Bradyrhizobium japonicum* USDA110 indeed code for functional haloalkane dehalogenases.[23–25] These strains have not been reported as haloalkane degraders. This raises questions about the physiological substrates and metabolic roles of these dehalogenating enzymes.

Much of the initial insight into the catalytic mechanism of haloalkane dehalogenases has come from X-ray structure work on the *Xanthobacter autotrophicus* enzyme (DhlA) (**Figure 2**).[26] (For details on the structures and mechanisms of the two other well-characterized haloalkane dehalogenases, DhaA and LinB, the reader is referred to a recent review that compares the three haloalkane dehalogenase enzymes.[27]) One striking example is the knowledge gained from soaking dehalogenase crystals with the substrate at different pH values and
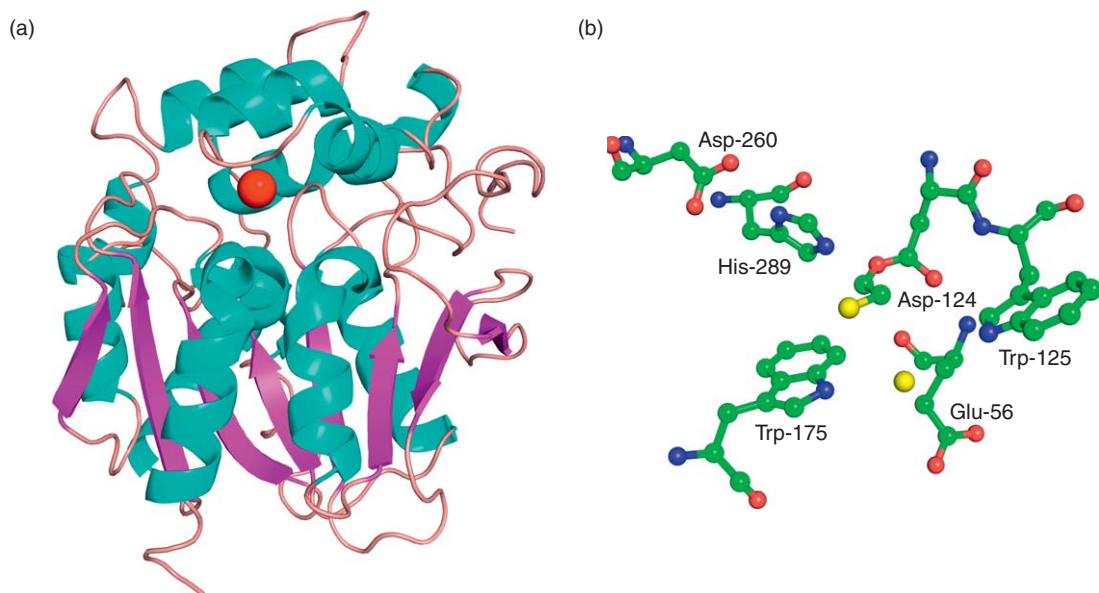
Figure 2 (a) Ribbon diagram of the monomeric structure of DhlA. For clarity, α-helices are shown in cyan and β-strands in purple. The location of the active site is indicated by the bound chloride ion, which is shown as a red sphere. (b) A close-up of the active site of DhlA where Asp-124 is covalently modified by a chloroethyl group, which results from the nucleophilic attack of Asp-124 on 1,2-dichloroethane (**Figure 3**). For clarity, both the chlorine atoms of the chloroethyl group and the displaced chloride ion are shown in yellow. The roles of the key active site residues (Glu-56, Asp-124, Trp-125, Trp-175, Asp-260, and His-289) and their interactions are discussed in the text. The figure was prepared with PyMOL (W. L. DeLano, *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, USA, 2002, http://www.pymol.org.).

solving the X-ray structures of the intermediates that accumulated.[28] DhlA crystals soaked in mother liquor containing $10 \, \text{mmol} \, \text{l}^{-1}$ 1,2-dichloroethane at pH 5 showed electron density consistent with the presence of both a covalent alkyl–enzyme intermediate and a chloride ion at the active site (**Figure 2(b)**). Crystals soaked at pH 6.2 showed only chloride ion in the active site, suggesting that the alkyl–enzyme intermediate could be hydrolyzed at a higher pH and that the 2-chloroethanol product is lost from the active site. Hence, these structures identified the active site, suggested interactions that might be responsible for the specific binding of substrate, and led to a working hypothesis for the catalytic mechanism.

The DhlA enzyme functions as a monomer ($\sim$35 kDa) and is composed of two domains: a main domain and a cap domain (**Figure 2(a)**).[26] The main domain consists of a mostly parallel eight-stranded β-sheet connected by α-helices on both sides of the sheet. The cap domain is composed of five α-helices with intervening loops. The active site is an occluded hydrophobic cavity located at the interface of the two domains. The overall fold of the main domain is the hallmark of the α/β-hydrolase fold superfamily of enzymes, to which lipases, esterases, carboxypeptidases, and acetylcholinesterases also belong.[26,28] These superfamily members catalyze the hydrolysis of ester and amide bonds via a two-step nucleophilic substitution mechanism similar to that of serine proteases.

DhlA uses a very similar two-step catalytic mechanism. The major players in catalysis are the three residues (Asp-124, His-289, and Asp-260) that form the catalytic triad, functionally similar to the classical catalytic triad in serine proteases, which is provided by the main domain.[28] The first step in catalysis is the nucleophilic attack of one carboxylate oxygen of Asp-124 on the halogen-bearing carbon atom to form a covalently bound ester intermediate via an S_N2 substitution mechanism (**Figures 2(b) and 3**). This step is facilitated by two tryptophan residues (Trp-125 and Trp-175) that form a halide binding site, and are involved in leaving group stabilization during the cleavage process. In the second step, the covalent alkyl–enzyme intermediate is hydrolyzed by a water molecule that is activated by His-289, yielding the corresponding alcohol and halide ion (**Figure 3**). Activation of the catalytic water molecule is assisted by Asp-260, which stabilizes the positive charge that develops on His-289. The charge that develops on the carbonyl oxygen of the covalent ester intermediate is
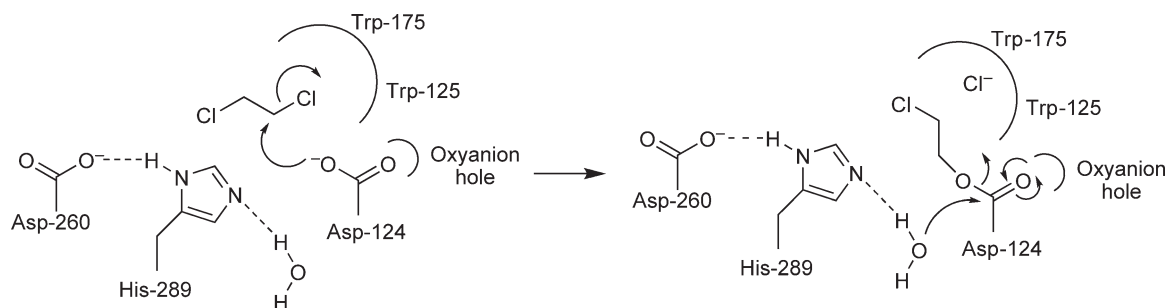
Figure 3 A schematic representation of the catalytic mechanism of haloalkane dehalogenase (DhlA).

stabilized through hydrogen bond interactions with the main chain amides of Glu-56 and Trp-125, which form an oxyanion hole. The last step in catalysis is the release of halide from the active site.

This proposed mechanism of covalent catalysis is supported by a plethora of experimental evidence. Incubation of the wild-type enzyme with a large excess of 1,2-dichloroethane in the presence of $H_2^{18}O$ resulted in the incorporation of ^{18}O in 2-chloroethanol and in the carboxylate group of Asp-124, whereas there is negligible incorporation of the label in the enzyme in the absence of substrate.[29] The results of this multiple turnover experiment are consistent with a reaction that proceeds via a nucleophilic attack of Asp-124 on the substrate molecule with the formation of an alkyl–enzyme intermediate and subsequent attack of solvent water on the carbonyl carbon of Asp-124. The role of Asp-124 was further investigated by analyzing the kinetic properties of three site-directed mutants (D124G, D124A, and D124E).[29] From these studies, it can be concluded that Asp-124 is essential for catalysis. Replacing this residue with a glycine, an alanine, or a glutamate results in no detectable activity. Additional evidence for covalent catalysis and the role of Asp-124 were derived from studies with the H289Q mutant.[30] This mutant was not catalytically active, but a halide ion burst stoichiometric to the amount of enzyme was observed upon incubation with the substrate. Using electrospray ionization mass spectrometry, accumulation of the covalent alkyl–enzyme species and binding of the alkyl moiety of the substrate to an Asp-124-containing tryptic peptide were demonstrated. From these studies it can be concluded that His-289 is the base catalyst for hydrolysis of the covalent ester intermediate and that Asp-124 is the nucleophile. There is an additional interesting aspect of the covalently trapped enzyme that deserves comment. The fluorescence of the two active site tryptophans is quenched when substrate or halide ion is bound by the enzyme.[31] This intrinsic fluorescence-quenching phenomenon was used to measure a dissociation constant for halide binding, and the experiments indicated that halide ions are strongly bound by the alkyl–enzyme species but not by the substrate-free enzyme.[30] This finding is consistent with the crystallographic observations and suggests that the halide ion released in the first catalytic step likely leaves the active site only after hydrolysis of the alkyl–enzyme intermediate.

The roles of Asp260, the third member of the catalytic triad, and the two tryptophans in the mechanism were further investigated by analyzing the properties of site-specific mutants (D260N, W125F, W125Q, W125R, W175Q, W175Y).[32–34] From these studies it can be concluded that Asp-260 is essential for catalysis. Replacing this residue with an asparagine results in no detectable activity.[32] Interestingly, the activity for brominated substrates was restored to the inactive D260N mutant by replacing Asn-148 with an aspartic or glutamic acid. Molecular modeling of these mutants showed that the repositioned catalytic triad member could indeed take over the interaction with His-289.[32] The two tryptophans are also important for catalysis. All tryptophan mutants, except W125F and W175Y, showed about a 10-fold reduced k_{cat} and much higher K_m values with 1,2-dichloroethane and 1,2-dibromoethane when compared to the wild-type enzyme.[33,34] In addition, fluorescence quenching experiments showed a decrease in the affinity of the mutant enzymes for halide ions. Furthermore, the ^2H kinetic isotope effect observed with the wild-type enzyme in deuterium oxide was lost in some of these tryptophan mutants, indicating that in these cases the rate-limiting step has shifted to a step before hydrolysis of the covalent alkyl–enzyme intermediate.[33,34] Collectively, these findings indicate that both tryptophans are involved in substrate and halide ion binding, and in stabilizing the transition state during the nucleophilic substitution step that causes carbon–halogen bond cleavage.

Pre-steady-state kinetic studies have unraveled further details of the catalytic mechanism.[35,36] Rapid quench flow experiments and stopped-flow fluorescence measurements indicated that the rate of substrate binding is fast ($>700\,s^{-1}$ at $5\,mmol\,l^{-1}$ 1,2-dibromoethane), whereas the rates of the chemical steps are quite slow (>130 and $\sim10\,s^{-1}$ for carbon–bromine bond cleavage and hydrolysis of the alkyl–enzyme intermediate, respectively). However, the rate of halide ion release (i.e., product release) turns out to be the last and slowest step in the catalytic cycle ($\sim4\,s^{-1}$ for bromide). The kinetics of halide ion binding and release further indicated that a slow enzyme isomerization step limited the overall rate of halide ion release. Janssen and co-workers proposed that this step involved a conformational change in the cap domain that is necessary to allow water to enter the normally occluded active site and solvate the halide ion.[35,36] This hypothesis is supported by kinetic and crystallographic studies on DhlA variants with cap domain mutations and a thermodynamic analysis of halide binding to wild-type DhlA.[37,38] The results of these studies are consistent with the occurrence of conformational changes upon bromide binding and release. Solution structures of DhlA with and without halide ions bound at the active site would provide further insight into the precise location and mechanism of these conformational changes.

As the first dehalogenase crystallized and studied in detail, the DhlA-catalyzed reaction became a paradigm for enzymatic carbon–halogen bond cleavage.[27] More recently, the structures of other dehalogenases have been solved and their catalytic mechanisms investigated. The cumulative body shows the diversity in catalytic strategies for cleaving the carbon–halogen bond. These studies are discussed below.

8.04.2.2 Haloacid Dehalogenases

HADs (also referred to as haloalkanoate dehalogenases or haloacetate dehalogenases) catalyze the hydrolysis of α-halogenated carboxylic acids to yield the corresponding α-hydroxycarboxylic acids and hydrogen halides. These enzymes are found in many strains of soil bacteria that are able to grow on halogenated aliphatic acids (**Figure 1**(**a**)), as well as in strains that are not known to degrade halogenated compounds.[4,39–43] Their roles in these latter strains are unknown. The widespread occurrence of these enzymes and the diversity in sequences may reflect the natural occurrence of haloacids in nature. The HAD genes are probably of ancient evolutionary origin and their widespread distribution may not be due to environmental contamination.

It is noteworthy that bacterial strains capable of utilizing simple haloacids such as 2-chloroacetate and 2-chloropropionate can be readily isolated from almost any soil sample. In contrast, synthetic organohalogens such as 1,2-dichloro- and 1,2-dibromoethane are much more difficult to biodegrade, and degradative organisms can only be isolated after prolonged adaptation or if a suitable inoculum is used in which the degradative activity has been pre-enriched from exposure to the synthetic organohalogen in the natural environment.[4] The products of the hydrolytic dehalogenation of haloacids are readily metabolizable hydroxyacids, so the presence of a single enzyme that recognizes and converts the organohalogen allows the microorganisms to take advantage of a new carbon source without the need to assemble a whole new catabolic pathway.

The HADs have been divided into two families on the basis of sequence similarities and are called group I and group II enzymes. (This classification into two families, which is based on molecular data, replaces previous classifications that focused on arbitrary characteristics such as substrate specificity and stereospecific action on 2-chloropropionic acid.[44]) The two families appear to be evolutionary unrelated and together represent almost all of the HADs described to date.[44] The group II enzymes are well characterized and define the so-called HAD superfamily, to which the magnesium-dependent phosphatases and P-type ATPases also belong.[44–46] L-2-HAD from *Pseudomonas* sp. YL (L-DEX)[45] and L-2-HAD from *X. autotrophicus* GJ10 (DhlB; **Figure 1**(**a**)),[46] two representative group II members whose crystal structures recently became available, will be discussed here. Both L-DEX and DhlB act on the L-isomers of their substrates, yielding products with inversion of configuration at the chiral C-2 carbon atom.

Sequence analysis and site-directed mutagenesis provided the first clues about the L-DEX mechanism.[47] A multiple sequence alignment of seven related L-2-HADs (36–70% sequence identity) from different bacterial strains was used to identify highly conserved residues. Site-directed mutagenesis of all the conserved charged and polar residues in L-DEX (36 out of 232 amino acids) identified several residues that were required for catalytic activity. Among these were two possible candidates for an active-site carboxylate nucleophile (Asp-10 and Asp-180) in a mechanism analogous to that catalyzed by haloalkane dehalogenase.[47]

The possibility of a catalytic mechanism that proceeds via a covalent alkyl–enzyme intermediate was investigated using two different techniques. First, ^{18}O-isotope labeling experiments similar to those described for haloalkane dehalogenase were performed.[48] Under multiple turnover conditions in H$_2^{18}$O, with a large excess of substrate over enzyme, the D-lactate produced from L-2-chloropropionate contained the ^{18}O label. However, under single turnover conditions in H$_2^{18}$O, using enzyme in excess over substrate, the product does not contain ^{18}O label, suggesting that an oxygen atom of the solvent water is first incorporated into the enzyme and then transferred to the product. To determine the site of incorporation, mass spectrometric analysis of the enzyme that had undergone multiple turnovers in H$_2^{18}$O was performed. This analysis revealed that Asp-10 was labeled with two ^{18}O atoms whereas no labeling of Asp-180 was detected. Hence, it was concluded that Asp-10 (corresponding to Asp-8 in DhlB) acts as the carboxylate nucleophile that attacks the α-carbon atom of the substrate to yield an ester intermediate, which is hydrolyzed by nucleophilic attack of a water molecule on the carbonyl carbon atom (**Figure 4**).[48]

Additional support for the mechanism came from a second approach involving chemical modification experiments with hydroxylamine.[49] Incubation of L-DEX with hydroxylamine in the presence of substrate led to the alkylation of Asp-10 and the concomitant loss of catalytic activity. The absence of substrate protected the enzyme from modification and inactivation by hydroxylamine. These findings are consistent with hydroxylamine attack on the carbonyl carbon of the ester intermediate, which results in the formation of catalytically inactive adducts at the site of the aspartate nucleophile.[49]

Extensive X-ray structure work has been carried out on both L-DEX and DhlB.[45,46,50,51] The enzymes are homodimers with two or three domains per subunit (**Figure 5(a)**). Both have a core domain with a Rossmann-fold-like six-stranded parallel β-sheet flanked by five α-helices and a subdomain inserted into the core domain consisting of a four-helix bundle. In DhlB but not in L-DEX, a dimerization domain of two antiparallel α-helices is present (**Figure 5(a)**). The active site is located between the core domain and the subdomain. The use of active site mutants or low pH (to slow down the reaction rate) allowed the determination of a series of crystal structures that captured reaction intermediates in the dehalogenation of haloalkanoates by L-DEX or DhlB (**Figure 5(b)**).[50,51] These structures provided detailed insight into the reaction mechanisms of the two enzymes. In the first step of the reaction, one carboxylate oxygen of Asp-8 (DhlB numbering) attacks the halogen-bearing C2 atom of the substrate to form a covalent enzyme–ester intermediate (**Figures 4 and 5(b)**). The interactions between the halogen atom and the side chains of Arg-39, Asn-115, and Phe-175, which form a halide-stabilizing cradle, facilitate this step by labilizing the carbon–halogen bond. In the next step of the reaction, the ester bond is hydrolyzed by nucleophilic attack of a water molecule on the Cγ atom of Asp-8, yielding the product and free enzyme (**Figure 4**). The negative charge that develops on the carbonyl oxygen atom of the ester intermediate is stabilized by an oxyanion hole formed by the side chain atoms from Thr-12, Asn-173, and Ser-171. This mechanism is analogous to that of haloalkane dehalogenases (**Figure 3**), but there is no histidine in the active site to activate a water molecule for nucleophilic attack. How the water molecule is activated is not known, but it has been suggested that another active site aspartate is involved.[51]

Intriguingly, HADs employ at least one other dehalogenating strategy. The reaction catalyzed by the DL-2-HAD from *Pseudomonas* sp. 113 (DL-DEX 113) appears to proceed without the formation of a covalent ester intermediate.[52,53] The results of ^{18}O-labeling studies indicate that a solvent water molecule directly attacks the
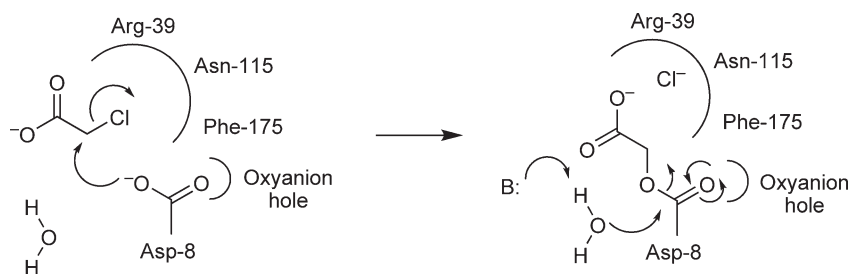
Figure 4 A schematic representation of the catalytic mechanism of haloacid dehalogenase (DhlB). For consistency with **Figure 5**, the mechanism for DhlB is shown. A similar mechanism has been proposed for L-DEX, in which residue Asp-8 in DhlB corresponds to Asp-10 in L-DEX.[48]
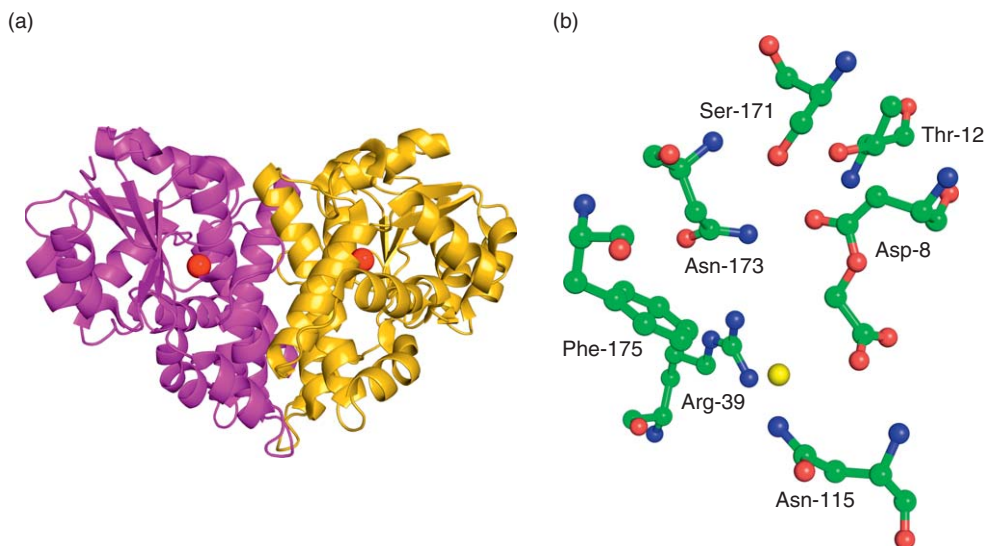
Figure 5 (a) Ribbon diagram of the homodimeric structure of DhlB. For clarity, one monomer is shown in purple and the other one in gold. The location of the active site in each monomer is indicated by the bound chloride ion, which is shown as a red sphere. (b) A close-up of the active site of DhlB where Asp-8 is covalently modified by acetate, which results from the nucleophilic attack of Asp-8 on 2-chloroacetate (**Figure 4**). The displaced chloride ion is shown as a yellow sphere. The roles of the key active site residues (Asp-8, Thr-12, Arg-39, Asn-115, Ser-171, Asn-173, and Phe-175) and their interactions are discussed in the text. The figure was prepared with PyMOL (W. L. DeLano, *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, USA, 2002, http://www.pymol.org.).

α-carbon atom of the substrate (2-haloalkanoic acid) to displace the halogen atom.[53] This observation suggests that the active site has a mechanism to activate the water and/or substrate molecule. Unfortunately, crystallographic data that could provide structural insight into the mode of action of this group I enzyme is still unavailable.

Fluoroacetate dehalogenase (DehH1) from *Moraxella* sp. B is another interesting enzyme. DehH1 does not belong to either group I or group II HADs, and is different from other HADs in that it is a defluorinating enzyme.[44,54] It shows high activity with 2-fluoroacetate but much lower activity with the 2-bromo- or 2-chloro-analogues. Initially it was thought that DehH1 might represent a third HAD family, but secondary structure predictions and a multiple sequence alignment of DehH1 with haloalkane dehalogenase DhlA (18% sequence identity) and a number of epoxide hydrolases revealed that DehHI is related to members of the α/β-hydrolase fold superfamily.[55] The relatedness between DehH1 and DhlA is apparent from short stretches of sequence that they have in common, including the conservation of Asp-124 in DhlA, which is the active site nucleophile, and His-289, which is the water-activating base, as Asp-105 and His-272 in DehH1.

The results of ^{18}O-labeling studies indicate that Asp-105 acts as the carboxylate nucleophile that attacks the α-carbon atom of the substrate to yield an ester intermediate, which is subsequently hydrolyzed by nucleophilic attack of a water molecule on the carbonyl carbon atom.[56] Site-directed mutagenesis experiments have confirmed that Asp-105 is essential for catalytic activity.[56] The catalytically inactive H272N mutant provided additional evidence for the existence of the ester intermediate. When the H272N enzyme is incubated with substrate, ion-spray mass spectrometry shows that the mutant enzyme contains the alkyl moiety of the substrate attached to the tryptic peptide containing Asp-105.[56] From these studies it can be concluded that Asp-105 is the nucleophile and His-272 is the base catalyst for hydrolysis of the covalent ester intermediate in a mechanism anologous to that of haloalkane dehalogenase. Further insight into the evolutionary relationships of DehH1 and into the mechanistic details of its ability to cleave the strong carbon–fluorine bond awaits the determination of the three-dimensional structure of this fascinating enzyme.

8.04.2.3   4-Chlorobenzoyl-CoA Dehalogenases

4-Chlorobenzoyl-CoA dehalogenase catalyzes the hydrolytic dehalogenation of 4-CBA-CoA to yield 4-hydroxybenzoate-CoA (4-HBA-CoA). This enzyme has been discovered in a number of soil bacteria where, along with 4-CBA-CoA ligase and 4-HBA-CoA thioesterase, it forms a fascinating three-step dehalogenation pathway for the conversion of 4-chlorobenzoate into 4-hydroxybenzoate (**Figure 6**(a)).[57–60] The 4-hydroxybenzoate is now poised for degradation by the central oxidative aromatic pathways. Such a multistep dehalogenation pathway is also involved in the catabolism of *cis-* and *trans-*3-chlorocrotonate in *Alcaligenes* sp. strain CC1 (**Figure 6**(b)).[61] The need for a multistep dehalogenation pathway for these substrates may reflect the fact that the carbon–halogen bond between a halogen and an arenic or vinylic carbon atom is much more difficult to cleave than the one between a halogen and an sp^3-hybridized carbon atom. Therefore, the substrate is first activated by conjugation to CoA, after which the halogen can be displaced by a nucleophilic substitution mechanism. In this context, the reactions catalyzed by the *cis*-CaaD and CaaD are of particular interest because these enzymes dehalogenate a vinylic substrate in one step without the need for substrate activation by CoA conjugation (**Figure 6**(c); see below).[62]

The three-dimensional structure of 4-CBA-CoA dehalogenase from *Pseudomonas* sp. strain CBS3 (in complex with 4-HBA-CoA) has been solved (**Figure 7**).[63] The enzyme functions as a homotrimer in which each subunit folds into two domains. The large N-terminal domain is characterized by a 10-stranded β-sheet, forming two nearly perpendicular layers, which are flanked by α-helices. The small C-terminal domain is composed of three amphiphilic α-helices, extends away from the body of the molecule, and is primarily involved in trimerization. The two domains of each subunit are linked together by a cation, presumably a calcium ion (**Figure 7(a)**).

The structure of the 4-CBA-CoA dehalogenase-4-HBA-CoA complex identified the active site and suggested interactions that might be responsible for the binding and activation of the substrate (**Figure 7**(b)).[63] In the structure of the complex, the benzoyl ring is surrounded by the aromatic rings of Phe-64, Phe-82, Trp-89, and Trp-137, while the 4-hydroxyl group of the benzoyl ring forms a hydrogen bond with the carboxylate group of Asp-145. The thioester carbonyl group forms hydrogen-bonding interactions with the backbone amide protons of Phe-64 and Gly-114 as well as an interaction with the positive dipole of the α-helix formed by residues 114–125. The phosphoryl groups of the CoA unit form ion pairs with the charged side chains of Arg-24, Arg-257, and Arg-67, while the adenine C-6 amino group hydrogen bonds with the backbone carbonyl oxygen of Phe-64.[63]
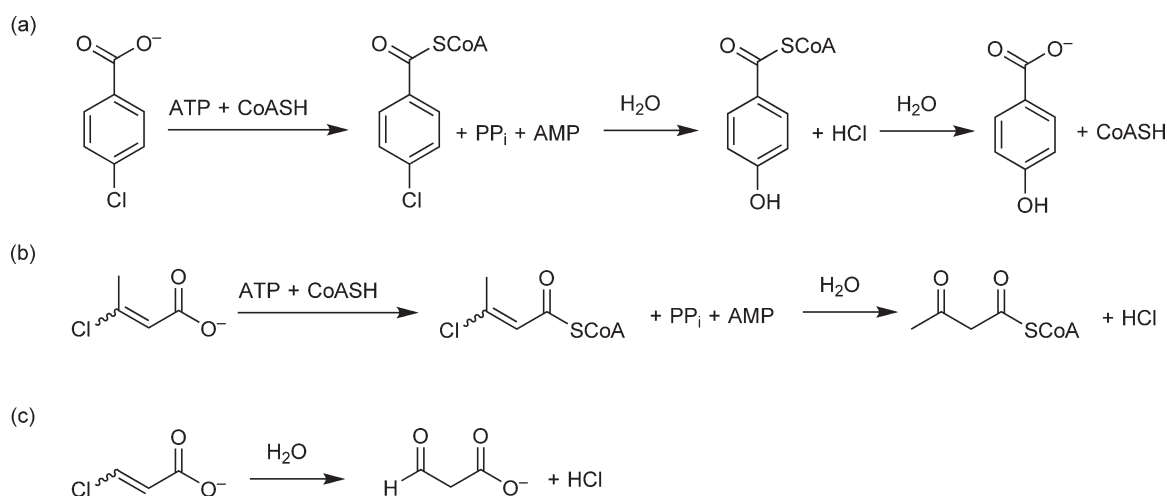
Figure 6   Dehalogenation pathways for arene and vinyl halides. (a) The three-step dehalogenation pathway for conversion of 4-chlorobenzoate into 4-hydroxybenzoate in *Pseudomonas* sp. strain CBS3. (b) The multistep dehalogenation pathway for the catabolism of *cis-* and *trans-*3-chlorocrotonate in *Alcaligenes* sp. strain CC1. (c) The one-step dehalogenation of *cis-* and *trans-*3-chloroacrylate by the *cis*-CaaD and CaaD.

(a)   (b)
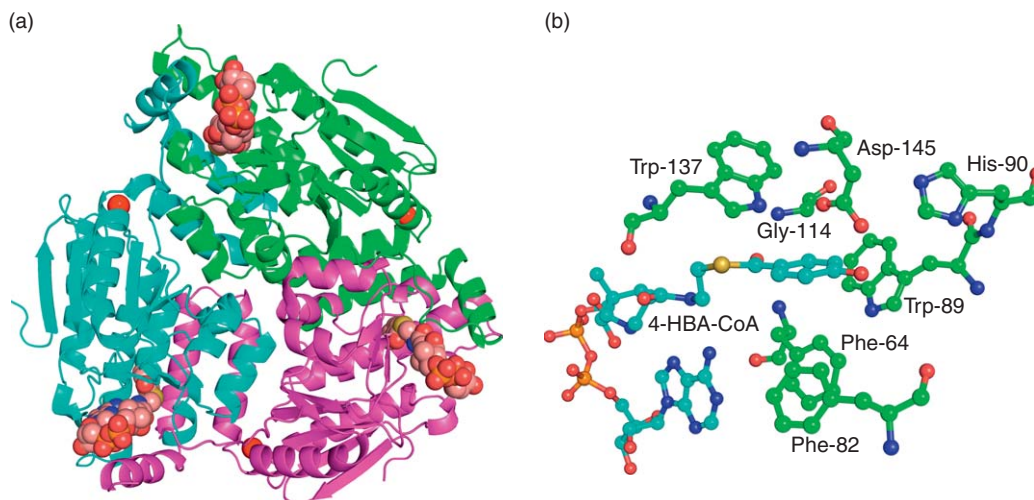
Figure 7   (a) Ribbon diagram of the homotrimeric structure of 4-CBA-CoA dehalogenase. For clarity, each monomer is shown in a different color (cyan, purple, and green). The location of the active sites is indicated by the bound 4-HBA-CoA product, which is shown in Corey–Pauling–Koltun (CPK) representation. The cations (most likely calcium ions) that link the two domains within each subunit are shown as red spheres. (b) A close-up of the active site of 4-CBA-CoA dehalogenase in complex with 4-HBA-CoA. For clarity, the carbon atoms of the active site residues are shown in green, whereas those of the 4-HBA-CoA ligand are shown in cyan. The roles of the key active site residues (Phe-64, Phe-82, Trp-89, His-90, Gly-114, Trp-137, and Asp-145) and their interactions are discussed in the text. The figure was prepared with PyMOL (W. L. DeLano, *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, USA, 2002, http://www.pymol.org.).

On the basis of the interactions observed in the enzyme–product complex and extensive mechanistic studies, a multistep catalytic mechanism for the 4-CBA-CoA dehalogenase-catalyzed reaction was proposed (**Figure 8**).[63–75] A key catalytic task for the enzyme is to activate the halogen-bearing carbon atom of the substrate for nucleophilic attack. This is accomplished by strong polarizing interactions between active site residues and the benzoyl portion of the substrate. The results of ultraviolet (UV)–visible, Raman, and NMR spectroscopic studies are consistent with a significant shift of electron density from the aromatic ring into the thioester C=O group when substrate analogues such as 4-methylbenzoyl-CoA or 4-methoxybenzoyl CoA are bound to the active site.[65–68] Major contributions come from the Gly-114 and Phe-64 backbone amide groups, which form hydrogen bonds with the substrate benzoyl C=O group. By measuring the changes in Gibbs free energy of the enzyme–substrate and enzyme–transition state complexes brought about by site-directed mutations, it was estimated that the binding interactions between the Gly-114 and Phe-64 backbone amide protons and the substrate benzoyl C=O group intensify in the rate-limiting transition state by $\sim$3.1 kcal mol^{-1}.[69] An additional polarizing contribution comes from the α-helix of residues 114–121, which provides a dipolar interaction. Collectively, these interactions comprise an oxyanion hole. The effect of both the helix dipole and the hydrogen bonds will polarize the benzoyl C=O bond, which, in turn, polarizes the electron distribution within the entire benzoyl moiety. The polarization effects are assisted by the hydrophobic residues surrounding the benzoyl group, providing a low-dielectric constant microenvironment.[63] This strong electron polarization promotes catalysis by reducing the electron density at the halogen-bearing carbon atom (inducing a partial positive charge), thereby activating this atom for nucleophilic attack.[69]

Substrate binding and activation are followed by attack of the carboxylate side chain of Asp-145 at the benzoyl C-4 atom to give an enzyme-stabilized Meisenheimer intermediate (EMc) (**Figure 8**). Indeed, a site-directed mutant in which Asp-145 has been replaced by an alanine is catalytically inactive.[70] Ketonization of the EMc results in rearomatization of the benzoyl ring and expulsion of the chloride. This nucleophilic addition–elimination mechanism (S$_N$Ar-type reaction) results in a second covalent (aryl–enzyme) intermediate, which is subsequently hydrolyzed by a water molecule that is activated by His-90 to give the free enzyme and the product.[70] The existence of a covalent aryl–enzyme intermediate has been inferred from ^{18}O-labeling studies (similar to those described for haloalkane and haloalcohol dehalogenase) and from the direct measurement of the aryl–enzyme
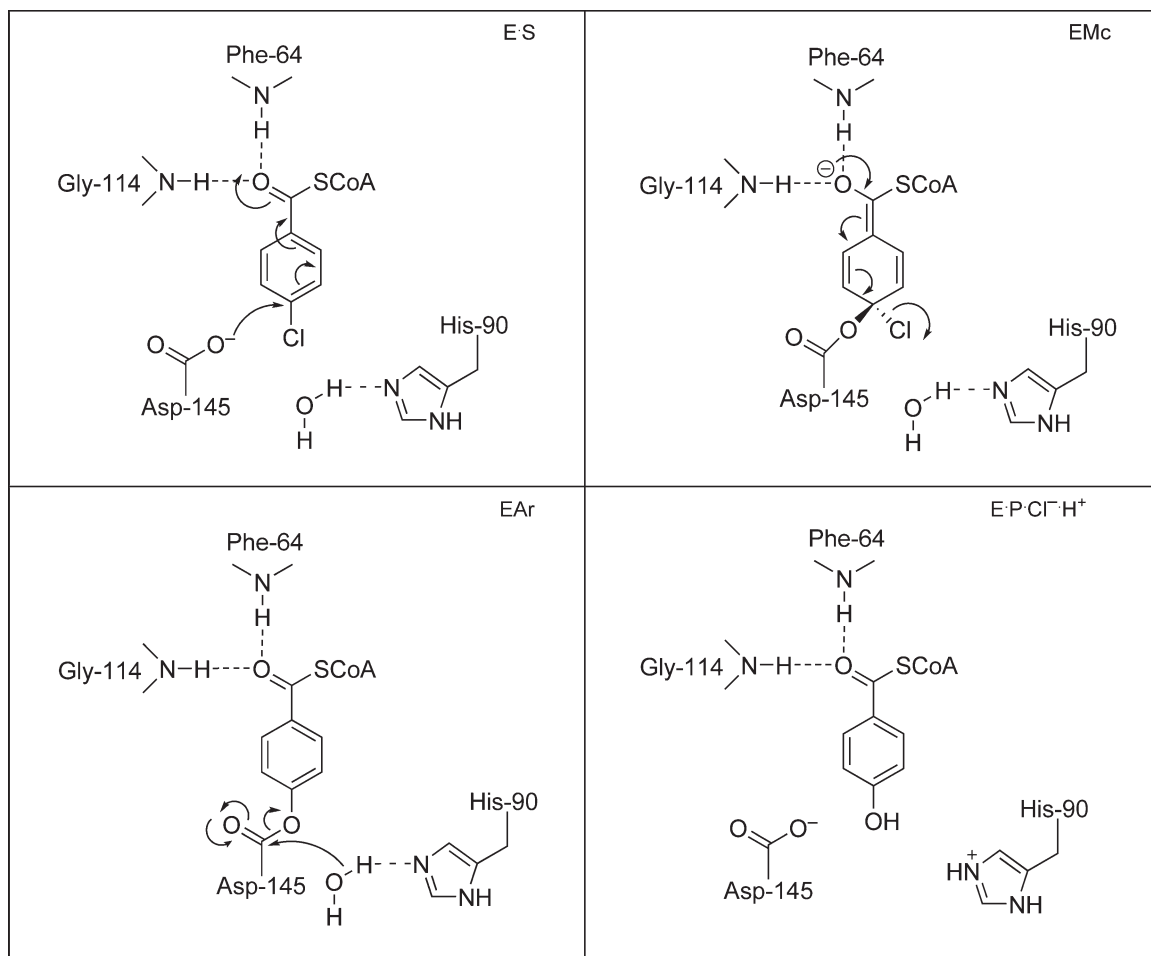
Figure 8   A schematic representation of the catalytic mechanism of 4-chlorobenzoyl-CoA dehalogenase. The catalytic residues functioning in the enzyme–substrate complex (E·S), Meisenheimer intermediate (EMc), arylated enzyme intermediate (EAr), and enzyme–product complex (E·P·Cl⁻·H⁺) are shown. The direction of the catalytic steps is as follows: from E·S to EMc to EAr and finally to E·P·Cl⁻·H⁺.

population by using $[^{14}C]$4-CBA-CoA and rapid quench techniques.[64] Results from a kinetic analysis of the H90Q mutant support the role of His-90 as general base catalyst in ester hydrolysis.[71] The indole proton of Trp-137 likely provides a hydrogen bond to the Asp-145 carbonyl group and thus might activate the Asp-145 aryl-ester and stabilize the oxyanion that is formed upon hydrolysis.[70]

Evidence for Meisenheimer complex formation comes from the very clever and sophisticated Raman spectroscopy work using substrate analogues that slowly undergo turnover.[72] With the native substrate 4-CBA-CoA, the EMc does not accumulate to a significant extent during turnover and is therefore difficult to observe. However, two alternative substrates, 4-fluorobenzoyl-CoA (4-FBA-CoA) and 4-nitrobenzoyl-CoA (4-NBA-CoA), with poor leaving groups at C-4 significantly reduce the forward partitioning rate of the Meisenheimer complex, that is, the formation of the arylated enzyme from the Meisenheimer complex is hindered. The reduced rates enabled the collection of Raman spectroscopic data that provide strong evidence for a population of Meisenheimer complexes in reaction mixtures containing the 4-FBA-CoA and 4-NBA-CoA substrates. The results of kinetic studies suggest that approximately 10–20% of the enzyme–substrate complexes in the reaction mixtures are present as Meisenheimer complex.[72] Importantly, Meisenheimer complex formation was not observed when the D145A mutant was used in reaction mixtures with 4-FBA-CoA and 4-NBA-CoA, consistent with the fact that the crucial active site nucleophile that initiates the reaction is no longer present.[72]

Mutagenesis and Raman spectroscopy studies have shown that the strength of the polarizing forces at the benzoyl carbonyl of the substrate correlate with the rate of EMc formation.[73] Thus, the polarizing forces at the thioester carbonyl are transmitted to the benzoyl C-4 position, five chemical bonds away, and correlate with the rate of aromatic nucleophilic addition at this position. Recent theoretical studies provide support for formation of the Meisenheimer complex during the substitution reaction as an intermediate, and strongly suggest that the formation of the Meisenheimer complex is the rate-limiting step, resolving a long-standing uncertainty in the (experimental) kinetic model.[74,75] These observations leave no doubt that the introduction of these polarizing interactions into the 4-CBA-CoA dehalogenase active site was an essential step in the evolution of an efficient aromatic dehalogenase.

8.04.3   Dehalogenation Via Noncovalent Catalysis

8.04.3.1   Haloalcohol Dehalogenases

Haloalcohol dehalogenases (also referred to as halohydrin hydrogen-halide lyases or halohydrin dehalogenases) catalyze the nucleophilic displacement of a halogen by a vicinal hydroxyl group in haloalcohols, yielding the corresponding epoxides and hydrogen halides.[76,77] The prototypical reaction shown in Equation (1) is the conversion of 1,3-dichloro-2-propanol to epichlorohydrin and hydrogen chloride. These enzymes are of considerable biocatalytic interest because they have broad substrate specificity and enantioselectivity toward aliphatic as well as aromatic vicinal haloalcohols.[78,79] They also efficiently catalyze the reverse reaction, the enantio- and β-regioselective epoxide ring opening by halides (chloride, bromide, and iodide).[79] Moreover, the broad nucleophile specificity of haloalcohol dehalogenases enables the use of a range of alternative nucleophiles such as azide, cyanide, and nitrite in the ring opening reaction.[80–82] These properties make haloalcohol dehalogenases promising biocatalysts for the synthesis of enantiopure epoxides and haloalcohols, as well as azido-, cyano-, and other β-substituted alcohols.[78–83] A striking example of the biocatalytic applicability of haloalcohol dehalogenase is its role in a sequential kinetic resolution of racemic 4-chloro-3-hydroxybutanoate methylester (in the presence of cyanide) to yield (S)-4-cyano-3-hydroxybutanoate methylester, a key building block for statins, which are used as cholesterol lowering agents (**Scheme 1**).[84]

$$\text{(1)}$$
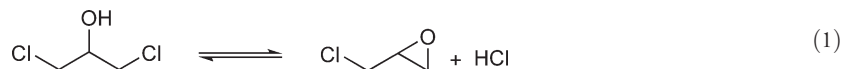
Haloalcohol dehalogenases have been found in both Gram-positive and Gram-negative bacteria, where they are part of degradation pathways for environmental pollutants such as 1,3-dichloro-2-propanol, epichlorohydrin, and 1,2-dibromoethane.[85,86] Haloalcohol dehalogenase homologues with pairwise sequence identities to known haloalcohol dehalogenases ranging from 19 to 53% have also been detected in organisms that have no known history of organohalogen degradation, although they are not as widespread as, for example, haloalkane and HAD homologues.[4] The haloalcohol dehalogenases that have been isolated and characterized can be grouped into three subtypes (A, B, and C) on the basis of sequence similarities.[86] The B-type haloalcohol dehalogenases share only 24% sequence identity with the A- and C-type haloalcohol dehalogenases, which themselves are 33% identical. On the basis of sequence and structural similarity, it has been proposed that the A- and C-type enzymes share a common ancestor, whereas the B-type haloalcohol dehalogenases have originated from a different precursor.[87] The different subtypes show considerable differences in their catalytic properties. The A- and B-type haloalcohol dehalogenases display only a modest enantiopreference whereas the C-type enzymes are highly enantioselective. The enzyme HheC[87] from *Agrobacterium radiobacter* AD1 (C-type)
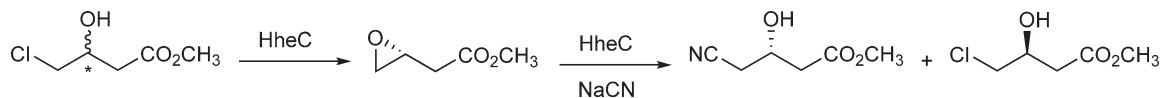
Scheme 1   The use of HheC for the enantioselective synthesis of (S)-4-cyano-3-hydroxybutanoate methylester, a key building block for statins.

(a)

(b)
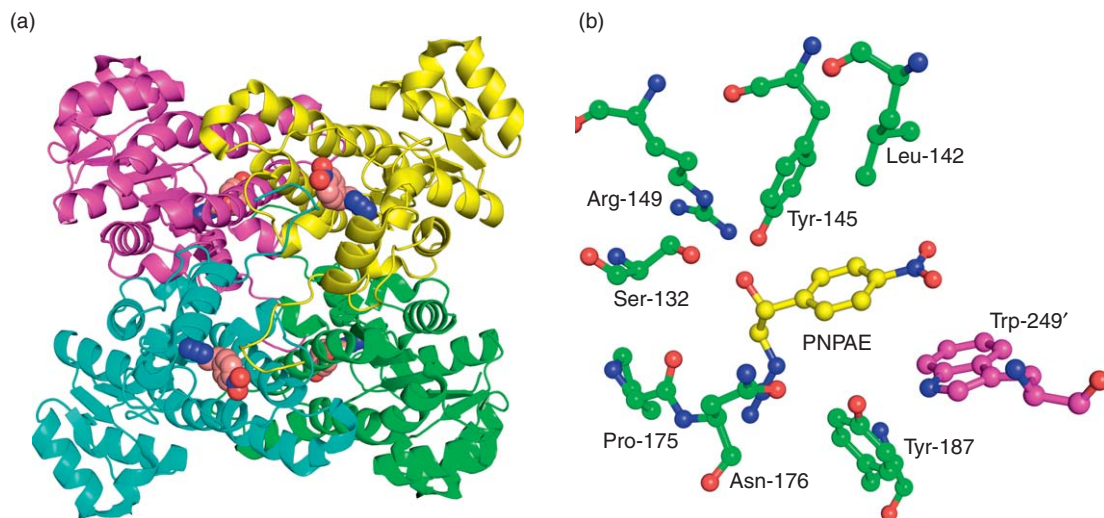
Figure 9 (a) Ribbon diagram of the homotetrameric structure of HheC. For clarity, each monomer is shown in a different color (purple, yellow, cyan, and green). The location of the active site in each monomer is indicated by the bound haloalcohol substrate mimic, (R)-1-*para*-nitro-phenyl-2-azido-ethanol (PNPAE), which is shown in CPK representation. (b) A close-up of the active site of HheC in complex with PNPAE . For clarity, the carbon atoms of the active site residues are shown in green or purple (monomer labeling), whereas those of the bound PNPAE are shown in yellow. The roles of the key active site residues (Ser-132, Leu-142, Tyr-145, Arg-149, Pro-175, Asn-176, Tyr-187, and Trp-249′) and their interactions are discussed in the text. The prime designation indicates that Trp-249 comes from an opposite monomer. The figure was prepared with PyMOL (W. L. DeLano, *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, USA, 2002, http://www.pymol.org.).

and the enzyme HheA[88] from *Arthrobacter* sp. strain AD2 (A-type) will be discussed here because these are the only two haloalcohol dehalogenases for which structural information is available.

The crystal structures of HheA and HheC (**Figure 9**) confirmed what was expected on the basis of their amino acid sequences, that is, these enzymes are related to the widespread family of NAD(P)H-dependent short-chain dehydrogenases/reductases (SDR family).[87,88] The members of the SDR family are structurally homologous proteins that share the well-known dinucleotide-binding Rossmann fold, as well as a Ser-Tyr-Lys/Arg catalytic triad.[89] The majority of the identified SDR family members are redox enzymes that catalyze the oxidation of a hydroxyl group or the reduction of a carbonyl group in a wide variety of alcohols, steroids, and sugars. Significantly, HheA and HheC possess the fold and catalytic triad (Ser-Tyr-Arg) of the SDR family (**Figure 9**).[87,88]

Both HheA and HheC function as a homotetramer, which can be viewed as a dimer of dimers (**Figure 9(a)**).[87,88] Each subunit consists of a seven-stranded parallel β-sheet flanked on both sides by α-helices, which resembles the characteristic Rossman fold. The dimer is formed by the interaction of the two longest α-helices and part of their connecting loops, forming an intermolecular, antiparallel four-helix bundle. Two of these dimers form the functional homotetramer, mainly through contacts between two β-strands and two α-helices, and their connecting loops. The active site is located in a loop-rich cavity that contains the catalytic triad (**Figures 9(a) and 9(b)**). The fold and dimerization interfaces of HheA and HheC are typical for members of the SDR family, which mainly occur as dimers or tetramers.[87,88]

A remarkable difference between the haloalcohol dehalogenases and the SDR family enzymes is that HheA and HheC lack the characteristic dinucleotide-binding Gly-X-X-X-Gly-X-Gly motif that is found in SDR enzymes.[87,88] In HheA and HheC several larger residues replace the smaller ones of the motif and occupy the space where part of the NAD(P)H cofactor is bound. Instead, a spacious halide binding site is found at the location of the NAD(P)H binding site in SDR proteins. This fascinating evolutionary lineage suggests that HheA and HheC are SDR superfamily members that have lost their NAD(P)H binding capability, and have recruited a halide-binding site that facilitates haloalcohol dehalogenation rather than a redox reaction.

To a large extent the knowledge about the catalytic mechanism of haloalcohol dehalogenases was obtained from the crystal structures of HheC complexed with either a bromide ion, (R)-1-para-nitro-phenyl-2-azido-ethanol (PNPAE) (a haloalcohol substrate mimic) (**Figure 9(b)**), or (R)-styrene oxide (an epoxide product) and a chloride ion.[87] These structures support a catalytic mechanism in which Tyr-145 of the catalytic triad (Ser-132, Tyr-145, Arg-149) deprotonates the haloalcohol hydroxyl function to generate an intramolecular nucleophile that attacks the vicinal carbon and displaces the halogen, yielding the epoxide product and a chloride ion (**Figure 10**). Deprotonation of the haloalcohol hydroxyl function is assisted by Arg-149, which likely lowers the pK_a of Tyr-145, and Ser-132, which positions the substrate and likely stabilizes the partial negative charge that develops on the haloalcohol hydroxyl oxygen. The residues of the catalytic triad were also shown to be important for catalysis because the individual S132A, Y145F, and R149N mutants are more than 10 000-fold less active than the wild-type enzyme.[86] Thus, whereas the aspartate-dependent hydrolytic dehalogenases discussed above employ covalent catalysis (a two-step hydrolysis reaction), haloalcohol dehalogenases catalyze an unusual one-step intramolecular substitution reaction that does not involve a covalent enzyme–substrate intermediate.

Further insight into the catalytic mechanism of HheC came from kinetic, product inhibition, and mutagenesis studies.[90–92] Kinetic analyses were performed with para-nitro-2-bromo-1-phenylethanol (PNSHH) as a model substrate (Equation (2)).[90] HheC displays a 150-fold lower catalytic efficiency for the (S)-enantiomer ($k_{cat} = 7\,s^{-1}$; $K_m = 430\,\mu mol\,l^{-1}$; $k_{cat}/K_m = 1.6 \times 10^4\,mol^{-1}\,l\,s^{-1}$) than for the (R)-enantiomer ($k_{cat} = 22\,s^{-1}$; $K_m = 9\,\mu mol\,l^{-1}$; $k_{cat}/K_m = 2.5 \times 10^6\,mol^{-1}\,l\,s^{-1}$) of PNSHH, resulting in a very high enantiopreference for the (R)-enantiomer. The conversion of (R)-PNSHH follows an ordered Uni−Bi mechanism, and the inhibition pattern of bromide ion as well as the occurrence of burst kinetics suggested that the bromide ion is first released from the enzyme, followed by the epoxide product (para-nitrostyrene oxide, PNSO). In addition, multiple turnover analyses showed that the binding of (R)-PNSHH occurs in a rapid equilibrium step and that the rate of formation of the enzyme–product ternary complex is ~380 s^{-1}. Bromide ion release (~21 s^{-1}) appeared to be rate-limiting in the overall catalytic cycle of the forward reaction.[90]

$$\text{(2)}$$

PNSHH   PNSO
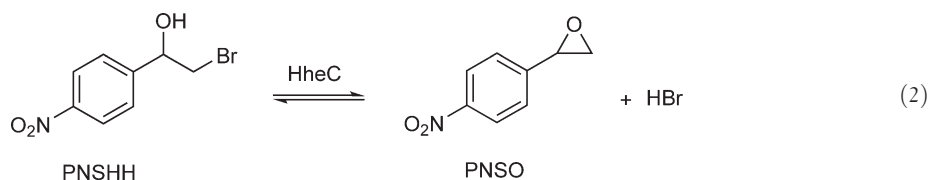
As the halide-binding site of HheC can accommodate a range of small negatively charged ions and because halide release is the rate-limiting step in the catalytic cycle of the ring closure reaction, the nature of the spacious halide-binding site in HheC warrants comment. The interactions of the halide ion are more extensive with the residues of this binding site and more hydrophobic in nature than those previously found for haloalkane dehalogenases and HADs. The binding site is mainly formed by the backbone of residues Pro-175, Asn-176, Tyr-177, and Leu-178, which are in a loop region.[87] The chloride ion also interacts with a
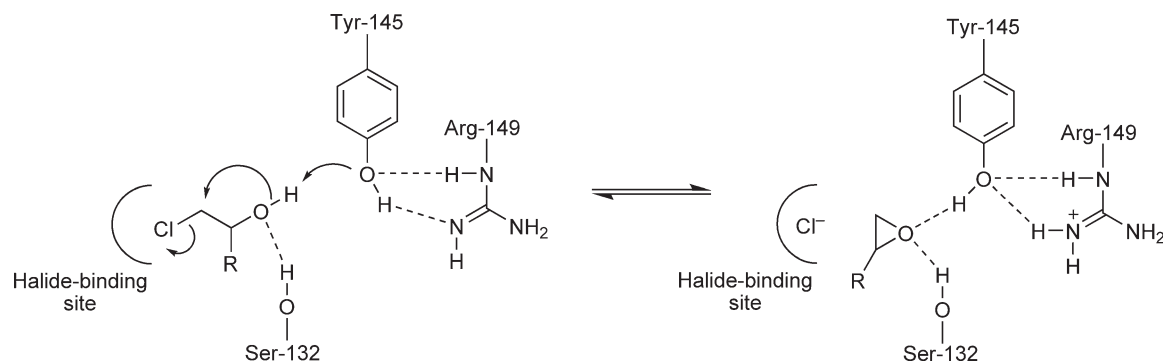
Figure 10 A schematic representation of the catalytic mechanism of haloalcohol dehalogenase (HheC). The proton of the catalytic tyrosine (Tyr-145) is proposed to be released to the solvent via Arg-149.[87]

water molecule and with the side chains of Pro-175, Leu-178, Phe-12, Phe-186, and Tyr-187. The loop that forms the halide-binding site is stabilized by hydrogen bonds between Asn-176 and Tyr-187 and between Tyr-187 and Trp-249 from another subunit (**Figure 9(b)**).[87] Janssen and co-workers hypothesized that deletion of these hydrogen bonds could possibly weaken the packing of the halide-binding site and thus facilitate the motions that are required for halide release and accelerate the catalytic cycle of the enzyme. Accordingly, they constructed and characterized two mutants of HheC, Y187F and W249F.[92] Strikingly, these mutants showed a higher catalytic activity toward both aliphatic and aromatic substrates and a higher rate of halide ion release than wild-type HheC. These results represent an elegant example of rational protein engineering, where a rate-limiting step had first been identified and then improved in order to increase the biocatalytic performance of the enzyme.

Another interesting aspect of the crystallographic work on HheC is the unusual way in which the molecular basis of the high enantioselectivity of this enzyme in nucleophilic ring opening reactions was uncovered. The HheC-catalyzed nucleophilic ring opening of racemic PNSO by azide only results in conversion of (R)-PNSO, but not of (S)-PNSO, even when the (R)-enantiomer has been fully converted.[93] A detailed kinetic analysis of this reaction, as described above for PNSHH, was not possible because of the low solubility of PNSO in water. Instead, Dijkstra and co-workers determined X-ray structures of complexes of HheC with the favored and unfavored enantiomers of PNSO.[93] Comparison of these structures demonstrated that the aromatic parts of the two enantiomers bind in a very similar way, but the oxirane ring of the unfavored (S)-enantiomer binds in a nonproductive manner, with the oxirane oxygen and the $C\beta$ atom positions inverted with respect to those of the favored (R)-enantiomer. These results, coupled with the calculated small difference in free energy of binding of the two enantiomers ($\Delta\Delta G = 4.7 \pm 3.1$ kJ mol^{-1}), suggest that the high enantioselectivity of HheC in nucleophilic ring opening reactions of aromatic epoxides by azide is mainly determined by the inability of the unfavored (S)-enantiomer to form a productive complex in the active site, rather than by a large difference in binding affinity of the different enantiomers.[93] This is in sharp contrast with the corresponding dehalogenation reaction, in which weaker binding of the unfavored (S)-enantiomer of PNSHH governs the enantioselectivity (vide supra). Manual docking of the (R)- and (S)-enantiomers of the haloalcohol substrate 1-*para*-nitrophenyl-2-chloroethanol in the active site of HheA shows that both enantiomers can be accommodated in a productive manner.[88] This is in agreement with the lack of enantiopreference of HheA.

8.04.3.2 3-Chloroacrylic Acid Dehalogenases

The 3-chloroacrylic acid dehalogenases catalyze the hydrolytic dehalogenation of either *cis*- or *trans*-3-haloacrylates to yield malonate semialdehyde and hydrogen halides (**Figure 6(c)**). These enzymes are interesting because they catalyze a rare and difficult cofactor-independent hydrolytic dehalogenation of a halogenated alkene. These dehalogenases have been purified from a few microorganisms that can utilize either 1,3-dichloropropene or 3-chloroacrylic acid as the sole source of carbon and energy.[62,94–96] The CaaD[62] from *Pseudomonas pavonaceae* 170 and the *cis*-CaaD[96] from coryneform bacterium strain FG41 are discussed as these are the only two 3-chloroacrylic acid dehalogenases for which detailed sequence and structural information are available.

8.04.3.2.1 Trans-3-chloroacrylic acid dehalogenase
CaaD is part of a pathway that is responsible for the degradation of the nematocide 1,3-dichloropropene in the soil bacterium *Pseudomonas pavonaceae* 170 (**Figure 1(b)**).[13,62] Its metabolic function is to convert *trans*-3-chloroacrylate into malonate semialdehyde ($k_{cat} \sim 3$ s^{-1}, $k_{cat}/K_m \sim 1.2 \times 10^5$ mol^{-1} l s^{-1}), which is probably the most difficult reaction of the pathway.[97] 3-Chloroacrylates are stable at room temperature and neutral pH, and nonenzymatic decomposition requires harsh conditions. From examination of rate constants at elevated temperatures, Horvat and Wolfenden[98] reported that the half-time at 25 °C and pH 7 for spontaneous hydrolytic dechlorination of *trans*-3-chloroacrylic acid is ~10 000 years, several orders of magnitude longer than half-times for the chemical decomposition of other notable environmental pollutants such as 1,2-dichloroethane (72 years), paraoxon (13 months), atrazine (5 months), and 1,3-dichloropropene (9.5 days).[98–101] As the uncatalyzed dechlorination reaction proceeds at a rate of 2.2×10^{-12} s^{-1} at 25 °C, CaaD provides a 10^{12}-fold rate enhancement.

The enzyme has a rather narrow substrate specificity: only the *trans*-isomers of 3-chloro- and 3-bromoacrylates and 3-halopropiolates undergo hydrolytic cleavage.[62,97] Structural analogues lacking either the carboxylate group (i.e., 3-chloroallylalcohol) or the halogen substituent (i.e., acrylate) are not processed. Those bearing pseudo-halogens (i.e., 3-cyano- and 3-methoxyacrylates) or chlorine at the C-2 position are also not processed. Three possible mechanisms for the CaaD reaction were initially proposed on the basis of its substrate specificity.[62,97,102] The first mechanism involves addition of water to the double bond, followed by either enzyme-catalyzed or nonenzymatic decomposition of the halohydrin intermediate to afford malonate semialdehyde (**Figure 11(a)**). The second mechanism involves a vinylic addition–elimination step, where the chlorine atom is displaced by a water-derived hydroxyl group, followed by tautomerization of the enol intermediate (**Figure 11(b)**). The third mechanism involves an addition–elimination step in which an active site nucleophile (e.g., a carboxylate group) displaces the halogen and generates a covalent alkenyl–enzyme intermediate (**Figure 11(c)**). Subsequent hydrolysis of the intermediate yields product and the free enzyme.

The first two mechanisms are fundamentally distinct from those of other hydrolytic dehalogenases because they do not involve the formation of a covalent (aspartyl) intermediate. The latter mechanism is somewhat analogous to that of the 4-chlorobenzoyl-CoA dehalogenase-catalyzed reaction (**Figure 8**). A single turnover experiment in $H_2^{18}O$ could be used to determine whether covalent catalysis is operative. Hydrolysis of the covalent intermediate would release the enzyme and product, and result in the incorporation of the ^{18}O label in the enzyme but not in the product. However, the outcome of this experiment would be ambiguous for the CaaD reaction due to the rapid formation of the hydrate of malonate semialdehyde, which would produce ^{18}O-labeled product. However, CaaD is not inactivated by hydroxylamine in the presence of substrate, which argues against this mechanism.[102] Albeit indirect, inactivation by hydroxylamine can be diagnostic of an alkyl–enzyme intermediate because hydroxylamine, a potent nucleophile that is much more reactive than water, can attack the ester intermediate and result in the formation of catalytically inactive adduct(s) at the site of the active site carboxylate.

Valuable insights about the catalytic mechanism of CaaD came from the observation and characterization of two promiscuous activities. First, it was found that CaaD shows a promiscuous hydratase activity and catalyzes
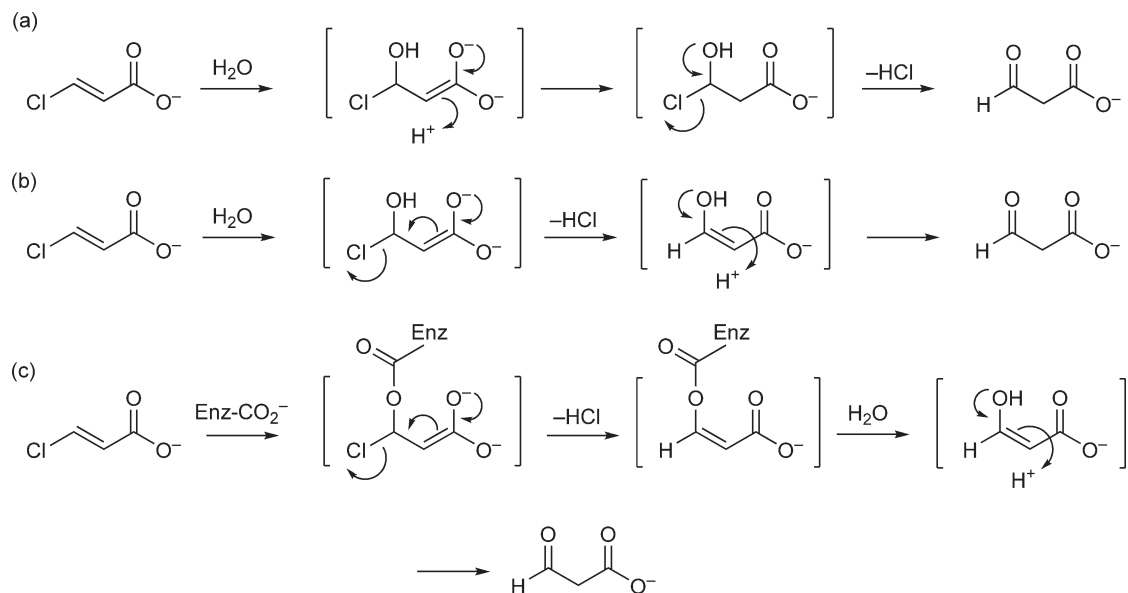
Figure 11   Three possible mechanisms for the dehalogenation of *trans*-3-chloroacrylate. (a) Addition of water to the double bond, followed by enzyme-catalyzed or chemical decomposition of a short-lived halohydrin intermediate to afford malonate semialdehyde. (b) Conjugate addition reaction where the chlorine atom is displaced by a water-derived hydroxyl group, followed by tautomerization of the enol intermediate. (c) Conjugate addition reaction where the chlorine atom is displaced by an active site carboxylate group, followed by hydrolysis of the covalent ester intermediate, and tautomerization of the enol intermediate.
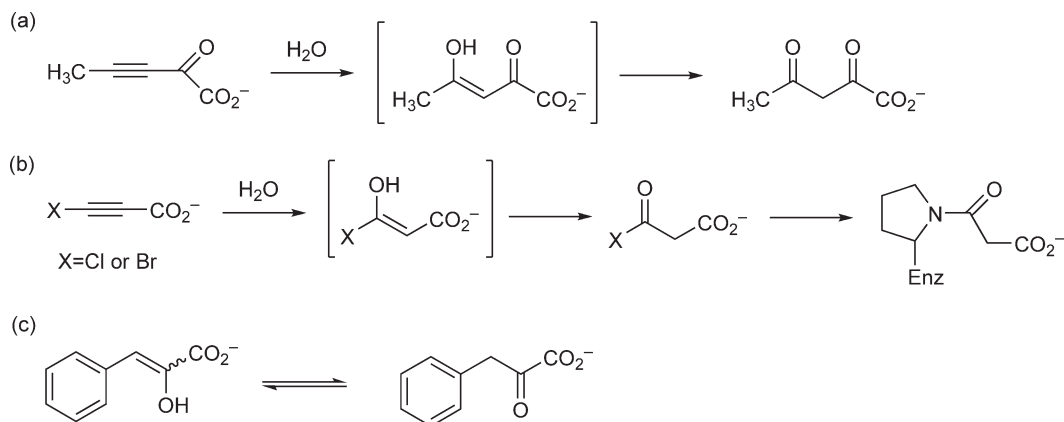
Figure 12 Three promiscuous activities of CaaD. (a) The CaaD-catalyzed hydration of 2-oxo-3-pentynoate to yield acetopyruvate. (b) The CaaD-catalyzed hydration of 3-halopropiolates to yield irreversible inhibitors of the enzyme. (c) The CaaD-catalyzed tautomerization of phenylenolpyruvate.

the hydration of 2-oxo-3-pentynoate and 3-chloro- and 3-bromopropiolate (**Figures 12(a) and 12(b)**).[97] Hydration of 2-oxo-3-pentynoate afforded acetopyruvate ($k_{cat}/K_m = 6.4 \times 10^3 \, \mathrm{mol^{-1} l s^{-1}}$), whereas the 3-halopropiolates are converted into potent irreversible inhibitors of the enzyme. Although not physiologically relevant, these promiscuous hydratase activities of CaaD were used to implicate a hydration reaction mechanism (**Figures 11(a) and 11(b)**) in its native activity, the hydrolytic dehalogenation of *trans*-3-haloacrylates.[97] Second, CaaD was found to display a significant phenylpyruvate tautomerase activity ($k_{cat}/K_m = 2.3 \times 10^4 \, \mathrm{mol^{-1} l s^{-1}}$) (**Figure 12(c)**).[103] Moreover, the reaction is stereoselective in D_2O, resulting in the formation of the 3S-isomer of [3-^{2}H]phenylpyruvate in a 1.8:1 ratio. This promiscuous activity is consistent with a mechanism involving a tautomerization step in the native catalytic cycle (**Figure 11(b)**).

The crystal structure of CaaD inactivated by 3-bromopropiolate provided further insight into the catalytic mechanism (**Figure 13**).[104] CaaD is a heterohexamer (~50 kDa) composed of three α-subunits (75 amino acids) and three β-subunits (70 amino acids). The enzyme forms a barrel-shaped hexamer, which can be viewed as a trimer of heterodimers (**Figure 13(a)**). The overall topology of CaaD is similar to that observed in all characterized members of the tautomerase superfamily, a group of structurally homologous proteins that share a characteristic β-α-β-fold as well as a catalytic N-terminal proline.[105,106] Till the discovery of CaaD, all characterized superfamily members were tautomerases that catalyzed keto–enol conversions of a pyruvoyl moiety using Pro-1, which has a low pK_a of ~6.4, as the proton-transferring base.[106] Significantly, both the α- and β-subunits of CaaD have an N-terminal proline, but mutagenesis and chemical modification studies have demonstrated that only Pro-1 of the β-subunit is critical for CaaD activity.[62]

CaaD contains three active sites, which are located at the interface between two ($\alpha\beta$)-heterodimers on one side of the hexamer, each harboring the catalytically important βPro-1 (**Figure 13**).[104] This catalytic proline is surrounded by charged residues including αArg-8, αArg-11, and αGlu-52. In the structure of the inactivated enzyme (**Figure 13(b)**), βPro-1 forms a covalent bond to the C-3 of a malonyl group, the adduct resulting from the enzyme-catalyzed transformation of 3-bromopropiolate.[104] The carboxylate group of the adduct interacts with the two arginines (αArg-8 and αArg-11), suggesting an orientation for substrate in the active site. Also, the adduct has hydrophobic contacts with αPhe-50, αLeu-57, and β-Ile-37. One carboxylate oxygen of αGlu-52 makes a hydrogen bond with the backbone carbonyl group of βIle-37, suggesting that the glutamate side chain is not ionized, which is likely the result of abstracting a proton from the water molecule during the hydration of 3-bromopropiolate. This observation strongly indicates αGlu-52 as the water activating residue in CaaD.[104]

On the basis of these observations, the following mechanism was proposed (**Figure 14**). The first step in catalysis is the nucleophilic attack of a water molecule on C-3 of *trans*-3-chloroacrylate to form an enzyme-stabilized enediolate intermediate. This step is facilitated by two additional actions: the activation of water by αGlu-52 and the alignment and polarization of the α,β-unsaturated carboxylate substrate by interactions with αArg-8 and αArg-11. After water addition, the enzyme-stabilized enediolate intermediate can undergo two
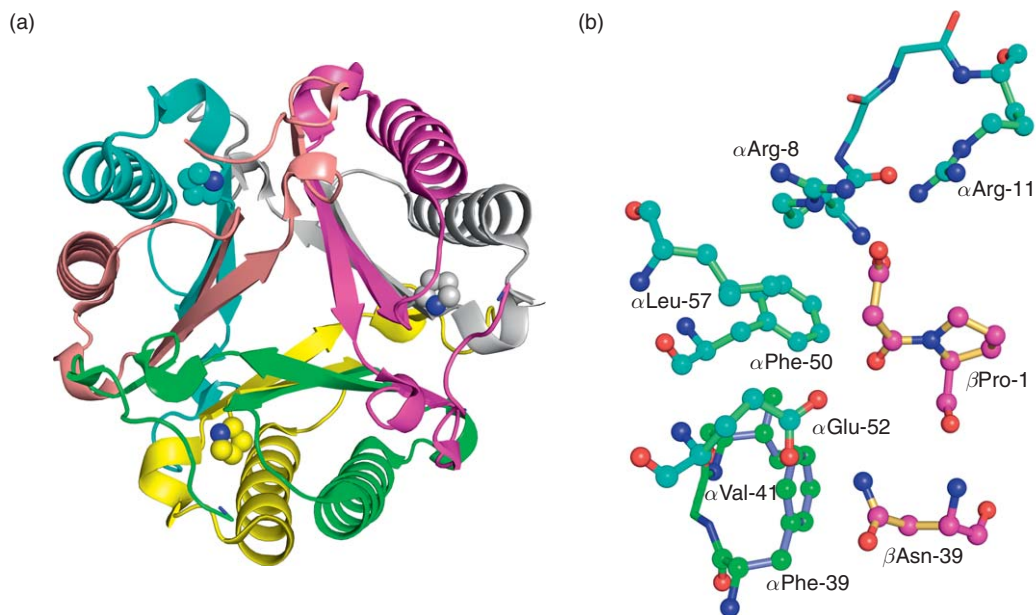
Figure 13 (a) Ribbon diagram of the heterohexameric structure of CaaD. For clarity, the three α-subunits are shown in green, brown, and purple, whereas the three β-subunits are shown in yellow, cyan, and gray. The catalytic Pro-1 of the β-subunit is shown in CPK representation. (b) A close-up of the active site of CaaD where βPro-1 is covalently modified by 3-oxopropanoate, which results from the CaaD-catalyzed hydration of 3-bromopropiolate (**Figure 12(b**)). The roles of the key active site residues (βPro-1, αArg-8, αArg-11, and αGlu-52) and their interactions are discussed in the text. The other residues (αPhe-39, β-Asn-39, α-Val-41, α-Phe-50, and αLeu-57) further define the active site. The figure was prepared with PyMOL (W. L. DeLano, *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, USA, 2002, http://www.pymol.org.).
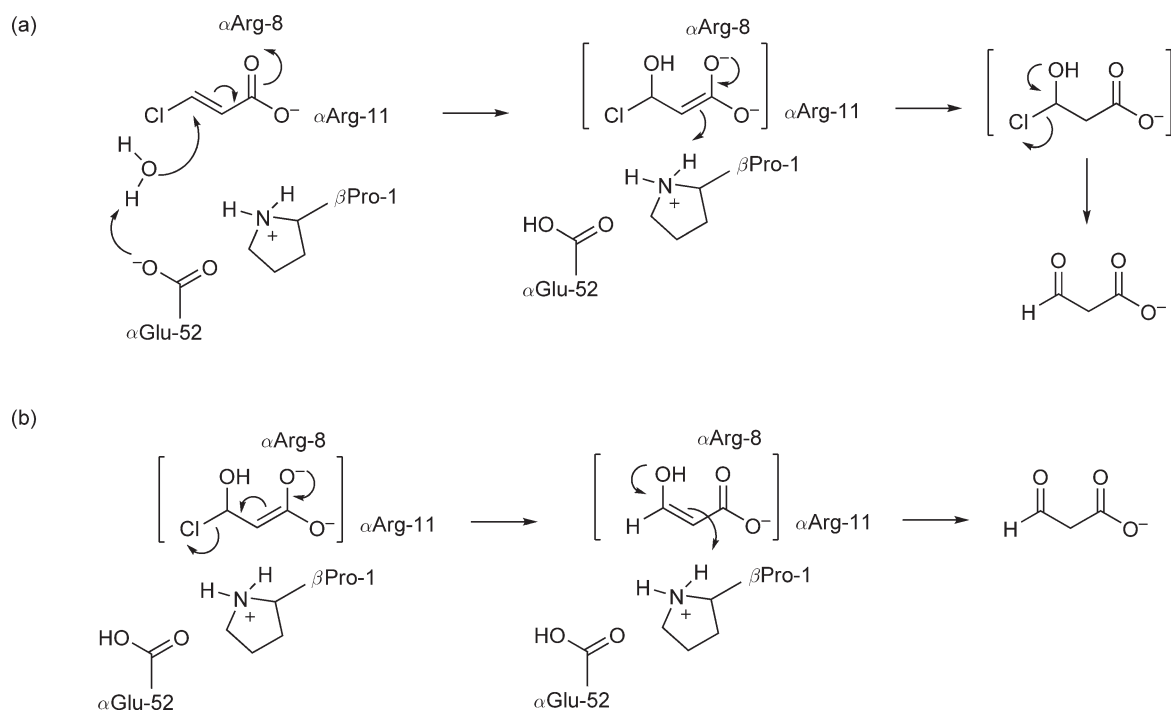
Figure 14 A schematic representation of the catalytic mechanism of *trans*-3-chloroacrylic acid dehalogenase (CaaD), showing the key participants in the reaction.

equally plausible fates. As shown in **Figure 14**(**a**), the enediolate can ketonize with protonation at C-2 by βPro-1. The resulting chlorohydrin intermediate can then undergo an enzyme-catalyzed or a nonenzymatic process to produce malonate semialdehyde. Alternatively, the enediolate can ketonize with elimination of chloride (**Figure 14**(**b**)). Subsequent, ketonization of the enol and protonation at C-2 by βPro-1 yields malonate semialdehyde.[104] The phenylpyruvate tautomerase activity of CaaD seemingly favors the α,β-elimination mechanism (**Figure 14**(**b**)), but this activity may also reflect the ketonization of the initially formed enediolate intermediate.[103]

The pH dependence of the kinetic parameters for the CaaD-catalyzed reaction is consistent with the proposed general acid/base mechanism. The pH rate profile implicates two groups on the free enzyme with pK_a values of 7.6 and 9.2 for optimal activity.[107] Moreover, direct pH titration of the uniformly ^{15}N-labeled CaaD and the αP1A-mutant, where the chemical shift of the prolyl nitrogen is followed as a function of pH by using ^{15}N NMR spectroscopy, indicated that βPro-1 is likely responsible for the pK_a value of 9.2.[107] The fact that the catalytically important βPro-1 of CaaD is charged at physiological pH, and not neutral as had been previously seen for the tautomerase superfamily enzymes, was an important step in defining the role of this residue in catalysis.

The role of the glutamate was further investigated by analyzing the kinetic properties of two site-specific mutants (αE52Q, αE52D).[104] Replacing this residue with a glutamine or an aspartate results in no detectable activity (i.e., αE52Q) or greatly reduced activity (i.e., αE52D). Along with the crystallographic evidence, this has led to the conclusion that αGlu-52 is the most likely candidate to activate a water molecule. Further evidence for involvement of the two arginine residues, αArg-8 and αArg-11, came from a series of [1]H–[15]N heteronuclear single quantum coherence (HSQC) NMR experiments. In these experiments, Azurmendi *et al.*[107] used a saturating amount of the competitive inhibitor 3-chloro-2-butenoic acid (3-CBA) and observed the NϵH signals for αArg-8 and αArg-11. In the absence of 3-CBA, these signals (and the NϵH resonances for the other arginine residues) disappear due to base-catalyzed NϵH exchange. The selective protection of the NϵH resonances for αArg-8 and αArg-11 in the presence of 3-CBA reflects an interaction between αArg-8 and αArg-11 and the carboxylate group of 3-CBA. Extrapolation of this observation to substrate binding supports the proposed roles for αArg-8 and αArg-11 in binding and activation of the substrate, and in stabilizing the enediolate intermediate in the conjugate addition mechanism.[107] The two arginine residues were also shown to be essential for catalysis as the individual αR8A and αR11A mutants are not able to process *trans*-3-haloacrylates.[97,104]

8.04.3.2.2   *Cis-3-chloroacrylic acid dehalogenase*

cis-CaaD is part of a degradative pathway for *cis*-3-chloroacrylate in coryneform bacterium strain FG41.[95] The enzyme, like CaaD, catalyzes the dehalogenation of 3-haloacrylates (in this case the *cis*-isomers) and the hydration of 2-oxo-3-pentynoate and 3-halopropiolates.[96] As expected on the basis of this functional relationship to CaaD, sequence analysis of *cis*-CaaD revealed that this enzyme belongs to the same protein superfamily (i.e., the tautomerase superfamily).[96] However, CaaD and *cis*-CaaD have low sequence identity ($\sim$20%) and different oligomerization states. While CaaD is a heterohexamer consisting of three 75-residue α-chains and three 70-residue β-chains, *cis*-CaaD functions as a homotrimer, where each subunit is composed of 149 amino acid residues. As a result, the two enzymes have been classified in two different families of the tautomerase superfamily.[96]

The structural relationship between *cis*-CaaD and CaaD underscores a common theme in the tautomerase superfamily. Recent X-ray structure work on the native *cis*-CaaD showed that each monomer consists of a four-stranded β-sheet that is formed by the antiparallel interaction of a pair of two-stranded parallel β-sheets (**Figure 15(a)**).[108] Two α-helices, each spanning the two strands of the parallel β-sheets, lie antiparallel to each other in the concave side of the β-sheet plane. Hence, each monomer is made up of two β-α-β building blocks and can be considered as the fusion product of the CaaD α- and β-chain, each coding the single β-α-β motif. Three *cis*-CaaD monomers form the functional barrel-like trimer, which contains three active sites with the N-terminal proline buried in the interior of a monomer on one side of the trimer (**Figure 15**(**a**)). A structural alignment shows that the two connected β-α-β structural motifs in *cis*-CaaD, and the α- and β-chains of CaaD align with rmsd values of 0.9 Å for 50 C$_\alpha$ atoms.[108] Thus, even though *cis*-CaaD and CaaD have low sequence identity and belong to two different families of the tautomerase superfamily, they are structurally very similar.
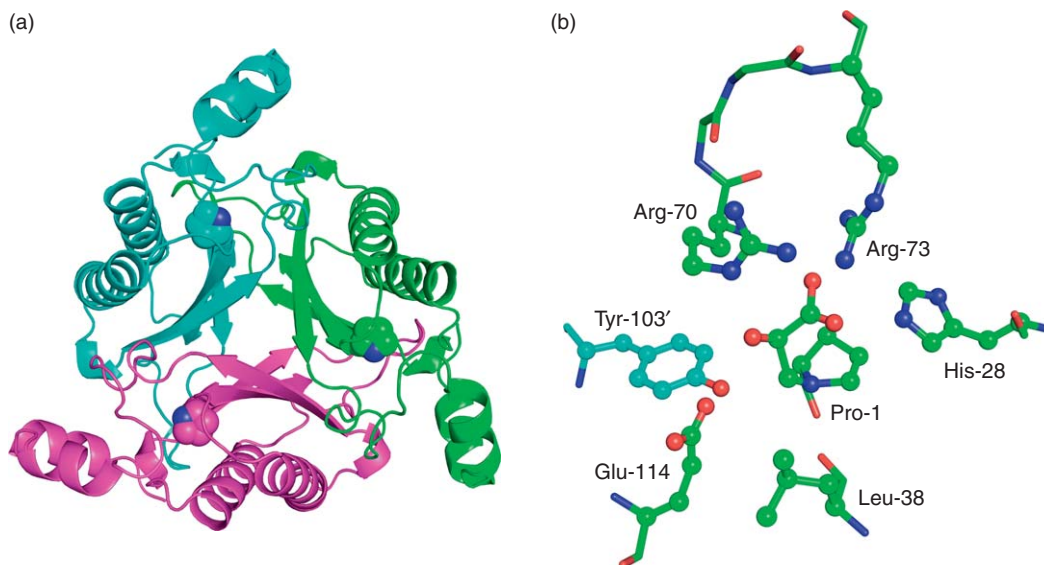
Figure 15 (a) Ribbon diagram of the homotrimeric structure of *cis*-CaaD. For clarity, each subunit is shown in a different color (cyan, purple, and green). The catalytic Pro-1 is shown in CPK representation. (b) A close-up of the active site of *cis*-CaaD where Pro-1 is covalently modified by (*R*)-2-hydroxypropanoate, which results from the affinity-labeling reaction between *cis*-CaaD and (*R*)-oxirane-2-carboxylate. The roles of the key active site residues (Pro-1, His-28, Arg-70, Arg-73, Tyr-103′, and Glu-114) and their interactions are discussed in the text. The prime designation indicates that Tyr-103 comes from an adjacent subunit. The figure was prepared with PyMOL (W. L. DeLano, *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, USA, 2002, http://www.pymol.org.).

Initially, much of the catalytic mechanism for *cis*-CaaD was deduced from mutagenesis and chemical modification studies, paralleling those used for CaaD.[96] For example, the importance of Pro-1 in catalysis was demonstrated by site-directed mutagenesis: the replacement of Pro-1 with an alanine resulted in an inactive enzyme. Reaction of *cis*-CaaD with 3-chloropropiolate, a mechanism-based inhibitor, led to the alkylation of Pro-1 and the concomitant loss of catalytic activity. The presence of substrate protected the enzyme from modification and inactivation by 3-chloropropiolate. Besides the presence of a catalytically important N-terminal proline, three other key functionalities are conserved. The two arginine residues (α-Arg-8 and αArg-11) in CaaD, which are responsible for binding and polarization of the substrate's carboxylate group, and the glutamate residue (αGlu-52), which is the presumed water-activating base, are conserved as Arg-70, Arg-73, and Glu-114 in *cis*-CaaD respectively. The crystal structure of *cis*-CaaD locates these residues in the active site (**Figure 15**(**b**)) and mutational analysis revealed that Arg-70, Arg-73, and Glu-114 are critical for catalysis.[96,108]

The subtle mechanistic nuances between *cis*-CaaD and CaaD were discovered by affinity labeling with (*R*)-oxirane-2-carboxylate and subsequent crystallographic analysis (**Figure 15**(**b**)).[108,109] The carboxylate side chain of (*R*)-oxirane-2-carboxylate enables it to bind at the active site of *cis*-CaaD, in a mode that results in the covalent modification of Pro-1 and the loss of catalytic activity. The affinity-labeling reaction is stereospecific, because the (*S*)-enantiomer of oxirane-2-carboxylate does not alkylate the enzyme. The rate of inactivation is also hindered by the presence of substrate.[109] Taken together, these findings support the active site nature of the inactivation process and the critical contribution of Pro-1 to activity.

The crystal structure of *cis*-CaaD inactivated by the (*R*)-oxirane-2-carboxylate shows two additional active site residues (His-28 and Tyr-103) that are not present in CaaD (**Figure 15**(**b**)).[108] On the basis of the interactions observed in the complex, a mechanism for *cis*-CaaD was proposed (**Figure 16**). A key catalytic task for *cis*-CaaD is to activate a water molecule. In contrast to CaaD, where the water molecule is activated by a single residue (αGlu-52), *cis*-CaaD uses the side chains of two residues, Glu-114 and Tyr-103, to activate the nucleophilic water molecule. The first step in catalysis is the nucleophilic attack of the activated water molecule on C-3 of *cis*-3-chloroacrylate to form an enediolate intermediate. This intermediate is presumably
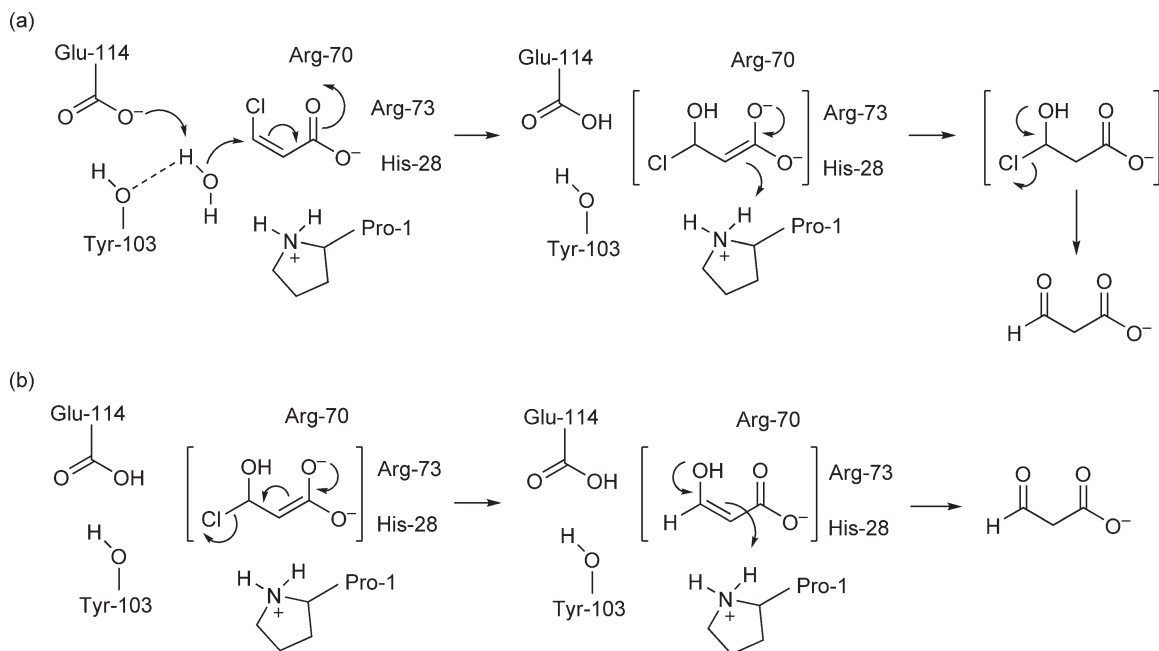
Figure 16 A schematic representation of the catalytic mechanism of *cis*-3-chloroacrylic acid dehalogenase (*cis*-CaaD), showing the key participants in the reaction.

stabilized by interactions with the side chains of Arg-70, Arg-73, and His-28. Ketonization of the enediolate may be assisted by Pro-1, which places a proton at the C-2 atom, and generates an unstable chlorohydrin intermediate (**Figure 16(a)**). Collapse of the proposed chlorohydrin could be an enzymatic or a nonenzymatic process. Alternatively, the enediolate can ketonize with elimination of chloride, followed by ketonization of the enol and protonation at C-2 by Pro-1, yielding the malonate semialdehyde product (**Figure 16(b)**).[108] Indeed, *cis*-CaaD also displays a phenylpyruvate tautomerase activity although not as robust as that of CaaD.[103]

The involvement of His-28 and Tyr-103 distinguishes the *cis*-CaaD mechanism from the otherwise parallel CaaD mechanism. The assisting roles of these two residues in activation of the substrate and the nucleophilic water molecule respectively are supported by site-directed mutagenesis: the Y103F and H28A mutants displayed substantially reduced *cis*-CaaD activities. (Neither mutant has been examined for structural damage which could contribute to the loss of activity.) Moreover, it has been suggested that the presence of His-28 and Tyr-103 in *cis*-CaaD but not in CaaD, could account for the individual substrate specificities.[108] The presence of the additional carboxylate-binding residue (His-28) in *cis*-CaaD results in differences in the orientation of the substrate in the active site with respect to CaaD. Tyr-103 seems to be responsible for the shape difference of the substrate-binding pockets, which fit the shape of their respective substrates. The presence of Tyr-103 in *cis*-CaaD effectively blocks the binding of the 3-chloro group of the *trans*-isomer of the substrate. Instead, a pocket formed by the side chains of Thr-34, Leu-38, Leu-119, and Arg-70 could favor the binding of the 3-chloro moiety of the *cis*-isomer.

The conservation of the β-α-β fold, the building block of the tautomerase superfamily, and key functionalities suggest that CaaD and *cis*-CaaD are related by divergent evolution from a common ancestor. The different oligomeric structures, the low sequence identity, and the differences in substrate specificity and reaction mechanism would further suggest that the two enzymes diverged quite some time ago. A gene duplication of a small gene encoding the β-α-β structural motif followed by co-evolution of the two genes could give rise to CaaD.[96,110] *cis*-CaaD might have evolved from an independent gene duplication event of the small ancestral gene followed by gene fusion.[96] Hence, the last common ancestor of CaaD and *cis*-CaaD was probably a small gene encoding the β-α-β structural motif. Despite their likely independent evolution, CaaD and *cis*-CaaD use similar catalytic strategies to process different isomers of 3-haloacrylates.

8.04.4   Structurally Undefined Dehalogenases

The six dehalogenases discussed above (haloalkane dehalogenase, HAD, 4-chlorobenzoyl-CoA dehalogenase, haloalcohol dehalogenase, CaaD, and *cis*-CaaD) have been studied by X-ray crystallography. These studies have provided details about their catalytic mechanisms and evolutionary relationships, and much insight into how nature cleaves the intrinsically difficult carbon–halogen bond. However, there are several additional dehalogenases that catalyze intriguing chemical reactions, but for whom structural studies have not been reported. For some of these, mechanistic insight has been obtained by sequence homology analysis in combination with site-directed mutagenesis. Some of the most interesting examples are described here.

8.04.4.1   Aliphatic Dehalogenases

An unusual dehalogenase (LinA) has been discovered as part of the γ-HCH catabolic pathway in *S. paucimobilis* UT26 (**Figure 1**(**c**)).[111] LinA catalyzes two sequential steps involving the elimination of HCl to form first 1,3,4,5,6-pentachlorocyclohexene and then 1,3,4,6-tetrachloro-1,4-cyclohexadiene. This elimination reaction is particulary interesting because most known aliphatic dehalogenases catalyze nucleophilic substitution reactions. LinA is a homotetrameric enzyme consisting of 16.5 kDa subunits and shares no sequence similarity with other known dehalogenases.[112,113]

Notably, the enzyme shows activity with α-, γ-, and δ-HCH, but not with β-HCH.[112] This observation suggests that LinA catalyzes the elimination of 1,2-biaxial HCl because β-HCH lacks such a pair (**Scheme 2**). Sequence analysis and site-directed mutagenesis provided the first mechanistic clues.[113,114] A PSI-BLAST search for homologous enzymes revealed that LinA shows a distant relationship with a structurally character-ized scytalone dehydratase, which catalyzes the dehydration of two intermediates in the biosynthesis of melanin. Although LinA and the dehydratase share no notable overall sequence identity, secondary and overall structure predictions suggest that the two proteins have a highly conserved α/β barrel structure with a distinctive hydrophobic active site cavity. Moreover, Asp-31 and His-85 in scytalone dehydratase, which form the essential catalytic dyad, are conserved as Asp-25 and His-73 in LinA. Site-directed mutagenesis experiments have confirmed that Asp-25 and His-73 are essential for LinA activity. From these studies, and from an analysis of the stereochemical course of the reaction, Damborský and co-workers[113,114] proposed that His-73 acts as the active site base that initiates the reaction by abstracting an axial proton, which results in the concomitant anti-elimination of chloride from the adjacent carbon atom (**Figure 17**). Asp-25 likely assists in proton abstraction by keeping His-73 in the proper orientation and by stabilizing the positive charge that develops on the His-73 imidazole ring. Further insight into the mechanistic details of this unusual dehalogena-tion reaction awaits the determination of the three-dimensional structure of the LinA enzyme.

Another intriguing dehalogenation reaction is catalyzed by dichloromethane dehalogenase (DcmA). This enzyme initiates the metabolism of dichloromethane, a widely used industrial solvent, by methylotrophic bacteria obtained from dichloromethane-contaminated soil.[115,116] The product of the reaction is formaldehyde, which can be oxidized to CO_2 to provide energy, or assimilated into biomass. DcmA is a homohexameric enzyme of 33 kDa subunits and requires glutathione (GSH) for activity.[117–121] This first suggested that DcmA could be mechanistically similar to GSH *S*-transferases, which catalyze the nucleophilic attack of GSH upon an electrophilic substrate to form a GSH conjugate. Although there is no published X-ray structure available for DcmA, sequence comparisons show that DcmA shares modest sequence identity with enzymes in the theta class
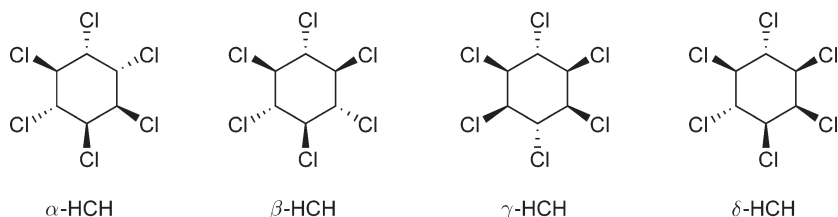
Scheme 2   Structures of the α, β, γ, and δ isomers of hexachlorocyclohexane (HCH).
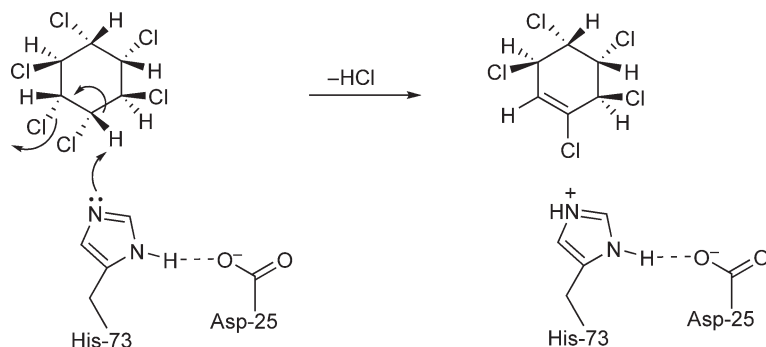
Figure 17 Proposed catalytic mechanism of γ-HCH dehydrochlorinase (LinA), highlighting the key step, the loss of the biaxial HCl.

of the GSH S-transferase superfamily (15–25% overall identity).[119,120] Also, the conserved and catalytically important serine in theta class GSH S-transferases, which is believed to enhance the nucleophilicity of the GSH thiol, is conserved as Ser-12 in DcmA. The observation that mutant enzymes in which Ser-12 has been replaced with alanine or threonine are inactive is consistent with a comparable GSH-activating role for Ser-12 in DcmA.[122]

The proposed mechanism for the enzyme is shown in **Figure 18**. DcmA likely catalyzes a nucleophilic substitution (S_N2) reaction, in which the GSH thiol (or thiolate) serves as the nucleophile and directly attacks the carbon atom of dichloromethane. The resulting S-(chloromethyl)GSH intermediate then undergoes facile hydrolysis, either enzymatically or free in solution, to form hydroxymethylglutathione. The latter compound decomposes to afford formaldehyde and regenerate the GSH. Evidence for S-(halomethyl)GSH formation comes from elegant ^{19}F-NMR spectroscopic studies using a substrate analogue that slowly undergoes turnover.[123] With the native substrate CH_2Cl_2, the S-(chloromethyl)GSH intermediate does not accumulate to a significant extent during turnover and therefore cannot be observed. However, the alternative substrate CH_2ClF, with fluorine as a poor leaving group, significantly reduces the rate of hydrolysis of the S-(halomethyl)GSH intermediate. The reduced rate of hydrolysis enabled the collection of ^{19}F-NMR spectroscopic data that provided evidence for the accumulation (and subsequent disappearance) of S-(fluoromethyl)GSH during turnover of CH_2ClF. The results of kinetic studies indicate that the initial addition of GSH to CH_2Cl_2 is fast and that the rate-limiting step in turnover is the release of the S-(chloromethyl)GSH species from the enzyme active site.[124] The release of this species occurs at a rate that is competitive with the rate of its spontaneous hydrolysis. These data are consistent with a S_N2 mechanism, where the role of the enzyme is simply to catalyze the nucleophilic attack of GSH upon dichloromethane to yield a S-(chloromethyl)thioester intermediate, which then decomposes nonenzymatically to form formaldehyde.

8.04.4.2 Aromatic Dehalogenases

A unique metal-dependent hydrolytic dehalogenase has been discovered in *Pseudomonas* sp. strain ADP by Wackett and co-workers.[125,126] The enzyme, atrazine chlorohydrolase (AtzA), initiates the metabolism of the herbicide atrazine (2-chloro-4-N-ethylamino-6-N-isopropylamino-1,3,5-triazine) by catalyzing a dechlorination reaction to yield hydroxyatrazine (**Figure 19(a)**). The overall reaction proceeds via a hydrolytic mechanism (rather than an oxygenase mechanism) as demonstrated in experiments using $H_2^{18}O$ in an argon atmosphere.[125] Accordingly, atrazine incubated with AtzA in $H_2^{18}O$ or $H_2^{16}O$ quantitatively yielded [18]O- or
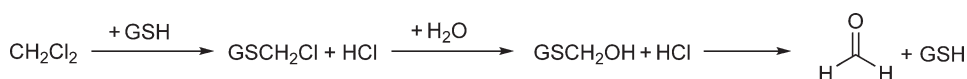
Figure 18 Proposed intermediates along the reaction pathway of dichloromethane dehalogenase (DcmA).
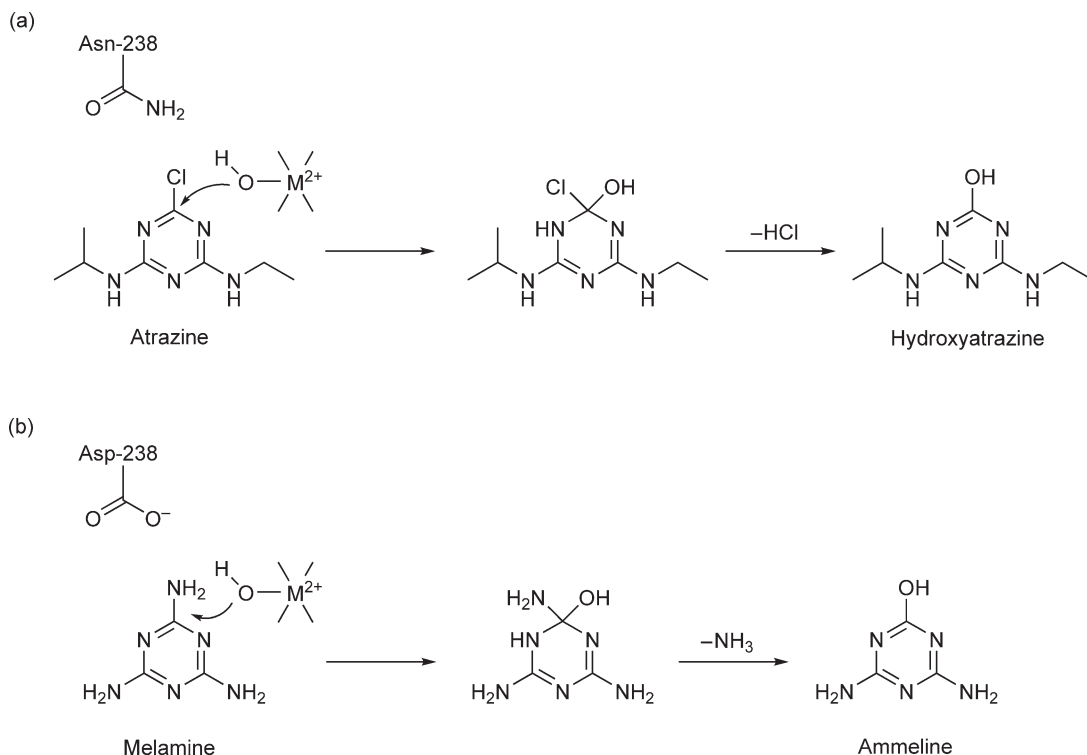
Figure 19   Models of (a) atrazine chlorohydrolase (AtzA) and (b) melamine deaminase (TriA) catalysis, picturing the catalytic metal functioning in water activation and the side chain of residue 238 (either Asn or Asp), which controls leaving group specificity.

^{16}O-labeled hydroxyatrazine, respectively. Hydroxyatrazine solubilized in $H_2^{18}O$ either in the presence or absence of AtzA did not yield detectable ^{18}O-hydroxyatrazine. These results rule out the possibility of a monooxygenation reaction and are consistent with the hydrolytic mechanism of the reaction catalyzed by AtzA.

The results of gel filtration chromatography experiments indicate that AtzA functions as either a tetramer or pentamer composed of 52.4 kDa subunits.[125] Initially, metal dependence was not reported. The observation that AtzA contains consensus amino acid sequences characteristic of enzymes in the amidohydrolase superfamily provided important clues about the enzyme and implicated a role for metal ions in its mechanism.[127,128] Members of the amidohydrolase superfamily are $(\alpha\beta)_8$ barrel proteins and many of them have conserved metal binding ligands and a common hydrolytic mechanism in which one or two metal ions are responsible for activating water for nucleophilic attack on the respective substrate.[129] A sequence alignment of AtzA with the structurally characterized superfamily members cytosine deaminase and adenosine deaminase, which are respectively involved in pyrimidine/purine metabolism, revealed that the metal binding ligands of the two deaminases are conserved in AtzA, despite the low overall sequence identity (20–30%) between AtzA and the deaminases.[128] This observation suggests a functional role for these conserved metal-binding residues (three histidines and one aspartate) in AtzA.

The presence of a conserved metal center in AtzA prompted Wackett and co-workers to perform detailed studies on the influence of metals on AtzA activity and the metal content of the native enzyme.[128] AtzA activity was shown to depend on an enzyme-bound, divalent transition-metal ion. The loss of activity obtained by incubating the enzyme with metal chelators was reversible upon addition of Fe^{II}, Mn^{II}, or Co^{II} salts. The results obtained from inductively coupled atomic plasma emission spectroscopy studies on the native enzyme indicate that there is about one iron atom per subunit. In the absence of an X-ray structure of AtzA, a minimal mechanism was proposed in which the catalytic iron atom is implicated in activating water for direct nucleophilic attack on the atrazine substrate (**Figure 19**(**a**)).[128]

It is noteworthy that melamine deaminase (TriA), an amidohydrolase superfamily member catalyzing the hydrolytic deamination of melamine (**Figure 19**(**b**)), shares 98% amino acid sequence identity with AtzA.[130] Each enzyme consists of 475 residues and the two enzymes differ by only nine amino acids. This very high overall sequence identity indicates a short evolutionary pathway connecting enzymes that catalyze physiologically relevant deamination and dehalogenation reactions, respectively. Since melamine is a related s-triazine that predates the use of atrazine, it was proposed that TriA may have served as the precursor for AtzA.[130] However, surprisingly, TriA has no significant dehalogenase activity toward chlorinated s-triazine substrates, whereas AtzA exhibits no detectable deaminase activity toward melamine (2,4,6-triamino-1,3,5-triazine) and comparable amino-substituted s-triazine substrates.[130] That TriA and AtzA discriminate between chloro and amino triazine substrates so well despite their sequences being 98% identical is most remarkable.

The divergence in function, from deamination to dechlorination, was postulated to arise from only a few amino acid changes in the TriA active site leading to the differential stabilization of amino and chloride leaving groups, respectively.[130] This hypothesis was supported by directed evolution experiments in which the *atzA* and *triA* genes were shuffled to generate hybrid proteins with one or a few amino acid changes compared to the parental enzymes and with different leaving group specificities in displacement from the s-triazine ring.[131] The results of these experiments suggested that leaving group specificity was largely controlled by the residue at position 238. An asparagine at this position was found in hybrid enzymes that are capable of only dechlorination, while an aspartate at position 238 was found in hybrid enzymes catalyzing any displacement reaction other than dechlorination. Apparently, the amide group assists in departure and stabilization of the chloride ion. Hence, it was proposed that the replacement of Asp-238 in TriA to an asparagine was an important step in the evolution of its AtzA activity (**Figures 19(a)** and **19(b)**).[5,131]

The removal of chlorine atoms from chlorinated aromatic compounds can also take place by a reductive reaction in which the chlorine atom is replaced by a hydrogen atom. The best studied reductive dehalogenase is tetrachlorohydroquinone (TCHQ) dehalogenase from *Sphingobium chlorophenolicum*, a soil bacterium that degrades pentachlorophenol, a widely used wood preservative. The enzyme catalyzes two sequential reductive dehalogenation steps to convert TCHQ first into trichlorohydroquinone (TriCHQ) and then into 2,6-dichlorohydroquinone (DCHQ) (**Scheme 3**).[132] The reducing equivalents for each step are provided by two molecules of GSH, which are oxidized to glutathione disulfide (GSSG).[133] TCHQ dehalogenase is a monomeric enzyme of 27 kDa that contains neither metal ions nor coenzymes. This observation first suggested that the reducing electrons must be transferred from GSH to TCHQ by some type of covalent interaction with GSH or an active site cysteine, rather than indirect transfer through a coenzyme or transition metal.[134]

A crystal structure of TCHQ dehalogenase is not yet available, but sequence homology analysis revealed that the dehalogenase has limited but significant (26–30% overall sequence identity) sequence similarity with several enzymes in the theta class of the GSH S-transferase superfamily. In fact, Copley and co-workers[135] have proposed that TCHQ dehalogenase is closely related to another member of this superfamily, maleylacetoacetate isomerase, which catalyzes the GSH-dependent isomerization of a *cis* double bond in maleylacetoacetate to the *trans* configuration during the catabolism of phenylalanine and tyrosine. The proposed evolutionary link between TCHQ dehalogenase and maleylacetoacetate isomerase was based on conservation of sequence in the active site region and the robust isomerization activity of TCHQ dehalogenase with maleylacetone, an analogue of maleylacetoacetate.[135]

The most recently proposed mechanism for TCHQ dehalogenase is shown in **Figure 20**.[136] The first step in catalysis is most likely the deprotonation of $TriCHQ^-$ to $TriCHQ^{-2}$, followed by a ketonization step and the
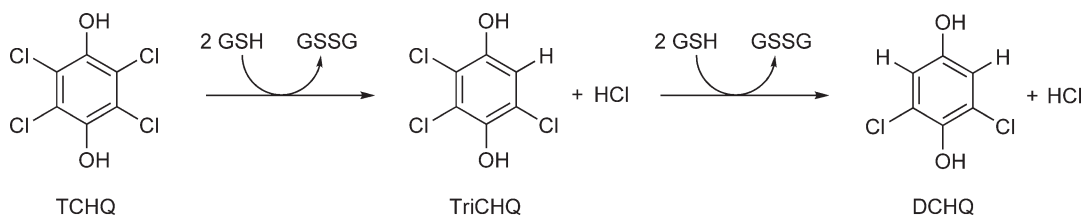
Scheme 3 Two sequential reductive dehalogenation steps catalyzed by TCHQ dehalogenase.
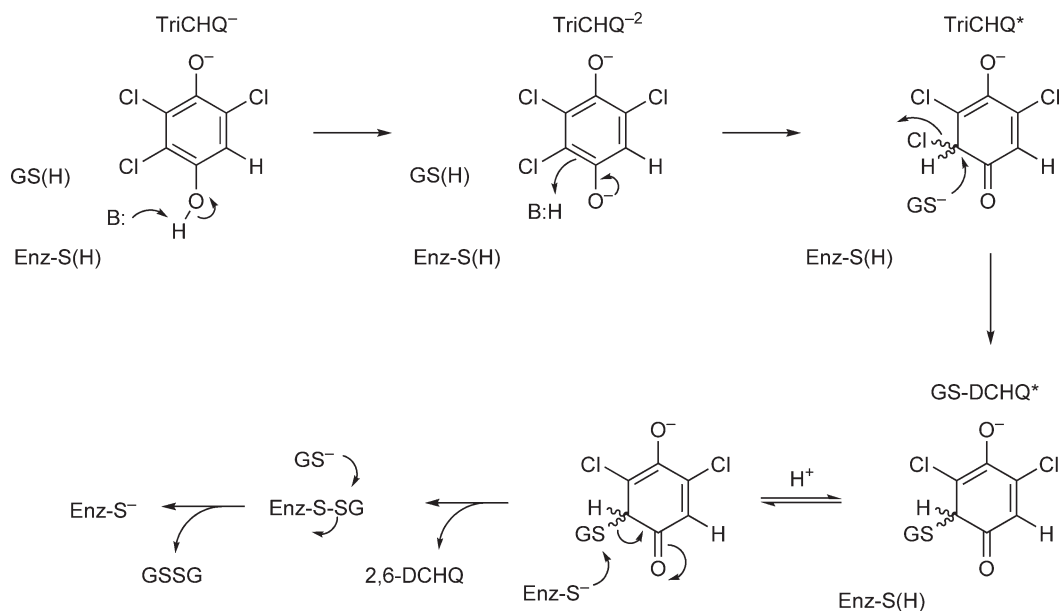
Figure 20 A schematic representation of the proposed catalytic mechanism of TCHQ dehalogenase, based on mechanistic studies and a pre-steady-state kinetic analysis.[136]

loss of aromaticity to give TriCHQ*. Nucleophilic attack of GSH upon TriCHQ* results in elimination of chloride and the formation of GS-DCHQ*. An active site cysteine (Cys-13) attacks the sulfur of the glutathionyl substituent, releasing the reduced and aromatic product, 2,6-DCHQ, and forming a mixed disulfide between the enzyme and GSH (Enz-S-SG). In the final step, a thiol–disulfide exchange reaction takes place where a second molecule of GSH regenerates the free enzyme and forms GSSG.

The last two steps of the proposed mechanism involving Cys-13 have significant experimental support. One of the major clues implicating an active site cysteine in catalysis was the observation that the enzyme undergoes oxidative damage during purification and that the damaged enzyme produces substantial amounts of the GSH conjugates 2,3,5-trichloro-6-S-glutathionylhydroquinone (GS-TriCHQ) and an unidentified isomer of dichloro-S-glutathionylhydroquinone (GS-DCHQ) in addition to the expected TriCHQ and DCHQ products.[134] The oxidative damage can be repaired by dithiothreitol (DTT), dramatically decreasing the amounts of GSH conjugates formed, suggesting that a cysteine or methionine residue is involved. Consequently, the roles of the two cysteines (Cys-13 and Cys-156) in TCHQ dehalogenase were investigated by analyzing the properties of site-specific mutants (C13S and C156S).[134] From these studies it can be concluded that Cys-13 is essential for the reductive dehalogenation of TCHQ. Although the rate of TCHQ conversion is similar for the C13S mutant- and the wild-type enzyme-catalyzed reactions, the C13S mutant converts TCHQ only into the GSH conjugates GS-TriCHQ and GS-DCHQ and does not produce DCHQ. Cys-156 is not important for catalysis. Replacing this residue with a serine results in a mutant enzyme that has the same properties as the wild-type dehalogenase.

Additional support for the proposed mechanism was provided by rapid-flow quench experiments that demonstrated the existence of a covalent adduct between GSH and TCHQ dehalogenase during turnover of TriCHQ.[137] Reaction mixtures containing TCHQ dehalogenase, [^{3}H]-GSH, and TriCHQ were quenched at variable times, and then subjected to multiple cycles of concentration and dilution by ultrafiltration to remove the excess free GSH. Analysis of the protein samples by scintillation counting indicated that an intermediate, the covalent adduct between GSH and the enzyme, accumulated to a maximum concentration at about 500 ms and later decreased. The nature of the covalent adduct was explored further using tryptic digestion and mass spectrometric analysis of the resulting peptides. This analysis revealed that the covalent adduct involves a disulfide bond formed between Cys-13 and GSH. Hence, it was concluded that Cys-13 attacks the sulfur of the

glutathionyl substituent of GS-DCHQ*, releasing the reduced product, 2,6-DCHQ, and forming a covalent mixed disulfide between the enzyme and GSH (Enz-S-SG).[137]

The mechanism of the first part of the reaction, formation of GS-DCHQ* (**Figure 20**), is less certain and rather complex. Earlier studies suggested a mechanism that involved nucleophilic aromatic substitution in the first step of the reaction.[134] In this mechanism the GSH thiolate serves as the nucleophile and directly attacks the aromatic substrate, TriCHQ, rather than on the nonaromatic TriCHQ*. The results of more recent mechanistic studies on the wild-type enzyme and the C13S mutant enzyme were used to argue against the possibility of a nucleophilic aromatic substitution reaction.[138] Instead, at that time it was believed that the reaction starts with ketonization of TriCHQ⁻, followed by 1,4-elimination of HCl to form dichlorobenzoquinone (DCBQ) (**Figure 21**).[139] Subsequently, DCBQ is attacked by GSH to form GS-DCHQ*.

Pre-steady-state kinetic studies led to the mechanism shown in **Figure 20**.[136] These studies provide evidence for the accumulation of discrete intermediates during turnover. The identities of these intermediates are uncertain, but it appears that binding and deprotonation of TriCHQ occur during a fast initial step and that the formation of TriCHQ* occurs in a slower subsequent step. In fact, the pre-steady-state rate for the initial (dehalogenation) part of the reaction (i.e., production of DCHQ from TriCHQ) is quite fast (~25 s^{-1} when the enzyme is saturated with TriCHQ and GSH), whereas the steady-state rate of formation of DCHQ is much slower (~0.1 s^{-1} in the presence of 200 μmol l^{-1} TriCHQ).[136] Intriguingly, the steady-state rate is limited by the slow rate of the thiol–disulfide exchange reaction required to regenerate the free enzyme after substrate turnover. Although this exchange reaction is intrinsically less difficult than the dehalogenation reaction and is estimated to occur with a k_{cat} of about 103 s^{-1} in the absence of an aromatic substrate, it is hampered by premature binding of TriCHQ (or TCHQ) to the active site prior to completion of the catalytic cycle.[136,140] Substrate inhibition studies have shown that TriCHQ (or TCHQ) binds to the Enz-S-SG form of the enzyme and prevents thiol–disulfide exchange by acting as a noncompetitive inhibitor.[141] In addition, the dissociation of the GSSG product from the active site is inhibited by the hydroquinone substrates. This severe inhibition of TCHQ dehalogenase by its aromatic substrates could be interpreted as evidence for a recent evolutionary origin of TCHQ dehalogenase. However, Warner and Copley[141] suggested that this substrate inhibition is physiologically irrelevant, since the metabolite flux through the pentachlorophenol pathway is limited by the poor performance of the first enzyme of the pathway, pentachlorophenol hydroxylase, which converts
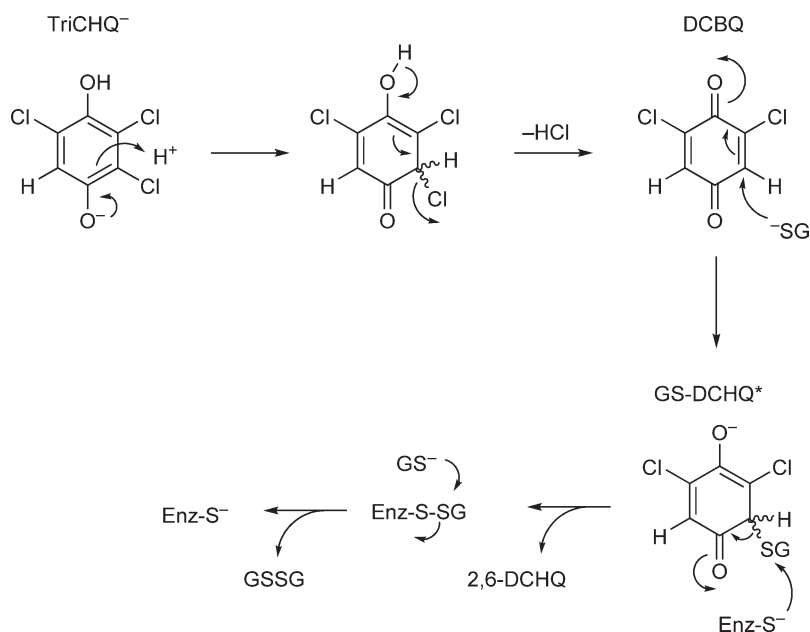
Figure 21 A schematic representation of a mechanism for TCHQ dehalogenase, showing ketonization, 1,4-elimination of HCl, and attack of GSH on DCBQ as the initial steps in the reaction.[138,139]

pentachlorophenol into TCHQ. The consequence of the metabolic flux being limited by the poor performance of the hydroxylase is a low concentration (in the low micromolar range) of TCHQ and TriCHQ in the cytoplasm of pentachlorophenol metabolizing cells.[142] At these low *in vivo* concentrations, inhibition of TCHQ dehalogenase by its aromatic substrates does not occur, and thus there is no selective pressure to evolve an enzyme that is not subject to substrate inhibition.

8.04.5 Conclusions

8.04.5.1 Diverse Catalytic Strategies for Carbon–Halogen Bond Cleavage

The examples of microbial dehalogenases discussed here clearly show that these enzymes use a variety of fundamentally different catalytic mechanisms to cleave carbon–halogen bonds. The best studied group of microbial dehalogenases is the one formed by the hydrolytic enzymes, for which two catalytic strategies have been observed. The most common approach employs an active site carboxylate to displace the halide and form an alkyl– or aryl–enzyme intermediate, followed by hydrolysis of the intermediate. This strategy of covalent catalysis is used by haloalkane dehalogenases, group II HADs, 4-chlorobenzoyl-CoA dehalogenase, and fluoroacetate dehalogenase. The second mechanistic approach used by the hydrolytic dehalogenases is the general base- or metal-catalyzed attack of water to displace the halide. This strategy of noncovalent catalysis is employed by the 3-chloroacrylic acid dehalogenases, AtzA, and presumably several group I HADs. The multiple occurrences of each catalytic strategy implies that attack of a carboxylate group and general base/metal-catalyzed attack of water are both effective ways to displace the halide ion.

The mechanism and energetic requirements for the aliphatic substitution reactions catalyzed by haloalkane dehalogenases and HADs are considerably different from those for the nucleophilic aromatic or vinylic substitutions catalyzed by 4-chlorobenzoyl-CoA dehalogenases, AtzA, and the 3-chloroacrylic acid dehalogenases, respectively. The hydrolytic dehalogenation reactions involving aromatic or vinylic substrates require the presence of a group on the substrate molecule that, when working in concert with the enzyme active site, can act as an electron sink. For example, the electron-withdrawing capabilities of the thioester portion of the substrate for 4-chlorobenzoyl-CoA dehalogenase are enhanced by the oxyanion hole (i.e., the polarizing interactions at the thioester carbonyl) provided by the enzyme. In the case of AtzA, the electron sink is provided by the N−1 nitrogen of the heteroaromatic ring, which can be protonated, effectively neutralizing the negative charge created by attack of the activated water molecule on the triazine ring. In the case of the 3-chloroacrylic acid dehalogenases, the electron sink is provided by the carboxylate group of the substrate working in concert with two active site arginine residues, which probably neutralize the negative charge that develops on one of the carboxylate oxygens upon attack of the hydroxyl on the vinylic substrate. In the absence of a mechanism for accommodating the negative charge formed by attack of the nucleophile, hydrolytic dehalogenation of aromatic or vinyl substrates is greatly reduced.

Another notable difference between the active sites of the haloalkane and HADs and those of the 4-chlorobenzoyl-CoA and 3-chloroacrylic acid dehalogenases is the absence of a halide binding site in the latter enzymes. In the haloalkane and HADs, the halide binding site promotes nucleophilic attack by polarizing the carbon–halogen bond, inducing a partial positive charge at the halogen-bearing carbon atom. In 4-chlorobenzoyl-CoA dehalogenase and the 3-chloroacrylic acid dehalogenases such a halide binding site may not be needed because the strong polarizing forces at the respective thioester carbonyl or carboxylate group of the conjugated substrate are transmitted to the halogen-bearing carbon atom, reducing the electron density at this atom, thereby activating it for nucleophilic attack. Thus, hydrolytic dehalogenation of these aromatic or vinylic substrates seems not to be dependent on extensive interactions with the halogen or halide ion.

Substitution reactions are also used in a different mechanistic context. Haloalcohol dehalogenases catalyze an unusual one-step intramolecular substitution reaction, where the chlorine atom is displaced by a vicinal hydroxyl group. This reaction does not involve the formation of a covalent enzyme–substrate intermediate. Another interesting approach is the use of GSH as nucleophile to displace the halide and form a glutathionyl intermediate. This strategy is employed by the GSH *S*-transferase-like enzymes DcmA and TCHQ dehalogenase. Further transformation of the glutathionyl intermediate is either an enzymatic (in the case of TCHQ dehalogenase) or a nonenzymatic process (in the case of DcmA). Whether these substitution reactions require

GSH as nucleophile (rather than a hydroxyl or carboxylate oxygen) for a mechanistic reason or whether the use of GSH simply reflects the fact that catalytically promiscuous GSH *S*-transferase superfamily members were recruited to serve as dichloromethane and TCHQ dehalogenases because they fortuitously had a useful level of dehalogenase activity is presently unknown.

8.04.5.2   Challenges for the Future

Detailed mechanistic studies and structural analyses of microbial dehalogenases have only been carried out for a few enzymes, and many more fascinating dehalogenases await the determination of their mechanisms and structures. The mechanisms of hydrolytic dehalogenases that do not involve covalent catalysis and those of nonhydrolytic dehalogenases need to be addressed more thoroughly. A notable gap in these studies is that crystal structures for γ-HCH dehydrochlorinase, DcmA, TCHQ dehalogenase, DL-2-HAD, fluoroacetate dehalogenase, and AtzA have not been solved. Placing the mechanistic studies in a structural context will undoubtedly give a better idea of how the pieces fit together.

One new area that deserves attention is the characterization of the large number of unexplored putative dehalogenase sequences in genomic databases. Microbial genome sequencing projects and the massive random sequencing of environmental DNA have shown that there is a large diversity of protein sequences that based on sequence similarities to known dehalogenating enzymes might be classified as putative dehalogenases. As these proteins may have novel functions and selectivities, cloning (or synthesis of) the corresponding gene sequences is expected to broaden the scope of available dehalogenases that may be useful for biocatalytic and enzyme-based waste treatment processes and in constructing new strains of bacteria for bioremediation. The current need for better genome sequence annotation and the promise of new and different dehalogenase enzymes make this work a priority. Importantly, the results of metagenome sequencing studies, coupled with insights obtained from molecular techniques, will also be critical to guide efforts to cultivate environmentally significant dehalogenating organisms.

A second area that merits attention is the design of new cell and protein catalysts using the existing dehalogenase sequences. Recent developments in directed evolution techniques may facilitate the construction of enzymes and organisms with degradation capacities that are not easily obtained by classical adaptation and enrichment. Studies with haloalkane dehalogenases have shown that enzymes with better activity toward alternative substrates can be obtained both by structure-based rational engineering and by random mutagenesis in combination with screening or selection.[143] A particularly important finding is that the activity and substrate specificity of these bacterial enzymes can be significantly influenced by engineering the size and shape of their entrance tunnels, which facilitate the transport of substrate from the solvent to the occluded active site.[144] A striking example of the directed evolution of a bacterial enzyme is the engineering of haloalcohol dehalogenase HheC with the potential for use in the manufacture of ethyl (*R*)-4-cyano-3-hydroxybutyrate, the starting material for the production of the cholesterol-lowering drug atorvastatin (Lipitor).[145] The 4000-fold improvement in the performance of HheC was required to meet the practical design criteria for this commercially relevant biocatalytic process, and was obtained by variants that had at least 35 mutations. Such directed evolution approaches should be exploited with other dehalogenases as well. This will not only give important insight into structure–function relationships, but the engineered enzymes with enhanced activity or different substrate specificities may also be useful for biocatalytic and bioremediation processes.

Finally, the origin and distribution of dehalogenase activities present a challenging area of research. Although some of the well-characterized dehalogenases belong to the same superfamily of enzymes, a striking observation is that none of the known dehalogenases are closely related to each other. Hence, each of these enzymes appears to have evolved independently from a pre-existing (superfamily) protein scaffold that provided an appropriate array of catalytic residues for the needed activity in the bacteria's environment. Owing to the common evolutionary history of the respective dehalogenases with members of the same superfamily, analysis of the superfamilies has been very informative for structural and functional studies on dehalogenases. However, it remains unknown to what degree the current functional dehalogenases differ from their recent evolutionary ancestors. Only for AtzA, a short evolutionary pathway has been suggested that could explain the rapid evolution of a functional dehalogenase from a closely related precursor that acts on a different

(nonhalogenated) substrate.[5,126] In the absence of a closely related sequence, one can only guess at the origins of the dehalogenase activities.

Acknowledgments

Abbreviations

3-CBA	3-chloro-2-butenoic acid
4-CBA-CoA	4-chlorobenzoyl-coenzyme A
4-FBA-CoA	4-fluorobenzoyl-CoA
4-HBA-CoA	4-hydroxybenzoate-CoA
4-NBA-CoA	4-nitrobenzoyl-CoA
atrazine	2-chloro-4-*N*-ethylamino-6-*N*-isopropylamino-1,3,5-triazine
AtzA	atrazine chlorohydrolase
BLAST	Basic Local Alignment Search Tool
CaaD	*trans*-3-chloroacrylic acid dehalogenase
CFCs	chlorofluorocarbons
***cis*-CaaD**	*cis*-3-chloroacrylic acid dehalogenase
CoA	coenzyme A
CPK	Corey–Pauling–Koltun
DCBQ	dichlorobenzoquinone
DCHQ	2,6-dichlorohydroquinone
DcmA	dichloromethane dehalogenase
DDTs	dichlorodiphenyltrichloroethanes
DehH1	fluoroacetate dehalogenase from *Moraxella* sp. B
DhaA	haloalkane dehalogenase from *Pseudomonas pavonaceae* 170
DhlA	haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10
DhlB	L-2-haloacid dehalogenase from *X. autotrophicus* GJ10
DL-DEX 113	DL-2-haloacid dehalogenase from *Pseudomonas* sp. 113
DTT	dithiothreitol
Enz-S-SG	mixed disulfide between enzyme and glutathione
GS-DCHQ	dichloro-*S*-glutathionylhydroquinone
GSH	glutathione
GSSG	glutathione disulfide
GS-TriCHQ	2,3,5-trichloro-6-*S*-glutathionylhydroquinone
HAD	haloacid dehalogenase
HCH	hexachlorocyclohexane
HheA	haloalcohol dehalogenase from *Arthrobacter* sp. strain AD2
HheC	haloalcohol dehalogenase from *Agrobacterium radiobacter* AD1
HSQC	heteronuclear single quantum coherence
L-DEX	L-2-haloacid dehalogenase from *Pseudomonas* sp. YL
LinA	γ-hexachlorocyclohexane dehydrochlorinase from *Sphingomonas paucimobilis* UT26
LinB	haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26
Melamine	2,4,6-triamino-1,3,5-triazine
NMR	nuclear magnetic resonance

PCBs	polychlorinated biphenyls
PNPAE	(*R*)-1-*para*-nitro-phenyl-2-azido-ethanol
PNSHH	*para*-nitro-2-bromo-1-phenylethanol
PNSO	*para*-nitrostyrene oxide
PSI-BLAST	position-specific iterative BLAST
SDR	short-chain dehydrogenases/reductases
S$_N$2	bimolecular nucleophilic substitution
S$_N$Ar	nucleophilic aromatic substitution
TCHQ	tetrachlorohydroquinone
TriA	melamine deaminase
TriCHQ	trichlorohydroquinone
UV	ultraviolet

Nomenclature

μ**mol l^{-1}**	micromolar
^{15}N	nitrogen-15
^{16}O	oxygen-16
^{18}O	oxygen-18
^{19}F	fluoride-19
^{1}H	hydrogen-1
^{2}H or D	deuterium
^{3}H	hydrogen-3
D$_2$O	heavy water
G	Gibbs free energy
kcal mol^{-1}	kilocalorie per mole
k_{cat}	turnover number
k_{cat}/K_m	catalytic efficiency
kDa	kilodalton
kJ mol^{-1}	kilojoule per mole
K_m	Michaelis constant
mmol l^{-1}	millimolar
mol	mole
mol^{-1} l s^{-1}	per molar per second
s	second
s^{-1}	per second

References

1. G. W. Gribble, *Chemosphere* **2003**, *52*, 289–297.
2. K. Ballschmiter, *Chemosphere* **2003**, *52*, 313–324.
3. B. Hileman, *Chem. Eng. News* **1993**, *71*, 11–20.
4. D. B. Janssen; I. J. T. Dinkla; G. J. Poelarends; P. Terpstra, *Environ. Microbiol.* **2005**, *7*, 1868–1882.
5. L. P. Wackett, *J. Biol. Chem.* **2004**, *279*, 41259–41262.
6. S. D. Copley, *Curr. Opin. Chem. Biol.* **1998**, *2*, 613–617.
7. S. D. Copley, *Trends Biochem. Sci.* **2000**, *25*, 261–265.

8. S. Keuning; D. B. Janssen; B. Witholt, *J. Bacteriol.* **1985**, *163*, 635–639.
9. D. B. Janssen; A. Scheper; L. Dijkhuizen; B. Witholt, *Appl. Environ. Microbiol.* **1985**, *49*, 673–677.
10. D. B. Janssen; F. Pries; J. van der Ploeg; B. Kazemier; P. Terpstra; B. Witholt, *J. Bacteriol.* **1989**, *171*, 6791–6799.
11. A. J. van den Wijngaard; K. W. van der Kamp; J. van der Ploeg; F. Pries; B. Kazemier; D. B. Janssen, *Appl. Environ. Microbiol.* **1992**, *58*, 976–983.
12. G. J. Poelarends; J. E. T. van Hylckama Vlieg; T. Bosma; D. B. Janssen, The Haloalkane Dehalogenase Genes *dhlA* and *dhaA* are Globally Distributed and Highly Conserved. In *Biotechnology for the Environment: Strategy and Fundamentals*; S. N. Agathos, W. Reineke, Eds.; Focus on Biotechnology Series; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2002; Vol. 3A, pp 59–66.
13. G. J. Poelarends; M. Wilkens; M. J. Larkin; J. D. van Elsas; D. B. Janssen, *Appl. Environ. Microbiol.* **1998**, *64*, 2931–2936.
14. G. J. Poelarends; W. H. Johnson, Jr.; A. G. Murzin; C. P. Whitman, *J. Biol. Chem.* **2003**, *278*, 48674–48683.
15. Y. Nagata; T. Nariya; R. Ohtomo; M. Fukuda; K. Yano; M. Takagi, *J. Bacteriol.* **1993**, *175*, 6403–6410.
16. G. J. Poelarends; J. E. T. van Hylckama Vlieg; J. R. Marchesi; L. M. Freitas dos Santos; D. B. Janssen, *J. Bacteriol.* **1999**, *181*, 2050–2058.
17. G. J. Poelarends; L. A. Kulakov; M. J. Larkin; J. E. T. van Hylckama Vlieg; D. B. Janssen, *J. Bacteriol.* **2000**, *182*, 2191–2199.
18. G. J. Poelarends; M. Zandstra; T. Bosma; L. A. Kulakov; M. J. Larkin; J. R. Marchesi; A. J. Weightman; D. B. Janssen, *J. Bacteriol.* **2000**, *182*, 2725–2731.
19. D. B. Janssen; J. Gerritse; J. Brackman; C. Kalk; D. Jager; B. Witholt, *Eur. J. Biochem.* **1988**, *171*, 67–72.
20. Y. Nagata; K. Miyauchi; J. Damborský; K. Manova; A. Ansorgova; M. Takagi, *Appl. Environ. Microbiol.* **1997**, *63*, 3707–3710.
21. J. Kmuníček; K. Hynková; T. Jedlicka; Y. Nagata; A. Negri; F. Gago; R. C. Wade; J. Damborský, *Biochemistry* **2005**, *44*, 3390–3401.
22. J. Damborský; E. Rorije; A. Jesenská; Y. Nagata; W. J. Peijnenburg, *Environ. Toxicol. Chem.* **2001**, *20*, 2681–2689.
23. A. Jesenská; M. Bartos; V. Czerneková; I. Rychlík; I. Pavlík; J. Damborský, *Appl. Environ. Microbiol.* **2002**, *68*, 3724–3730.
24. Y. Sato; M. Monincová; R. Chaloupková; Z. Prokop; Y. Ohtsubo; K. Minamisawa; M. Tsuda; J. Damborský; Y. Nagata, *Appl. Environ. Microbiol.* **2005**, *71*, 4372–4379.
25. A. Jesenská; M. Pavlová; M. Strouhal; R. Chaloupková; I. Tesínská; M. Monincová; Z. Prokop; M. Bartos; I. Pavlík; I. Rychlík; P. Möbius; Y. Nagata; J. Damborský, *Appl. Environ. Microbiol.* **2005**, *71*, 6737–6745.
26. S. M. Franken; H. J. Rozeboom; K. H. Kalk; B. W. Dijkstra, *EMBO J.* **1991**, *10*, 1297–1302.
27. D. B. Janssen, *Curr. Opin. Chem. Biol.* **2004**, *8*, 150–159.
28. K. H. Verschueren; F. Seljée; H. J. Rozeboom; K. H. Kalk; B. W. Dijkstra, *Nature* **1993**, *363*, 693–698.
29. F. Pries; J. Kingma; M. Pentenga; G. van Pouderoyen; C. M. Jeronimus-Stratingh; A. P. Bruins; D. B. Janssen, *Biochemistry* **1994**, *33*, 1242–1247.
30. F. Pries; J. Kingma; G. H. Krooshof; C. M. Jeronimus-Stratingh; A. P. Bruins; D. B. Janssen, *J. Biol. Chem.* **1995**, *270*, 10405–10411.
31. K. H. Verschueren; J. Kingma; H. J. Rozenboom; K. H. Kalk; D. B. Janssen; B. W. Dijkstra, *Biochemistry* **1993**, *32*, 9031–9037.
32. G. H. Krooshof; E. M. Kwant; J. Damborský; J. Koca; D. B. Janssen, *Biochemistry* **1997**, *36*, 9571–9580.
33. C. Kennes; F. Pries; G. H. Krooshof; E. Bokma; J. Kingma; D. B. Janssen, *Eur. J. Biochem.* **1995**, *228*, 403–407.
34. G. H. Krooshof; I. S. Ridder; A. W. Tepper; G. J. Vos; H. J. Rozeboom; K. H. Kalk; B. W. Dijkstra; D. B. Janssen, *Biochemistry* **1998**, *37*, 15013–15023.
35. J. P. Schanstra; J. Kingma; D. B. Janssen, *J. Biol. Chem.* **1996**, *271*, 14747–14753.
36. J. P. Schanstra; D. B. Janssen, *Biochemistry* **1996**, *35*, 5624–5632.
37. J. P. Schanstra; I. S. Ridder; G. J. Heimeriks; R. Rink; G. J. Poelarends; K. H. Kalk; B. W. Dijkstra; D. B. Janssen, *Biochemistry* **1996**, *40*, 13186–13195.
38. G. H. Krooshof; R. Floris; A. W. J. W. Tepper; D. B. Janssen, *Protein Sci.* **1999**, *8*, 355–360.
39. B. Schneider; R. Muller; R. Frank; F. Lingens, *J. Bacteriol.* **1991**, *173*, 1530–1535.
40. J. van der Ploeg; G. van Hall; D. B. Janssen, *J. Bacteriol.* **1991**, *173*, 7925–7933.
41. H. Kawasaki; K. Tsuda; I. Matsushita; K. Tonomura, *J. Gen. Microbiol.* **1992**, *138*, 1317–1323.
42. H. Kawasaki; T. Toyama; T. Maeda; H. Nishino; K. Tonomura, *Biosci. Biotechnol. Biochem.* **1994**, *58*, 160–163.
43. V. Nardi-Dei; T. Kurihara; T. Okamura; J. Q. Liu; H. Koshikawa; H. Ozaki; Y. Terashima; N. Esaki; K. Soda, *Appl. Environ. Microbiol.* **1994**, *60*, 3375–3380.
44. K. E. Hill; J. R. Marchesi; A. J. Weightman, *J. Bacteriol.* **1999**, *181*, 2535–2547.
45. T. Hisano; Y. Hata; T. Fujii; J.-Q. Liu; T. Kurihara; N. Esaki; K. Soda, *J. Biol. Chem.* **1996**, *271*, 20322–20330.
46. I. S. Ridder; H. J. Rozenboom; K. H. Kalk; D. B. Janssen; B. W. Dijkstra, *J. Biol. Chem.* **1997**, *272*, 33015–33022.
47. T. Kurihara; J. Q. Liu; V. Nardi-Dei; H. Koshikawa; N. Esaki; K. Soda, *J. Biochem.* **1995**, *117*, 1317–1322.
48. J.-Q. Liu; T. Kurihara; M. Miyagi; N. Esaki; K. Soda, *J. Biol. Chem.* **1995**, *270*, 18309–18312.
49. J.-Q. Liu; T. Kurihara; M. Miyagi; S. Tsunasawa; M. Nishihara; N. Esaki; K. Soda, *J. Biol. Chem.* **1997**, *272*, 3363–3368.
50. Y.-F. Li; Y. Hata; T. Fujii; T. Hisano; M. Nishihara; T. Kurihara; N. Esaki; K. Soda, *J. Biol. Chem.* **1998**, *273*, 15035–15044.
51. I. S. Ridder; H. J. Rozenboom; K. H. Kalk; B. W. Dijkstra, *J. Biol. Chem.* **1999**, *274*, 30672–30678.
52. V. Nardi-Dei; T. Kurihara; C. Park; N. Esaki; K. Soda, *J. Bacteriol.* **1997**, *179*, 4232–4238.
53. V. Nardi-Dei; T. Kurihara; C. Park; M. Miyagi; S. Tsunasawa; K. Soda; N. Esaki, *J. Biol. Chem.* **1999**, *274*, 20977–20981.
54. H. Kawasaki; H. Yahara; K. Tonomura, *Agric. Biol. Chem.* **1984**, *48*, 2627–2632.
55. R. Rink; M. Fennema; M. Smids; U. Dehmel; D. B. Janssen, *J. Biol. Chem.* **1997**, *272*, 14650–14657.
56. J.-Q. Liu; T. Kurihara; S. Ichiyama; M. Miyagi; S. Tsunasawa; H. Kawasaki; K. Soda; N. Esaki, *J. Biol. Chem.* **1998**, *273*, 30897–30902.
57. P. H. Liang; G. Yang; D. Dunaway-Mariano, *Biochemistry* **1993**, *32*, 12245–12250.
58. J. D. Scholten; K.-H. Chang; P. C. Babbitt; H. Charest; M. Sylvestre; D. Dunaway-Mariano, *Science* **1991**, *253*, 182–185.
59. P. C. Babbitt; G. L. Kenyon; B. M. Martin; H. Charest; M. Sylvestre; J. D. Scholten; K. H. Chang; P. H. Liang; D. Dunaway-Mariano, *Biochemistry* **1992**, *31*, 5594–5604.
60. K.-H. Chang; P.-H. Liang; W. Beck; J. D. Scholten; D. Dunaway-Mariano, *Biochemistry* **1992**, *31*, 5605–5610.

61. D. Kohler-Staub; H.-P. E. Kohler, *J. Bacteriol.* **1989**, *171*, 1428–1434.
62. G. J. Poelarends; R. Saunier; D. B. Janssen, *J. Bacteriol.* **2001**, *183*, 4269–4277.
63. M. M. Benning; K. L. Taylor; R.-Q. Liu; G. Yang; H. Xiang; G. Wesenberg; D. Dunaway-Mariano; H. M. Holden, *Biochemistry* **1996**, *35*, 8103–8109.
64. G. Yang; P. H. Liang; D. Dunaway-Mariano, *Biochemistry* **1994**, *33*, 8527–8531.
65. K. L. Taylor; R.-Q. Liu; P. H. Liang; J. Price; D. Dunaway-Mariano; P. J. Tonge; J. Clarkson; P. R. Carey, *Biochemistry* **1995**, *34*, 13881–13888.
66. K. L. Taylor; H. Xiang; R.-Q. Liu; G. Yang; D. Dunaway-Mariano, *Biochemistry* **1997**, *36*, 1349–1361.
67. J. Clarkson; P. J. Tonge; K. L. Taylor; D. Dunaway-Mariano; P. R. Carey, *Biochemistry* **1997**, *36*, 10192–10199.
68. J. Dong; H. Xiang; L. Luo; D. Dunaway-Mariano; P. R. Carey, *Biochemistry* **1999**, *38*, 4198–4206.
69. L. Luo; K. L. Taylor; H. Xiang; Y. Wei; W. Zhang; D. Dunaway-Mariano, *Biochemistry* **2001**, *40*, 15684–15692.
70. G. Yang; R.-Q. Liu; K. L. Taylor; H. Xiang; J. Price; D. Dunaway-Mariano, *Biochemistry* **1996**, *35*, 10879–10885.
71. W. Zhang; Y. Wei; L. Luo; K. L. Taylor; G. Yang; D. Dunaway-Mariano; M. M. Benning; H. M. Holden, *Biochemistry* **2001**, *40*, 13474–13482.
72. J. Dong; P. R. Carey; Y. Wei; L. Luo; X. Lu; R.-Q. Liu; D. Dunaway-Mariano, *Biochemistry* **2002**, *41*, 7453–7463.
73. J. Dong; X. Lu; Y. Wei; L. Luo; D. Dunaway-Mariano; P. R. Carey, *Biochemistry* **2003**, *42*, 9482–9490.
74. D. Xu; Y. Wei; J. Wu; D. Dunaway-Mariano; H. Guo; Q. Cui; J. Gao, *J. Am. Chem. Soc.* **2004**, *126*, 13649–13658.
75. J. Wu; D. Xu; X. Lu; C. Wang; H. Guo; D. Dunaway-Mariano, *Biochemistry* **2006**, *45*, 102–112.
76. A. J. van den Wijngaard; P. T. Reuvekamp; D. B. Janssen, *J. Bacteriol.* **1991**, *173*, 124–129.
77. T. Nakamura; T. Nagasawa; F. Yu; I. Watanabe; H. Yamada, *Appl. Environ. Microbiol.* **1994**, *60*, 1297–1301.
78. P. E. Swanson, *Curr. Opin. Biotechnol.* **1999**, *10*, 365–369.
79. E. J. de Vries; D. B. Janssen, *Curr. Opin. Biotechnol.* **2003**, *14*, 414–420.
80. T. Nakamura; T. Nagasawa; F. Yu; I. Watanabe; H. Yamada, *Biochem. Biophys. Res. Commun.* **1991**, *180*, 124–130.
81. J. H. Spelberg; J. E. T. van Hylckama Vlieg; L. Tang; D. B. Janssen; R. M. Kellog, *Org. Lett.* **2001**, *3*, 41–43.
82. D. B. Janssen; M. Majerić-Elenkov; G. Hasnaoui; B. Hauer; J. H. Lutje Spelberg, *Biochem. Soc. Trans.* **2006**, *34*, 291–295.
83. R. M. Haak; C. Tarabiono; D. B. Janssen; A. J. Minnaard; J. G. de Vries; B. L. Feringa, *Org. Biomol. Chem.* **2006**, *5*, 318–323.
84. M. Majerić-Elenkov; L. Tang; B. Hauer; D. B. Janssen, *Org. Lett.* **2006**, *8*, 4227–4229.
85. F. Yu; T. Nakamura; W. Mizunashi; I. Watanabe, *Biosci. Biotechnol. Biochem.* **1994**, *58*, 1451–1457.
86. J. E. T. van Hylckama Vlieg; L. Tang; J. H. Lutje Spelberg; T. Smilda; G. J. Poelarends; T. Bosma; A. E. J. van Merode; M. W. Fraaije; D. B. Janssen, *J. Bacteriol.* **2001**, *183*, 5058–5066.
87. R. M. de Jong; J. J. W. Tiesinga; H. J. Rozenboom; K. H. Kalk; L. Tang; D. B. Janssen; B. W. Dijkstra, *EMBO J.* **2003**, *22*, 4933–4944.
88. R. M. de Jong; K. H. Kalk; L. Tang; D. B. Janssen; B. W. Dijkstra, *J. Bacteriol.* **2006**, *188*, 4051–4056.
89. H. Jörnvall; B. Persson; M. Krook; S. Atrian; R. Gonzalez Duarte; J. Jeffery; D. Ghosh, *Biochemistry* **1995**, *34*, 6003–6013.
90. L. Tang; J. H. Lutje Spelberg; M. W. Fraaije; D. B. Janssen, *Biochemistry* **2003**, *42*, 5378–5386.
91. L. Tang; A. E. J. van Merode; J. H. Lutje Spelberg; M. W. Fraaije; D. B. Janssen, *Biochemistry* **2003**, *42*, 14057–14065.
92. L. Tang; D. E. Torres Pazmino; M. W. Fraaije; R. M. de Jong; B. W. Dijkstra; D. B. Janssen, *Biochemistry* **2005**, *44*, 6609–6618.
93. R. M. de Jong; J. J. W. Tiesinga; A. Villa; L. Tang; D. B. Janssen; B. W. Dijkstra, *J. Am. Chem. Soc.* **2005**, *127*, 13338–13343.
94. S. Hartmans; M. W. Jansen; M. J. van der Werf; J. A. M. de Bont, *J. Gen. Microbiol.* **1991**, *137*, 2025–2032.
95. J. E. T. van Hylckama Vlieg; D. B. Janssen, *Biodegradation* **1992**, *2*, 139–150.
96. G. J. Poelarends; H. Serrano; M. D. Person; W. H. Johnson, Jr.; A. G. Murzin; C. P. Whitman, *Biochemistry* **2004**, *43*, 759–772.
97. S. C. Wang; M. D. Person; W. H. Johnson, Jr.; C. P. Whitman, *Biochemistry* **2003**, *42*, 8762–8773.
98. C. M. Horvat; R. V. Wolfenden, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16199–16202.
99. P. M. Jeffers; N. L. Wolfe, *Environ. Toxicol. Chem.* **1996**, *15*, 1066–1071.
100. S. A. Khan; A. J. Kirby, *J. Chem. Soc. B* **1970**, 1172–1182.
101. Z. Lei; M. Ye; X. J. Wang, *J. Environ. Sci. (China)* **2001**, *13*, 99–103.
102. G. J. Poelarends; C. P. Whitman, *Bioorg. Chem.* **2004**, *32*, 376–392.
103. G. J. Poelarends; W. H. Johnson, Jr.; H. Serrano; C. P. Whitman, *Biochemistry* **2007**, *46*, 9596–9604.
104. R. M. de Jong; W. Brugman; G. J. Poelarends; C. P. Whitman; B. W. Dijkstra, *J. Biol. Chem.* **2004**, *279*, 11546–11552.
105. A. G. Murzin, *Curr. Opin. Struct. Biol.* **1996**, *6*, 386–394.
106. C. P. Whitman, *Arch. Biochem. Biophys.* **2002**, *402*, 1–13.
107. H. F. Azurmendi; S. C. Wang; M. A. Massiah; G. J. Poelarends; C. P. Whitman; A. S. Mildvan, *Biochemistry* **2004**, *43*, 4082–4091.
108. R. M. de Jong; P. Bazzacco; G. J. Poelarends; W. H. Johnson, Jr.; Y. J. Kim; E. A. Burks; H. Serrano; A.-M. W. H. Thunnissen; C. P. Whitman; B. W. Dijkstra, *J. Biol. Chem.* **2007**, *282*, 2440–2449.
109. G. J. Poelarends; H. Serrano; W. H. Johnson, Jr.; C. P. Whitman, *Biochemistry* **2004**, *43*, 7187–7196.
110. G. J. Poelarends; J. J. Almrud; H. Serrano; J. E. Darty; W. H. Johnson, Jr.; M. L. Hackert; C. P. Whitman, *Biochemistry* **2006**, *45*, 7700–7708.
111. R. Imai; Y. Nagata; M. Fukuda; M. Takagi; K. Yano, *J. Bacteriol.* **1991**, *173*, 6811–6819.
112. Y. Nagata; T. Hatta; R. Imai; K. Kimbara; M. Fukuda; K. Yano; M. Takagi, *Biosci. Biotechnol. Biochem.* **1993**, *57*, 1582–1583.
113. L. Trantírek; K. Hynková; Y. Nagat; A. G. Murzin; A. Ansorgová; V. Sklenár; J. Damborský, *J. Biol. Chem.* **2001**, *276*, 7734–7740.
114. Y. Nagata; K. Mori; M. Takagi; A. G. Murzin; J. Damborský, *Proteins* **2001**, *45*, 471–477.
115. T. Leisinger; R. Bader; R. Hermann; M. Schmid-Appert; S. Vuilleumier, *Biodegradation* **1994**, *5*, 237–248.
116. S. Vuilleumier, Coping with a Halogenated One-Carbon Diet: Aerobic Dichloromethane-Mineralising Bacteria. In *Biotechnology for the Environment: Strategy and Fundamentals*; S. N. Agathos, W. Reineke, Eds.; Focus on Biotechnology Series; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2002; Vol. 3A, pp 105–130.
117. D. Kohler-Staub; T. Leisinger, *J. Bacteriol.* **1985**, *162*, 676–681.
118. R. Scholtz; L. P. Wackett; C. Egli; A. M. Cook; T. Leisinger, *J. Bacteriol.* **1988**, *170*, 5698–56704.
119. S. D. La Roche; T. Leisinger, *J. Bacteriol.* **1990**, *172*, 164–171.
120. R. Bader; T. Leisinger, *J. Bacteriol.* **1994**, *176*, 3466–3473.

121. G. Stucki; R. Gälli; H. R. Ebersold; T. Leisinger, *Arch. Microbiol.* **1981**, *130*, 366–371.
122. S. Vuilleumier; T. Leisinger, *Eur. J. Biochem.* **1996**, *239*, 410–417.
123. F. A. Blocki; M. S. Logan; C. Baoli; L. P. Wackett, *J. Biol. Chem.* **1994**, *269*, 8826–8830.
124. N. V. Stourman; J. H. Rose; S. Vuilleumier; R. N. Armstrong, *Biochemistry* **2003**, *42*, 11048–11056.
125. M. L. de Souza; M. J. Sadowsky; L. P. Wackett, *J. Bacteriol.* **1996**, *178*, 4894–4900.
126. J. L. Seffernick; L. P. Wackett, *Biochemistry* **2001**, *40*, 12747–12753.
127. M. J. Sadowsky; Z. Tong; M. de Souza; L. P. Wackett, *J. Bacteriol.* **1998**, *180*, 152–158.
128. J. L. Seffernick; H. McTavish; J. P. Osborne; M. L. de Souza; M. J. Sadowsky; L. P. Wackett, *Biochemistry* **2002**, *41*, 14430–14437.
129. L. Holm; C. Sander, *Proteins* **1997**, *28*, 72–82.
130. J. L. Seffernick; M. L. de Souza; M. J. Sadowsky; L. P. Wackett, *J. Bacteriol.* **2001**, *183*, 2405–2410.
131. S.-A. Raillard; A. Krebber; Y. Chen; J. E. Ness; E. Bermudez; R. Trinidad; R. Fullem; C. Davis; M. Welch; J. L. Seffernick; L. P. Wackett; W. P. C. Stemmer; J. Minshull, *Chem. Biol.* **2001**, *8*, 891–898.
132. L. Xun; E. Topp; C. S. Orser, *J. Bacteriol.* **1992**, *174*, 8003–8007.
133. L. Xun; E. Topp; C. S. Orser, *Biochem. Biophys. Res. Commun.* **1992**, *182*, 361–366.
134. D. L. McCarthy; S. Navarette; W. S. Willett; P. C. Babbitt; S. D. Copley, *Biochemistry* **1996**, *35*, 14634–14642.
135. K. Anandarajah; P. M. Kiefer, Jr.; B. S. Donohoe; S. D. Copley, *Biochemistry* **2000**, *39*, 5303–5311.
136. J. R. Warner; S. D. Copley, *Biochemistry* **2007**, *46*, 13211–13222.
137. D. L. McCarthy; D. F. Louie; S. D. Copley, *J. Am. Chem. Soc.* **1997**, *119*, 11337–11338.
138. P. M. Kiefer, Jr.; D. L. McCarthy; S. D. Copley, *Biochemistry* **2002**, *41*, 1308–1314.
139. P. M. Kiefer, Jr.; S. D. Copley, *Biochemistry* **2002**, *41*, 1315–1322.
140. J. R. Warner; S. L. Lawson; S. D. Copley, *Biochemistry* **2005**, *44*, 10360–10368.
141. J. R. Warner; S. D. Copley, *Biochemistry* **2007**, *46*, 4438–4447.
142. D. L. McCarthy; A. A. Claude; S. D. Copley, *Appl. Environ. Microbiol.* **1997**, *63*, 1883–1888.
143. T. Bosma; J. Damborský; G. Stucki; D. B. Janssen, *Appl. Environ. Microbiol.* **2002**, *68*, 3582–3587.
144. R. Chaloupková; J. Sýkorová; Z. Prokop; A. Jesenská; M. Monincová; M. Pavlová; M. Tsuda; Y. Nagata; J. Damborský, *J. Biol. Chem.* **2003**, *278*, 52622–52628.
145. R. J. Fox; S. C. Davis; E. C. Mundorff; L. M. Newman; V. Gavrilovic; S. K. Ma; L. M. Chung; C. Ching; S. Tam; S. Muley; J. Grate; J. Gruber; J. C. Whitman; R. A. Sheldon; G. W. Huisman, *Nat. Biotechnol.* **2007**, *25*, 338–344.

Biographical Sketches

Dr. Gerrit J. Poelarends received his Ph.D. degree from the University of Groningen, The Netherlands in 2001 under the supervision of Professor Dick B. Janssen on the topic of bacterial degradation of halogenated aliphatic compounds. He has worked as a postdoctoral fellow with Professor Wil N. Konings in the Microbiology Department at the University of Groningen and then with Professor Christian P. Whitman in the Division of Medicinal Chemistry at the University of Texas at Austin, USA. After a short postdoctoral period in the Biochemistry Department, University of Groningen, he was appointed as an assistant professor in Pharmaceutical Biotechnology at the same University in 2006. His research interests are in the field of biocatalysis with a focus on the discovery and design of new protein catalysts. He has (co)authored more than 50 publications and book chapters. He obtained the prestigious VENI-CW and VIDI-CW NWO grants for his research work in 2004 and 2006, respectively.

Dr. Christian P. Whitman received his Ph.D. degree from the University of California, San Francisco in 1984 under the supervision of Professor George L. Kenyon. At UCSF, he worked on the catalytic mechanism of mandelate racemase. This enzyme would later pique the interests of Professors John Gerlt, John Kozarich, Greg Petsko, and others, and expand our understanding of divergent evolution to include enzymes that catalyze different reactions but are derived from a common ancestor. Later he worked as a postdoctoral fellow with Professor Kozarich at the University of Maryland in College Park, where he tackled mechanistic and evolutionary questions about the enzymes comprising the β-ketoadipate pathway. He was appointed as an assistant professor in the Division of Medicinal Chemistry, College of Pharmacy at the University of Texas at Austin in 1987, subsequently he was promoted to the rank of professor and division head in 1998. His research interests focus on the answers to two fundamental questions: How do enzymes work and how do they evolve?

8.05 Guanidine-Modifying Enzymes in the Pentein Superfamily

Thomas W. Linsky and Walter Fast, University of Texas, Austin, TX, USA

8.05.1   Introduction to the Pentein Superfamily

The penteins are a functionally diverse superfamily of proteins grouped together because they share a similar structural motif. These proteins are important to a myriad of biological processes as varied as gene regulation, protein translation, cell–cell signaling, arginine metabolism, natural product biosynthesis, and bacterial survival mechanisms. The shared motif that unites these enzymes, called the pentein fold, consists of repeated $\alpha\beta\beta\alpha\beta$ subdomains arranged around a fivefold pseudosymmetrical axis.[1] Although the primary sequences of these five subdomains are not homologous, their structural homology supports the proposition that they derived from an ancient gene duplication.[2] The pentein fold has also been described as a β/α propeller, acknowledging its structural parallels to β-propeller proteins.[3] The pentameric arrangement of the five subdomains around a central axis results in a central hollow, a structural feature also conserved throughout the superfamily (**Figure 1**. All figures containing protein structures were constructed using the UCSF Chimera package from the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco.[4]).[1]

   Although proteins in this superfamily share the same fold, they are quite diverse in sequence and function, and are grouped into two main families. One family consists of noncatalytic proteins, and the remaining enzymes comprise a diverse family of guanidine-modifying enzymes.[5,6] The pentein fold is most easily seen in the noncatalytic proteins, in which the propeller fold is relatively unadorned by loop inserts and extensions. The catalytic proteins, however, carry more extensive structural elaborations, sometimes complicating auto-mated structure matching.[2] Although pentameric symmetry is less apparent in the enzymes, a conserved pentein fold serves as a unifying structural feature, suggesting that the catalytic and noncatalytic proteins derived from a common ancestor and can be classified as two distinct families within the pentein superfamily.[2] Enzymes within the guanidine-modifying family can be further subcategorized into at least three distinct
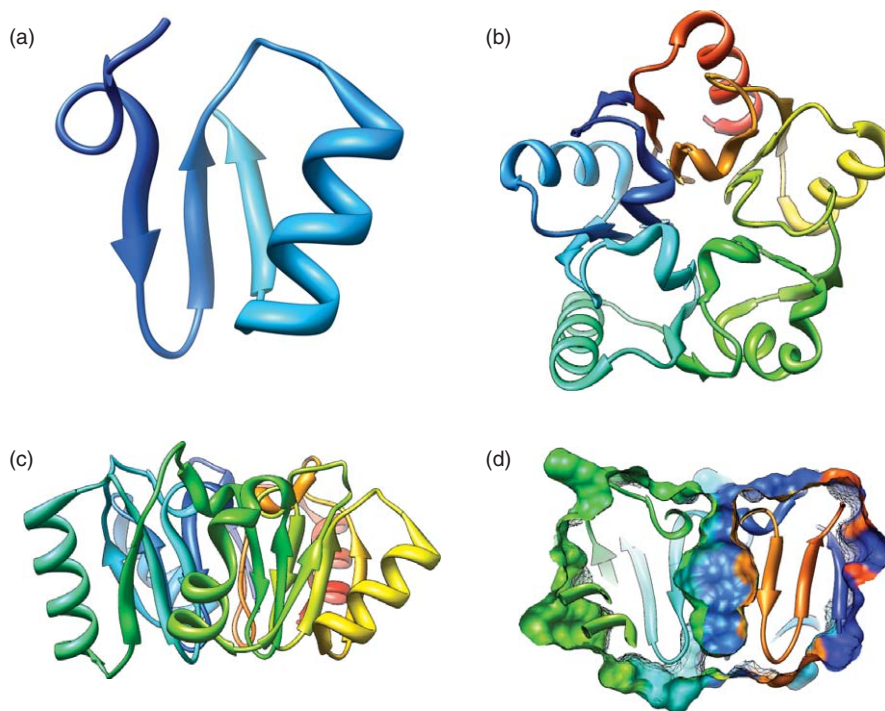
Figure 1   The pentein fold can be described as a β/α propeller. (a) Ribbon diagram of one $\alpha\beta\beta\alpha\beta$ subdomain (Ile2012 – Thr2051 of aIF6). (b) 'Top' view of a ribbon diagram showing pentameric pseudosymmetry of the β/α propeller. (c) 'Side' view of a ribbon diagram showing relatively flat surfaces on either side of the β/α propeller structure. (d) Cutaway 'side' view of a surface-coated ribbon diagram showing the central hollow. The structures above depict the structure of the ribosome antiassociation factor aIF6 from *Methanococcus jannaschii* (PDB accession code 1G61) and are rainbow-color coded from the N-terminus (blue) to the C-terminus (red).
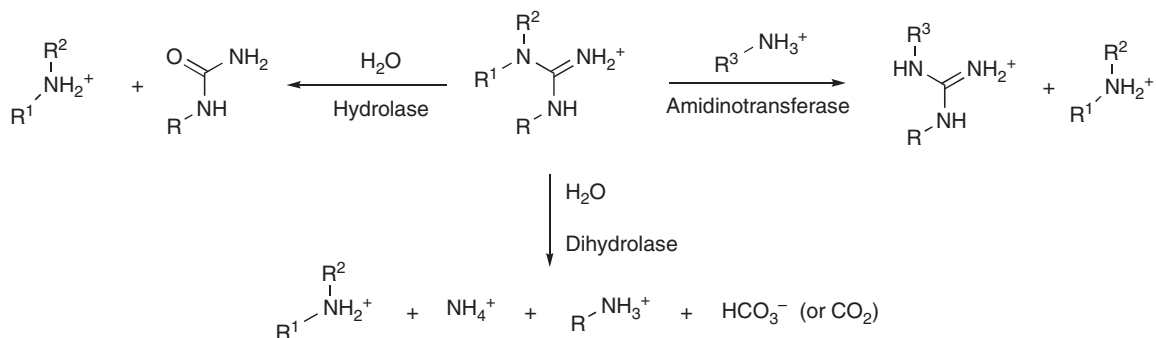
Figure 2 Three enzymatic activities observed in the guanidine-modifying enzyme family.

groups based here on their functions: hydrolases, dihydrolases, and amidinotransferases (**Figure 2**). Hydrolases catalyze the hydrolysis of guanidine derivatives to a ureido compound; dihydrolases catalyze the hydrolysis of a guanidine derivative to yield a primary amine, ammonia, and bicarbonate (or carbon dioxide); and amidino-transferases catalyze the transfer of an amidino group ($-C(=NH)NH_2$) from one substrate to an acceptor amine. This chapter categorizes proteins of the pentein superfamily into broad categories based on their function, and describe their biological use, structural characterization, catalytic mechanism, and regulation.

8.05.2 Noncatalytic Proteins

8.05.2.1 Ribosome Antiassociation Factors

8.05.2.1.1 Biological function

The noncatalytic proteins in the pentein superfamily are involved in regulating ribosome biogenesis and protein translation. Both known examples are ribosome antiassociation factors found in eukaryotes (eIF6) and in archaea (aIF6). These proteins bind to the 60S ribosomal subunit, preventing its association with the 40S subunit to form intact 80S ribosomes, and are thought to be required for proper pre-rRNA processing during the formation of 60S subunits.[7,8] eIF6 may be required for efficient initiation of translation and is the only known 60S-associated factor located downstream of extracellular growth signals. It has been proposed that eIF6 may be important in cell-cycle regulation, with a direct role in tumorigenesis.[9] In support of this model, reduced levels of eIF6 impair cell-cycle progression in mice,[9] and increased eIF6 levels are found in many cancers.[10,11] It may also participate in adhesion processes; human eIF6 can bind β_4 integrin. In addition to their role in ribosome formation and translation, a multiprotein complex including RISC (RNA-induced silencing complex) and eIF6 can mediate microRNA silencing in humans.[12] Posttranslational modifications of this protein are possible; the yeast isoform can be specifically phosphorylated on Ser174 by a casein kinase I α homolog.[13] Despite these numerous protein–protein interactions, the exact amino acids that mediate protein binding are not known. However, based on sequence conservation and protein structures, the interaction between *Saccharomyces cerevisiae* eIF6 and the 60S ribosomal subunit is thought to occur on the relatively flat C-terminal face of the protein.[1]

8.05.2.1.2 Protein structure

The structures of two ribosome antiassociation factors have been reported: aIF6 from the archaeon *Methanococcus jannaschii* and eIF6 from *S. cerevisiae*.[1] The β/α propeller fold unexpectedly shared by these proteins and those of the guanidine-modifying enzymes led to the subsequent naming of the pentein fold and the pentein superfamily.[1,2] The structures of these factors are the simplest expression of the pentein fold (**Figure 1**), and lack the long loop inserts and additional domains found in their catalytic relatives. Of note, the antiassociation factors have a solvent-filled cavity in the center of their structures. However, access to this cavity is somewhat restricted, the utility of this cavity to the noncatalytic proteins is not obvious. These noncatalytic proteins lack the conserved 'core' active-site residues found in catalytic members of this superfamily.

8.05.3   Hydrolases

Guanidine-modifying hydrolases catalyze the exchange of one guanidino nitrogen with an oxygen derived from water, forming a urea derivative. Hydrolases are the largest branch of the guanidine-modifying enzyme family, with biological roles ranging from energy production to cell signaling. There are four known types of these hydrolases: arginine deiminase (ADI), which catalyzes the hydrolysis of free arginine to form citrulline and ammonia; dimethylarginine dimethylaminohydrolase (DDAH), which catalyzes the hydrolysis of mono- and di-methylated derivatives of arginine to yield citrulline and an alkylamine; peptidylarginine deiminase (PAD), which catalyzes the hydrolysis of arginine residues found in proteins; and agmatine deiminase (AgDI), which catalyzes the conversion of agmatine into N-carbamoylputrescine. Although sequence identity is low among these enzymes, they retain the overall pentein fold and the 'core' catalytic machinery. The hydrolases share common mechanistic features, which include attack by an active-site Cys residue, cleavage of a guanidino C^ζ–N bond, formation of a covalent enzyme intermediate, and subsequent attack of a water (or hydroxide) molecule to form a urea derivative.

8.05.3.1   Arginine Deiminase

8.05.3.1.1   Biological function

One of the best-studied enzymes in the pentein superfamily is involved in arginine catabolism. Of the several different known microbial arginine degradation pathways, the ADI pathway (also called the arginine dihydrolase pathway) is the most widespread.[14,15] This pathway consists of three enzymes: ADI, which hydrolyzes arginine to citrulline (**1**) and ammonia; ornithine transcarbamoylase (OTC), which forms carbamoyl phosphate (**2**) and ornithine (**3**); and carbamate kinase (CK), which uses carbamoyl phosphate to phosphorylate ADP to yield ATP, ammonia, and bicarbonate or CO_2 (**Figure 3**). This is an energy-producing pathway; for every mole of arginine consumed, one mole of ATP is generated. A parallel phylogenetic analysis of ADI, OTC, and CK sequences suggests that the ADI pathway may have been assembled by enzyme recruitment.[14] In bacteria, genes for these enzymes usually appear in clusters, often along with a gene encoding an amino acid transporter.[14] ADI, the first enzyme of the pathway, is a member of the pentein superfamily. ADIs have been found in bacteria, archaea, and anaerobic eukaryotes, but not in humans or other higher eukaryotes. The ADI pathway is a major energy-producing pathway for several human pathogens and may contribute to the acid tolerance of others due to its ammonia production.[16] One notable example is the primitive eukaryote and human pathogen *Giardia intestinalis*, which uses the ADI pathway as a major energy source.[17,18] Therefore, ADI is emerging as an attractive target for antimicrobial drug design.

8.05.3.1.2   Protein structure

ADI enzymes from various bacteria have been structurally characterized. The ADI enzymes from *Mycoplasma arthritidis* and *Mycoplasma arginini* are homodimers and the ADI from *Pseudomonas aeruginosa* is a homotetra-mer.[19–21] ADI enzymes from *P. aeruginosa* and *M. arginini* have been crystallized while harboring different reaction intermediates.[22–24] The pentein fold is conserved in each monomer, but is elaborated by several insertions that disrupt the symmetry evident in the ribosome antiassociation factors (**Figure 4**).[23] The overall shape of ADI has been colorfully described as a 'clip-on fan' with the β/α propeller representing the fan blades

$R = (CH_2)_3CHNH_3^+COO^-$
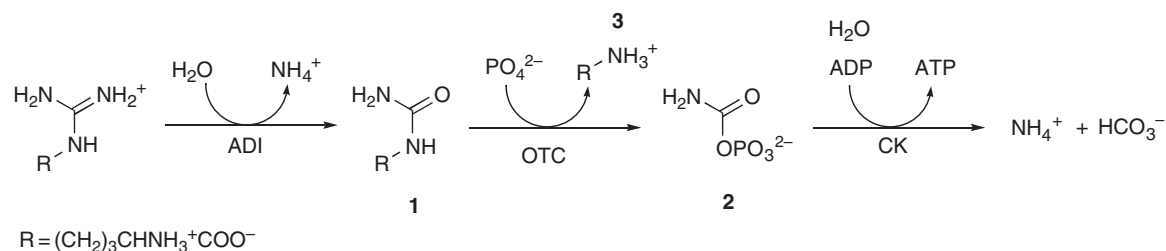
Figure 3   The arginine deiminase pathway consists of three enzymes: arginine deiminase (ADI), ornithine transcar-bamoylase (OTC), and carbamate kinase (CK).
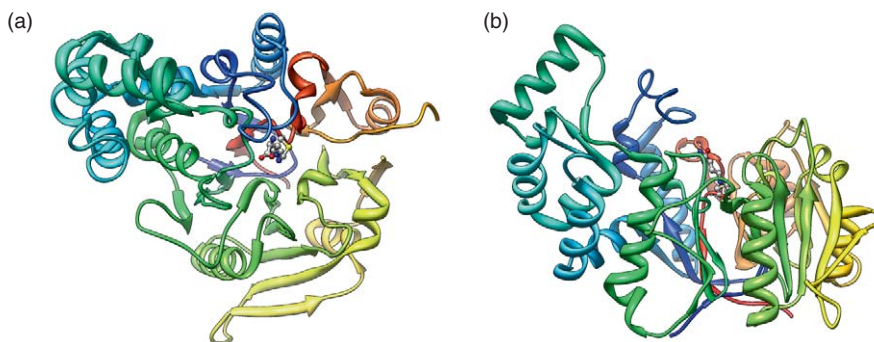
Figure 4   Structure of arginine deiminase. The pentein fold is visible in the 'top' (a) and 'side' (b) views, but is elaborated by a large insert in the first subdomain. A ball-and-stick model of one trapped reaction intermediate indicates that the substrate-binding site is located at the center of the propeller, near the face opposite the enzyme's N- and C-termini. These ribbon diagrams depict a H278A mutant of *Pseudomonas aeruginosa* ADI (PDB accession code 2AAF).
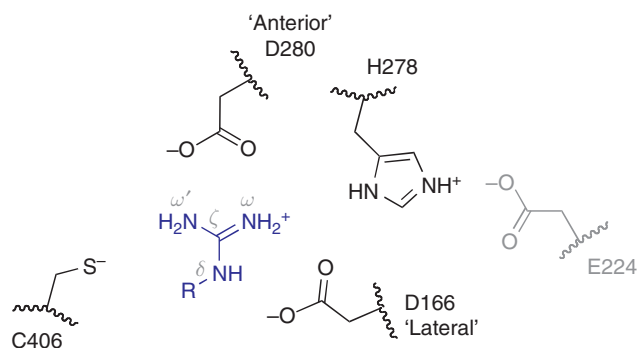
Figure 5   Active-site residues of ADI that interact with the guanidine of the substrate, L-arginine (shown in blue). 'Core' residues are shown in black and a neighboring Glu in gray. D166 is designated the 'lateral' Asp and D180 the 'anterior' Asp, with respect to the substrate's N^δ–C^ζ bond. Amino acid numbering is taken from *Pseudomonas aeruginosa* ADI.

and a large sequence insert as the 'clip'.[24] The substrate, L-arginine, binds at the center of the propeller structure, near the face opposite the enzyme's N- and C-termini.

Specific interactions of the substrate's guanidinium with two active-site carboxylate residues help to position this group for catalysis, with conserved Cys and His residues poised nearby (**Figure 5**). These conserved active-site residues have been dubbed 'core' residues[25] because they play a significant role in binding the guanidinium and catalyzing its hydrolysis; we will adopt this terminology here. The substrate's guanidinium is bound in a very polar environment, with its N^δ and N^ω atoms held by an interaction with an Asp (D166) positioned lateral to the N^δ–C^ζ bond of the substrate. This Asp presumably also modulates the electrophilicity of the substrate's guanidinium. The substrate's N^ω and $N^{\omega\prime}$ atoms are held by a second Asp (D280) located anterior to the substrate. Because most of the active-site residues are conserved throughout the superfamily, these two carboxylates will be termed herein the 'lateral' and 'anterior' Asp residues, with respect to the substrate's N^δ–C^ζ bond, to avoid confusion based on numbering differences. The Glu residue (E224) found within hydrogen-bonding distance to the core His residue is not absolutely conserved throughout the superfamily and does not directly contact the substrate, so it is not considered a core residue.

8.05.3.1.3   Catalytic mechanism

The detailed catalytic mechanisms for ADI enzymes from various bacteria have been reported, and the essential features are compiled here (**Figure 6**).[22–32] Most of the mechanistic details are conserved among isoforms. When the L-arginine substrate binds, the active site positions the amino acid moiety and carbon
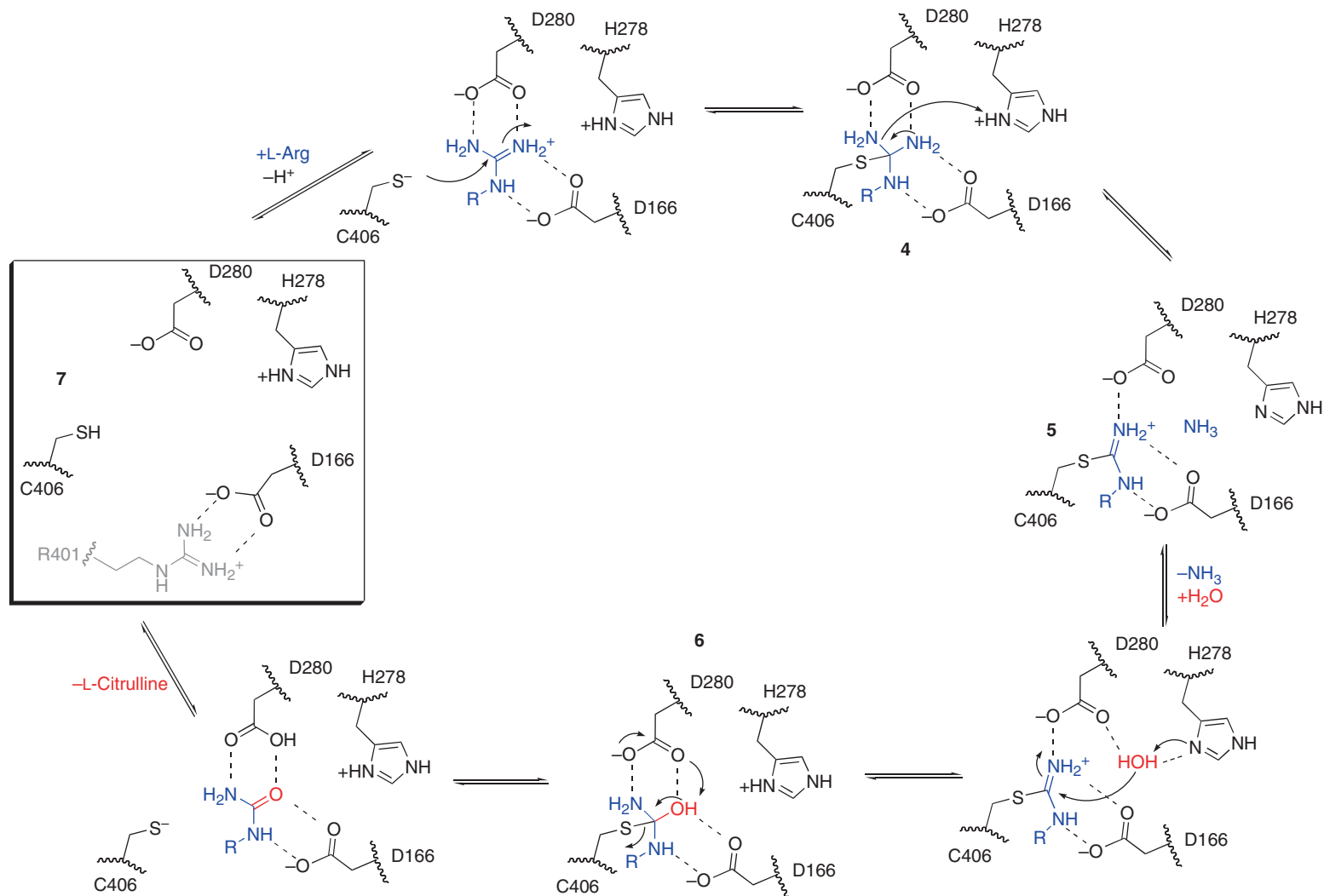
Figure 6 Proposed catalytic mechanism of ADI. Amino acid numbering and the autoinhibitory R401 (shown in gray) is from *Pseudomonas aeruginosa* ADI. Substrate L-arginine is shown in blue, 'core' residues in black, and the hydrolytic water in red. The apo form of the enzyme is boxed.

backbone of the substrate through polar and hydrophobic interactions, while the guanidinium is held by the core residues (**Figure 5**). Although most of the binding interactions are conserved, the ADI from *P. aeruginosa* shows a unique autoinhibitory conformation in the resting enzyme in which an arginine residue partially occludes the substrate-binding site and must swing away to allow the substrate binding.[22]

The core Cys and His residues at the active site are positioned on opposite faces of the bound guanidine. The Cys acts as a nucleophile and attacks the guanidino carbon (C^ζ), forming the first tetrahedral adduct (**4**). Hydrogen bonding and electrostatic interactions of one of the N^ω atoms with the anterior Asp and donation of a proton from the active-site cationic His residue promote C^ζ–N^ω bond cleavage and loss of NH_3 to form a planar *S*-alkyl thiouronium covalent intermediate (**5**). Interactions of the remaining N^δ and N^ω nitrogens of the intermediate with the lateral Asp are maintained. The anterior Asp maintains an interaction with the remaining N^ω nitrogen, but its second carboxylate oxygen is now available to bind to the water molecule that exchanges for the NH_3. The active-site His, now neutral, acts as a general base to deprotonate this water for attack on the C^ζ carbon of the planar thiouronium intermediate, forming the second tetrahedral adduct (**6**). Finally, collapse of this adduct and expulsion of the active-site Cys regenerate the resting enzyme (**7**). The importance of the active-site Cys has been demonstrated by structural studies, mutagenesis, and thiol-modifying reagents.[23–25,28,30] Dunaway–Mariano's group has used rapid quench techniques to demonstrate the formation and catalytic competence of the thiouronium covalent intermediate.[28] Crystal structures have been determined for the substrate-bound form, a tetrahedral intermediate (**8**), the thiouronium intermediate (**5**), and the active-site mutants, and are consistent with the interactions detailed in the proposed mechanism (**Figures 6** and **7**).[23,24]

Although the core Cys acts as a nucleophile in the reaction, there is growing evidence that this residue is predominantly protonated and therefore not very reactive in the resting state of the enzyme. Computational and functional studies indicate that the bound L-arginine substrate uses its positively charged guanidinium to stabilize the deprotonated anionic thiolate form of the core Cys, making it a more effective nucleophile.[27,32] This unusual mechanistic feature, an active-site Cys nucleophile that is predominantly protonated and relatively inert in the resting state of the enzyme, appears to be a mechanistic feature that is conserved throughout the enzymes in this superfamily.

8.05.3.1.4   Regulation, inhibitors, therapeutic application

ADI has been proposed to enable survival of some bacteria in acidic environments due to ammonia production.[16,33] Optimal activity of some ADI enzymes at acidic pH values is consistent with this proposal; ADIs from various sources show widely different pH rate dependencies, with optimums ranging from pH 5.6 to 7.5.[32,34] Low pH optimums may suggest a self-regulating mechanism that allows hydrolytic activity when needed, but would limit the enzyme's activity at neutral pH values to avoid unnecessary wasting of cellular arginine pools. The autoinhibition mechanism of *P. aeruginosa* ADI has also been suggested to serve a similar regulatory role in avoiding arginine wasting.[22,23]
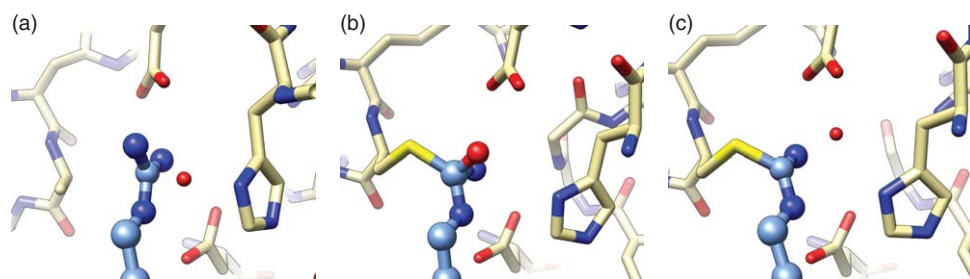
Figure 7   Three ADI structures with active-site ligands bound. (a) A C406A mutant of *Pseudomonas aeruginosa* ADI shows the substrate-binding orientation. (b) *Mycoplasma arginini* ADI complexed with a tetrahedral species, proposed to be the second tetrahedral adduct (**6**) shown in **Figure 6**. (c) *M. arginini* ADI complexed with a planar *S*-alkyl thiouronium intermediate (**5**). In all cases, the protein is shown in tan, and the active-site ligand in blue, with heteroatoms colored for clarity. Protein residues are shown as sticks, ligands as ball-and-sticks, and ordered water molecules as small red spheres. Panels a, b, and c are adapted from the PDB accession codes 2A9G, 1S9R, and 1LXY, respectively.
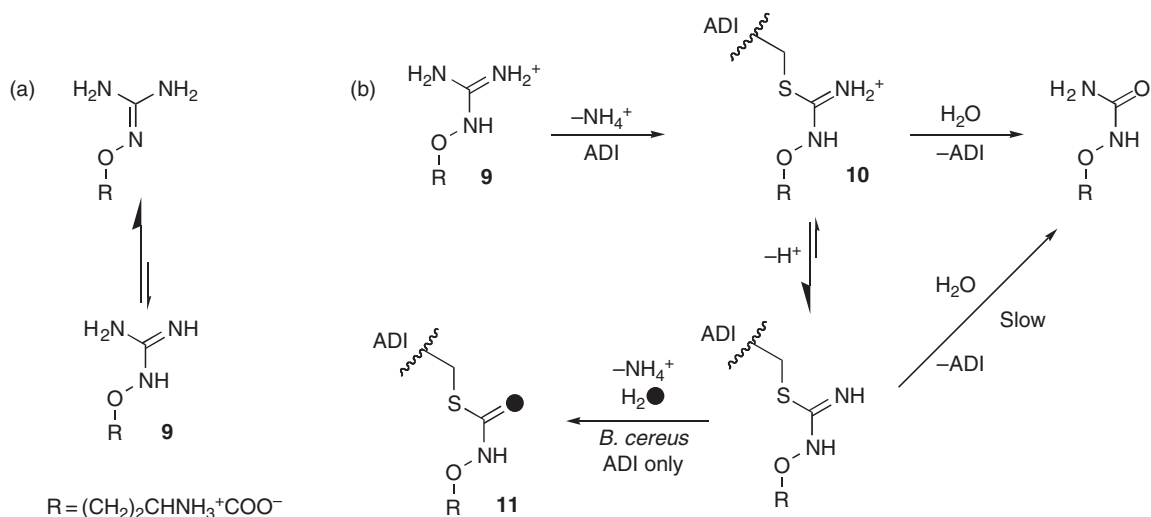
Figure 8   L-Canavanine (**9**) is a slow substrate and *de facto* inhibitor of ADI and an irreversible inactivator of the ADI from *Bacillus cereus*. (a) Canavanine, in its neutral form, favors the amino tautomer. (b) Proposed mechanism of slow turnover of canavanine by ADI and irreversible inactivation of *B. cereus* ADI. Filled circles represent ^{18}O.

Small molecule inhibitors of ADI have been investigated because this enzyme is an attractive drug target. One of the most interesting inhibitors from a mechanistic standpoint is L-canavanine (**9**), a natural product of the leguminous plants.[35] Canavanine is the δ-oxa analog of L-arginine, and its oxyguanidinium group has a significantly lower pK_a value (7.0) than L-arginine (12.5).[36] The neutral form of canavanine favors the amino (rather than the imino) tautomer (**Figure 8(a)**).[36] When reacted with ADI, canavanine acts essentially as a slow substrate. The formation of the oxythiouronium intermediate (**10**) is approximately 10-fold faster than its decay, leading to a time-controlled inactivation of the enzyme that is slowly reversible (**Figure 8(b)**).[34,37] One possible explanation of this phenomenon is that the electronic effects of the δ-oxa substitution may significantly decrease the electrophilicity of the intermediate and selectively affect the second half reaction more than the first.[34] Based on model compounds, the pK_a of the oxythiouronium intermediate formed with canavanine is expected to drop as low as 3.0, resulting in a deprotonated, less electrophilic adduct.[34] Canavanine is a slow substrate of ADI enzymes from *P. aeruginosa*, *Escherichia coli*, *Burkholderia mallei*, and *G. intestinalis*, but can result in irreversible inhibition of the ADI from *Bacillus cereus*.[34] Oxygen isotope labeling studies have suggested that partitioning into this irreversible pathway is due to hydrolysis of the oxythiouronium intermediate to form a thiocarbamate adduct (**11**) rather than releasing the usual ureido product.[34]

ADI also has interesting applications as a therapeutic anticancer enzyme. ADI treatment of various cell lines inhibits proliferation by inducing apoptosis after arresting the cell cycle.[38] The use of ADI to deplete L-arginine is especially useful in the case of human melanomas, renal cell carcinomas, and hepatocellular carcinomas, many of which are known to be auxotrophic for arginine due to an absence of argininosuccinate synthetase; treatment with ADI shows potent inhibition of these cancers.[39,40] Phase I and II trials of pegylated ADI for metastatic melanomas and hepatocellular carcinomas were successful and warranted further use.[41,42] In these trials, pegylated ADI was well tolerated and decreased plasma arginine levels to undetectable levels during treatments as long as 3 months in duration. Because arginine is the bioprecursor of nitric oxide (NO), an important modulator of angiogenesis, the antitumor effects of ADI presumably owe their efficacy to both the antiproliferative activity and the ability to block NO-mediated angiogenesis.[43]

8.05.3.2   Dimethylarginine Dimethylaminohydrolase

8.05.3.2.1   *Biological function*

DDAH catalyzes the hydrolysis of N^{ω}-methyl-L-arginine (**12**) and asymmetric N^{ω},$N0^{\omega}$-dimethyl-L-arginine (**13**) to yield L-citrulline (**1**) and the corresponding alkylamine (**Figure 9**).[44] In vertebrates, two tissue-specific
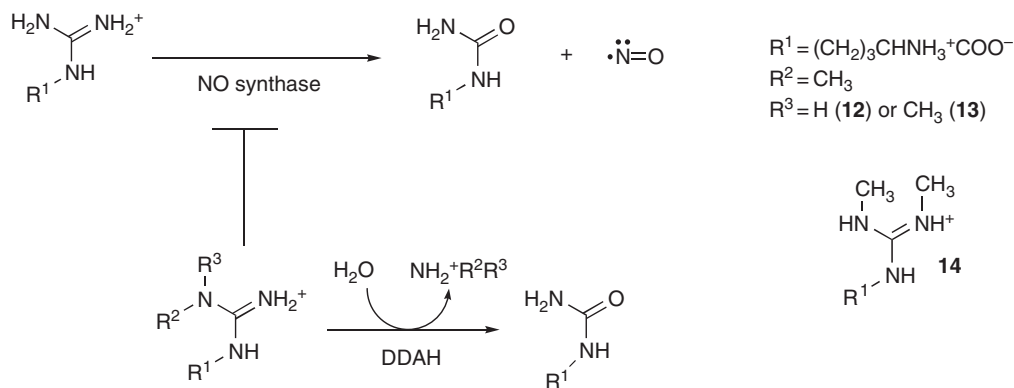
Figure 9 Endogenously produced, asymmetrically N^ω-methylated arginines (**12**, **13**) inhibit NO synthase, but **14** does not. Hydrolysis of asymmetrically methylated arginine residues by dimethylarginine dimethylaminohydrolase (DDAH) can relieve this inhibition and promote NO biosynthesis.

isoforms of DDAH (DDAH-1 and DDAH-2) serve as control valves for NO production.[45–47] Asymmetrically methylated arginine residues are derived from hydrolysis of endogenously methylated proteins and are found in healthy human plasma at low micromolar concentrations (for a review, see Knipp[48]). These methylated arginines serve as endogenous inhibitors of NO synthase and regulate NO production.[49–51] Specifically, N^ω-methyl-L-arginine is a mechanism-based inactivator and N^ω,N^ω-dimethyl-L-arginine is a reversible inhibitor of NO synthase.[52,53] In contrast, the regioisomer, symmetrical $N^\omega,N^{\omega\prime}$-dimethyl-L-arginine (**14**) is neither an inhibitor of NO synthase, nor a substrate of DDAH. In certain disease states marked by endothelial dysfunction, such as renal failure,[54] hypercholesterolemia,[55] diabetes mellitus,[56] and hyperhomocysteinemia,[57] plasma concentrations of N^ω,N^ω-dimethyl-L-arginine are increased and are proposed to cause pathological inhibition of NO synthase. The resulting decrease in NO production leads to the subsequent inability of blood vessels to expand appropriately.[58] There is also evidence that suggests that DDAH promotes angiogenesis during tumor growth, as DDAH overexpression has been linked to neovascularization of C6 gliomas *in vivo*.[59] Although there is some debate regarding the relevant inhibitory concentration of methylarginines *in vivo*, the importance of DDAH in regulating NO-mediated processes has been clearly demonstrated by using small molecule inhibitors of DDAH,[60] interference RNA studies,[47] and transgenic mice carrying either a heterozygous deletion of DDAH-1 (the homozygous deletion is lethal)[51] or an overexpressed human DDAH-1 knock-in.[61]

Oddly, DDAH is also found in various bacteria, including *P. aeruginosa*, *Streptomyces coelicolor*, and *Mycobacterium tuberculosis*,[62] which are not known to produce N^ω-methylated arginine residues. The biological functions of these microbial DDAH enzymes are not obvious. Bacterial DDAH genes are sometimes found in clusters that also contain a gene encoding an amino acid transporter.[62] These bacterial DDAH homologs could allow bacteria to use exogenous methylarginines as a carbon or nitrogen source or decrease their toxicity. There is also the intriguing possibility that bacteria use these enzymes to manipulate their host's NO production during an infection. Promoting NO production as a result of increased DDAH activity might seem counterintuitive because high concentrations of NO are cytotoxic. However, it is known that *P. aeruginosa*, a bacterium often found in chronic infections of inflamed lung tissue, expresses protective NO reductases, can use nitrate (an NO metabolite) as its terminal electron acceptor, and can also uptake and metabolize methylammonium.[63–65] Accordingly, increased concentrations of N^ω,N^ω-dimethyl-L-arginine have been found to block some of the tissue damage associated with *P. aeruginosa* infections.[66] *Mycobacterium tuberculosis* has also a complicated relationship with host-derived NO and related metabolites, and microbial DDAH could potentially play a supporting role in these infections.[67] Further study is required to clarify the biological functions of microbial DDAH enzymes.

8.05.3.2.2   Protein structure

Crystal structures of *P. aeruginosa* DDAH, bovine DDAH-1, and human DDAH-1 have been reported with various ligands bound at their active sites.[51,68–70] To date, only limited functional data and no structural studies have been reported for the DDAH-2 isoform. Bovine DDAH-1 and rat DDAH-1 are monomeric, but
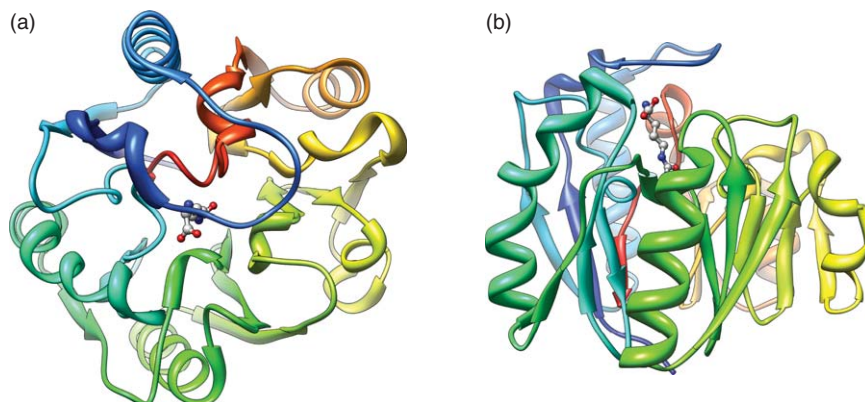
Figure 10 Structure of DDAH. The pentein fold is visible in the 'top' (a) and 'side' (b) views, but is elaborated by a loop insertion in the first subdomain. A ball-and-stick model of bound product indicates that the substrate-binding site is located at the center of the propeller, near the face opposite the enzyme's N- and C-termini. This ribbon diagram is constructed using coordinates from a product-bound C249S mutant of *Pseudomonas aeruginosa* DDAH (PDB accession code 1H70).
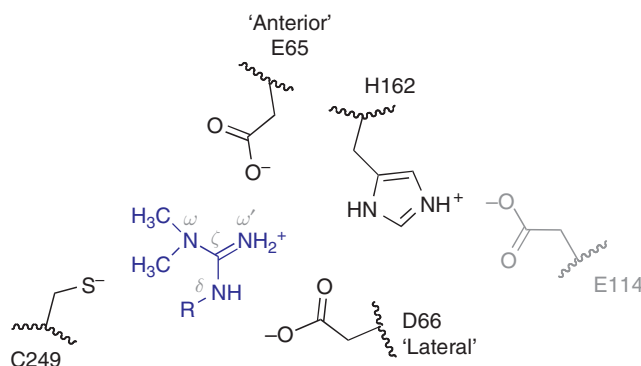
Figure 11 Active-site residues of DDAH that interact with the guanidine of the substrate, N^ω,N^ω-dimethyl-L-arginine (shown in blue). 'Core' residues are shown in black and a neighboring Glu in gray. D66 is designated the 'lateral' Asp, and E65 the 'anterior' Glu, to denote their coordination of the substrate's guanidine. The active-site Cys and His are located on the *re*- and *si*-faces, respectively, of the substrate's guanidinium. Amino acid numbering is taken from *Pseudomonas aeruginosa* DDAH.

P. aeruginosa DDAH is dimeric, with two independent active sites.[44,71,72] A modified pentein fold is obvious in the structure of DDAH (**Figure 10**). The product-bound structure of a mutant *P. aeruginosa* DDAH indicates that ligands bind at the center of the propeller near the face opposite the N- and C-termini in a similar orientation to that seen in ADI.[69] A loop insertion in the first subdomain folds over the bound ligand and is thought to facilitate binding. The core active-site residues observed in ADI are conserved in DDAH, with one notable exception (**Figure 11**): In primary sequence alignments, the anterior Asp of ADI does not align with a corresponding residue in DDAH. Instead, a Glu residue (E65) from a different part of the protein reaches toward the active site and makes a monodentate interaction with the unsubstituted N^ω of the substrate. It is proposed that the bidentate interaction of the anterior Asp of ADI is what excludes binding of the methylated arginine substrates, whereas the extra space created by the monodentate chelation in DDAH may provide extra space for binding methyl substituents.[73] The remaining core residues appear to be quite similar, with the active-site Cys and His residues poised on the *re*- and *si*-face, respectively, of the N^ω,N^ω-dimethyl-L-arginine's guanidinium.

8.05.3.2.3 Catalytic mechanism

The proposed catalytic mechanism of DDAH is very similar to that of ADI (**Figure 12**).[69,74] The active-site Cys attacks the guanidine carbon (C^ζ), leading to a tetrahedral adduct (**15**). Donation of a proton from the active-site cationic His residue to the alkyl-substituted N^ω atom promotes C^ζ–N^ω bond cleavage and loss of an alkylamine to
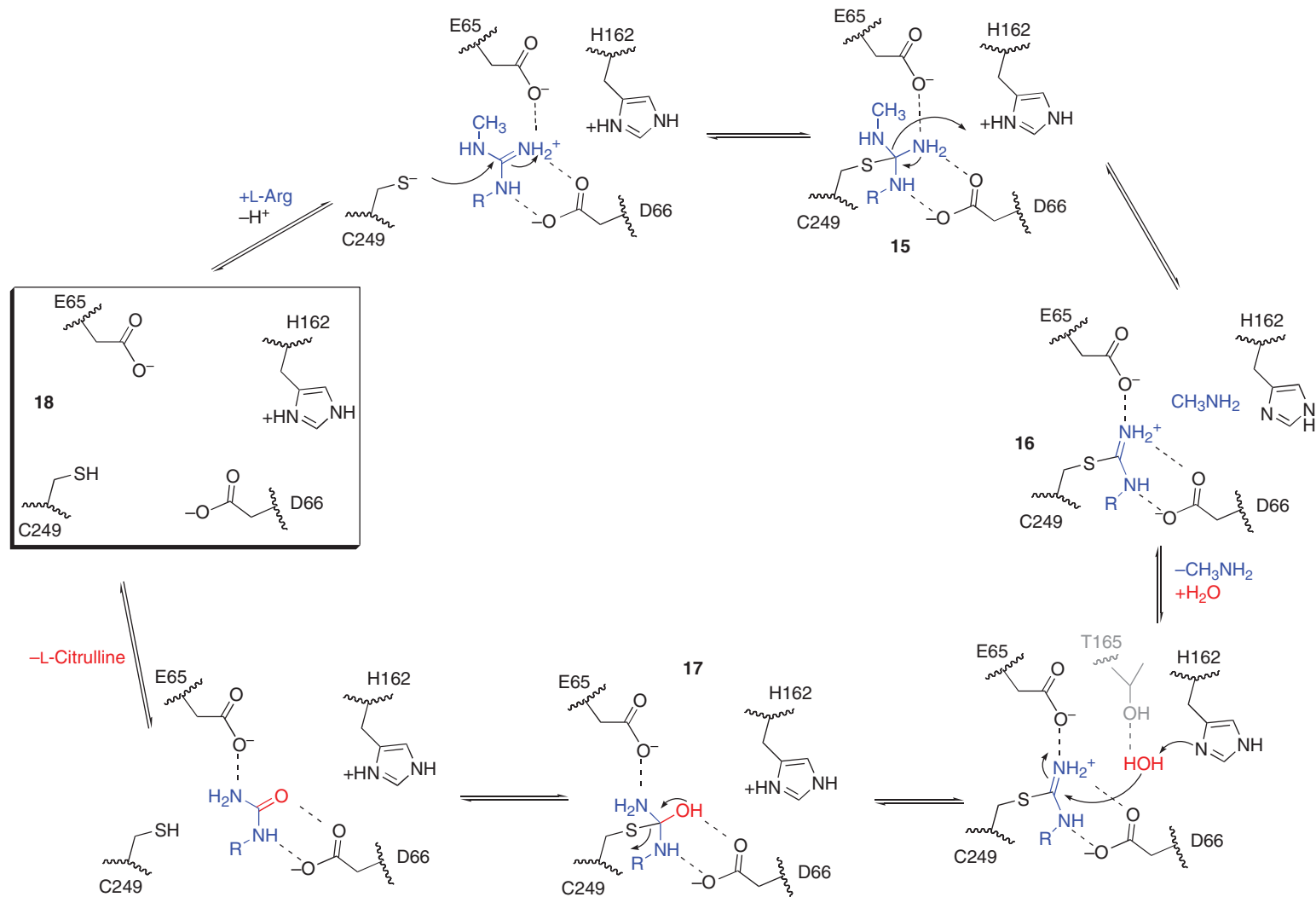
Figure 12 Proposed catalytic mechanism of DDAH. An N^ω-methyl-L-arginine substrate is shown in blue, the 'core' residues in black, the hydrolytic water in red, and a putative interaction with Thr165 in gray. Amino acid numbering is from *Pseudomonas aeruginosa* DDAH. The apo form of the enzyme is boxed.

form the planar S-alkyl thiouronium covalent intermediate (**16**). Interactions of the remaining N^δ and N^ω nitrogens of the intermediate with the lateral Asp are maintained. The anterior Glu maintains its monodentate interaction with the remaining N^ω nitrogen. Because the anterior Glu does not bind to the leaving group, no stabilization of the hydrolytic water by this residue is expected. Instead, the hydrolytic water may be stabilized by the active-site His and a conserved Ser/Thr residue (T165 in *P. aeruginosa* DDAH). This contrasts with the proposed ADI mechanism (**Figure 6**), in which the anterior Asp can help position the hydrolytic water. The core His of DDAH, now neutral, can act as a general base to deprotonate this water for attack on the C^ζ carbon of the planar thiouronium intermediate, forming the second tetrahedral adduct (**17**). Finally, collapse of this adduct and elimination of the active-site Cys regenerate the resting enzyme (**18**). Covalent catalysis has been demonstrated using mass spectrometry; the thiouronium covalent intermediate was trapped by an acid quench during steady-state turnover of both natural and synthetic substrates.[74,75] The importance of each of the core residues to catalysis is consistent with mutagenesis studies.[69,70,74,75] Crystal structures representing the substrate complex, product complex, and covalent thiouronium intermediate have all been reported and support the proposed mechanism.[69,70]

DDAH appears to share an additional unusual mechanistic feature with ADI. The active-site Cys is predominantly protonated, and hence much less reactive in the resting state of the enzyme.[74] The bound guanidinium of the substrate appears to stabilize the anionic thiolate form of the Cys, increasing its nucleophilicity and thereby participating in the mechanism. The high pK_a of the active-site Cys in the resting enzyme has important implications for understanding both the normal and pathological redox regulation of this enzyme (see below).

8.05.3.2.4 *Regulation, inhibitors, therapeutic application*

The redox sensitivity of DDAH has received considerable attention because of the enzyme's physiological role in regulating NO production. The active-site Cys residue can undergo S-nitrosylation, albeit under high concentration of reactive nitrogen species.[76,77] The resulting pathological S-nitrosylation would inhibit DDAH activity, raise levels of the endogenous NO synthase inhibitors, and decrease NO production, resulting in a feedback regulation loop. However, although it is often claimed that DDAH is unusually sensitive to oxidative stress, quantitative studies have shown that the core Cys residue is relatively insensitive to oxidation.[78,79] Human DDAH-1 is surprisingly insensitive to inactivation by H_2O_2, a powerful biological oxidant, and is also relatively insensitive to many reactive nitrogen species.[78,79] The resistance of DDAH to inactivation by most reactive oxygen and nitrogen species is not surprising when the catalytic mechanism is considered. Because the active-site Cys nucleophile is predominantly protonated in the resting state of the enzyme, this thiol is much less reactive and much less susceptible to oxidation than an anionic thiolate.[78] This scenario contrasts with that of other enzymes that use an active-site Cys nucleophile, such as caspase-3 and the phosphatases Cdc25B and PTP1B, in which the active-site Cys has a low pK_a in the resting enzyme. These enzymes are readily inactivated by reactive oxygen and nitrogen species, with inactivation rate constants several orders of magnitude larger than the human DDAH-1.[78–86] Tuning the Cys pK_a value in resting DDAH also makes sense physiologically to ensure that the enzyme is not inactivated by normal fluctuations in the reactive oxygen or nitrogen species, while allowing feedback inhibition when these redox species reach pathological concentrations. This is a good illustration of how a small mechanistic detail (the protonation state of one residue) can have important physiological ramifications.

Despite the insensitivity of DDAH to inactivation by most reactive oxygen and nitrogen species, the enzyme is readily inhibited by a select group of biologically relevant compounds. Redox modulation of DDAH activity has been proposed to cause pathological increases in N^ω,N^ω-dimethyl-L-arginine, subsequent NO inhibition and the resulting endothelial dysfunction found in diabetes and cardiovascular pathologies. Cardounel's group has demonstrated that irreversible inactivation of human DDAH-1 by the lipid peroxidation product 4-hydroxy-2-nonenal occurs at physiologically relevant concentrations through specific covalent modification of the active-site His.[79] Notably, the active-site Cys does not readily react with this aldehyde, presumably due to its high resting pK_a value. Additionally, human and bovine DDAH-1 can be inactivated by the reactive nitrogen species *S*-nitroso–L-homocysteine (HcyNO) (**19**).[68,78,86,87] In contrast to the results with H_2O_2, inactivation by HcyNO occurs with rate constants similar to physiologically relevant inactivation of other enzymes by S-nitrosylation and may be biologically relevant.[78,86] Vasak and coworkers have determined that inactivation of DDAH by HcyNO does not occur as expected, by transnitrosation or formation of a disulfide with the active-site Cys. Rather, inactivation results in the formation of a *N*-thiosulfoximide adduct (**20**) with the active-site Cys through an unusual mechanism (**Figure 13**).[68,87] Additional conformation of this unusual *N*-thiosulfoximide adduct was obtained through crystal
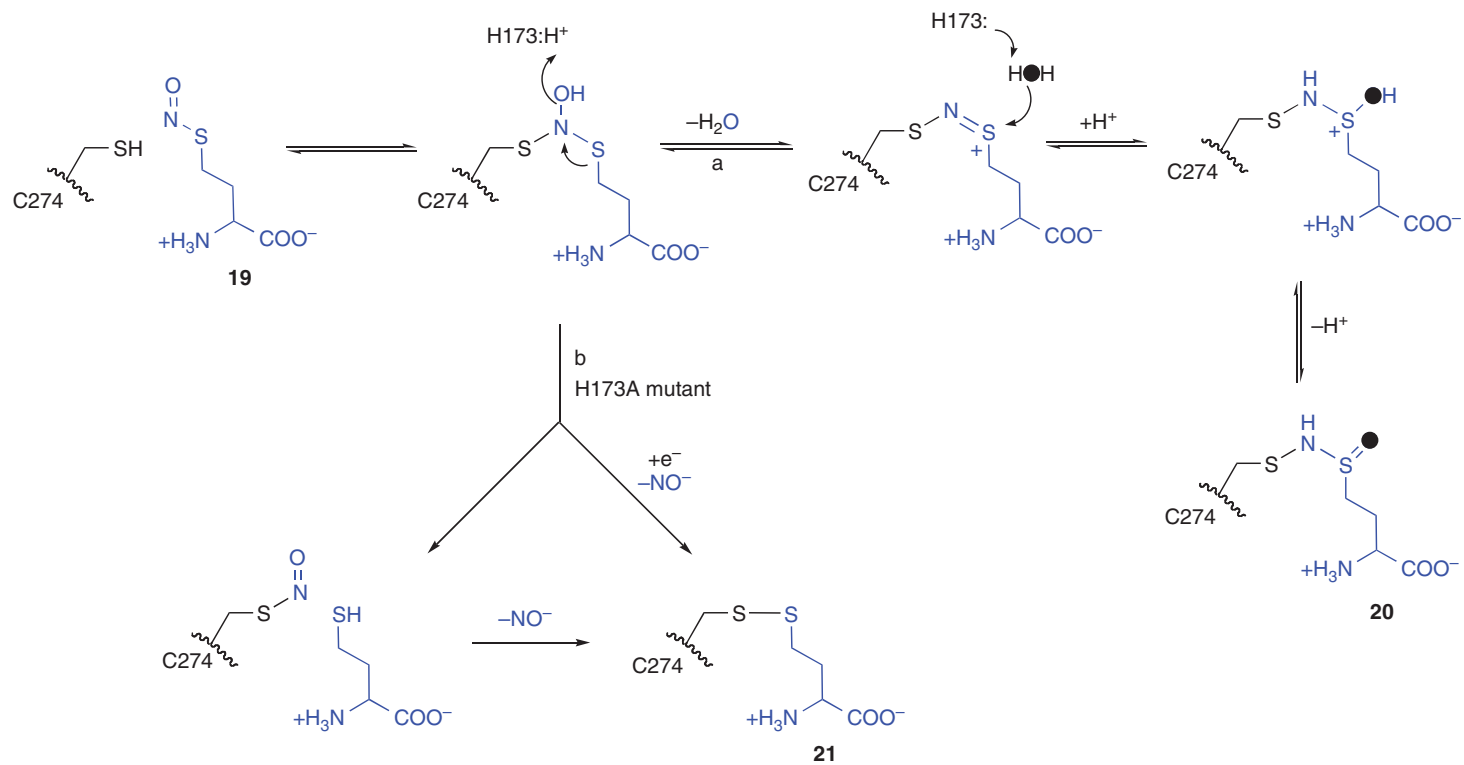
Figure 13 Proposed mechanism-based inactivation of DDAH-1 by *S*-nitroso-ʟ-homocysteine (**19**) (HcyNO, in blue). Reaction pathways for HcyNO inactivation of a wild-type mutant (path a) and a core H173A mutant (path b) are shown. Filled circles represent ^{18}O labeling. Numbering corresponds to that of human DDAH-1.
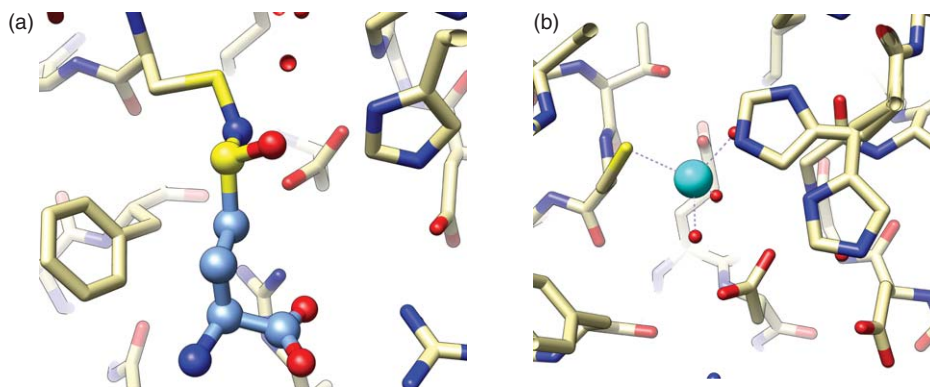
Figure 14 Structures of DDAH-1 with endogenous inhibitors. (a) Bovine DDAH-1 (tan, stick model) bearing an *N*-thiosulfoximide adduct (**20**) (blue, ball-and-stick model) at the active-site Cys. Only one of the two alternative adducts is shown here. (b) Zinc(II) bound (light blue sphere) at the active site of bovine DDAH-1 (tan) between the core Cys and His residues. An alternative conformer of His is also shown. Ordered water molecules are shown as small red spheres. Panels (a) and (b) were constructed using coordinates from PDB accession codes 2CI1 and 2CI7, respectively.

structures of bovine DDAH-1 inactivated by HcyNO (**Figure 14(a)**).[68] Further work with human DDAH-1 has shown that some of the active-site residues involved in the normal catalytic mechanism also participate in the formation of this unusual adduct and may help to explain this unexpected partitioning (Y. Wang, L. Hong, and W. Fast, unpublished observations). In particular, mutation of the core His residue in human DDAH-1 blocks *N*-thiosulfoximide (**20**) formation and instead leads to the formation of one of the more typical products, the mixed disulfide (**21**). Therefore, because aspects of the normal catalytic mechanism are required for inactivation by HcyNO, this compound can be classified as a mechanism-based inactivator of DDAH-1.

In addition to these covalent modifications, bovine and *P. aeruginosa* DDAH can also be inhibited by Zn(II) binding at the active site (**Figure 14(b)**).[68,88] Therefore, release of zinc during periods of oxidative stress may also be a potential cause of DDAH inhibition.[48] Further studies will be required to determine which of these inhibition pathways, or others, occur *in vivo* during normal and pathophysiology.

Apart from naturally occurring inhibitors, DDAH inhibition by synthetic small molecules is a therapeutic goal for cancer due to this enzyme's involvement in tumor angiogenesis and other disease states in which NO levels are elevated. A series of substrate analogs has been characterized as micromolar inhibitors of DDAH.[60] Of these, N^{ω}-(2-methoxyethyl)-L-arginine (**22**) (**Figure 15**) was characterized by X-ray crystallography in complex with human DDAH-1.[68] Although this inhibitor is chemically similar to normal substrates, a rearrangement of active-site residues upon binding holds the inhibitor's guanidinium in a nonproductive orientation and displaces the core His residue. This rearrangement is seen easily in comparison with the citrulline-bound structure (**Figure 16**). A few alternative scaffolds for developing reversible DDAH inhibitors have also been reported, including pentafluorophenyl sulfonates (**23**) and indolyl barbiturates (**24**) (**Figure 15**).[89,90] The indole fragment is proposed to bind at the hydrophobic surface normally occupied by the substrate's backbone, but the precise way that these inhibitors interact with DDAH remains to be determined. An irreversible inactivator of DDAH based on an α-halo amidine structure (**25**) has also been reported.[91] This affinity label covalently modifies the active-site Cys specifically, even in the presence of four other Cys residues. Inactivation occurs by a mechanism similar to that of α-halo ketone inhibitors of cysteine proteases (**Figure 15**). Although this reactive moiety has limited affinity and selectivity, further structural elaboration will likely improve its usefulness.

DDAH itself may also have potential as a therapeutic protein in disease states marked by excess N^{ω},N^{ω}-dimethyl-L-arginine. For example, overexpression of DDAH has beneficial effects in transgenic mouse models of graft coronary artery disease and can enhance sensitivity to insulin.[61,92] DDAH may also have beneficial effects in treating chronic kidney disease, as overexpressed DDAH appears to slow progression of renal dysfunction in rat models.[93,94] Elevated asymmetric dimethylarginine (ADMA) levels have also been identified in the development of chronic lung diseases, specifically pulmonary fibrosis. However, the causal relationship between ADMA and these conditions has yet to be explored in detail.[95]
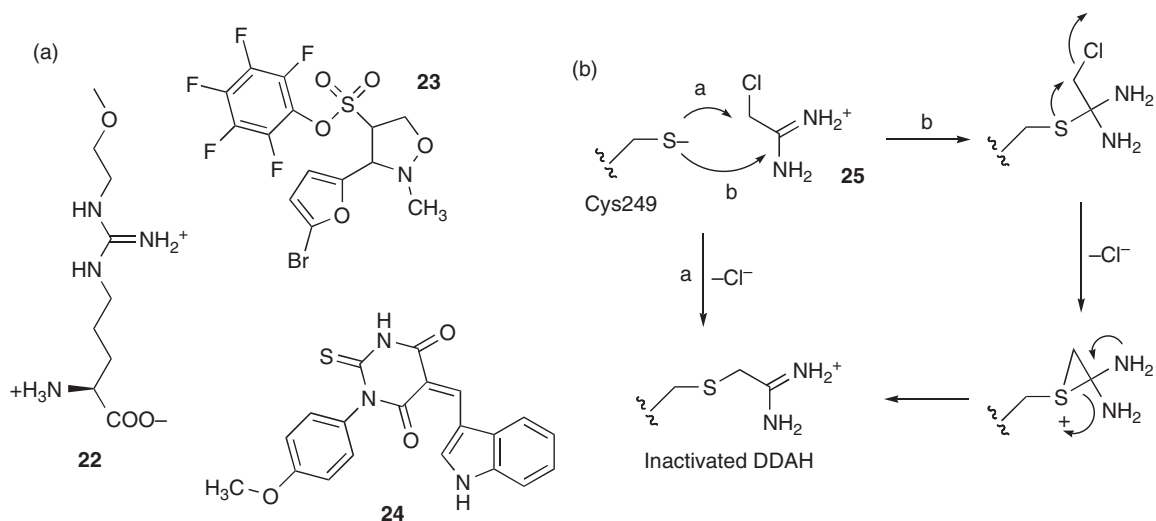
Figure 15 Selected inhibitors of DDAH. (a) Selected reversible inhibitors, including representative structures of N^{ω}-substituted-L-arginines (**22**), pentafluorophenyl sulfonates (**23**), and indolyl barbiturates (**24**). (b) Irreversible inactivation of *Pseudomonas aeruginosa* DDAH by 2-chloroacetamidine (**25**).
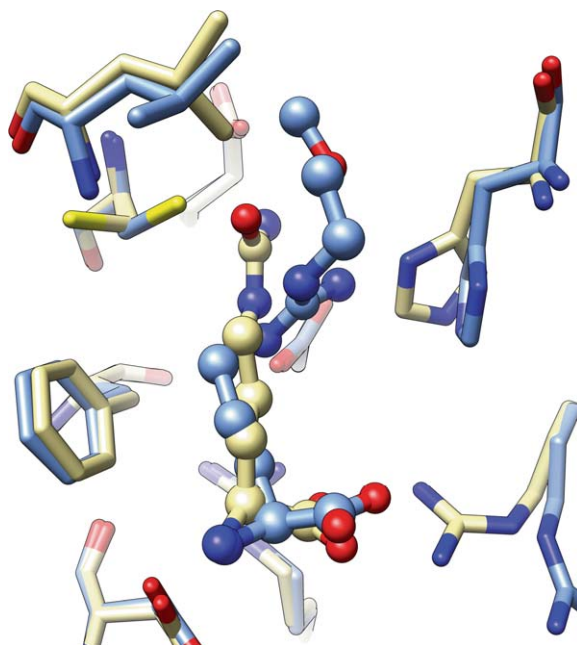
Figure 16 Overlay of human DDAH-1 structures bound to the product L-citrulline (tan) and N^{ω}-(2-methoxyethyl)-L-arginine (**22**) (blue). Protein residues are shown as stick models and ligands as ball-and-stick models. Rearrangements of active-site Leu, His, and Arg residues accommodate the inhibitor, and hold the guanidinium in a nonproductive orientation. The figure was constructed using coordinates from PDB accession codes 2JAJ and 2JAI, respectively.

8.05.3.3   Peptidylarginine Deiminase

8.05.3.3.1   *Biological function*

Enzymes in the pentein superfamily are also capable of catalyzing posttranslational modifications. (PAD also called protein arginine deiminase (PADI)) catalyzes hydrolysis of arginine side chains found in proteins and
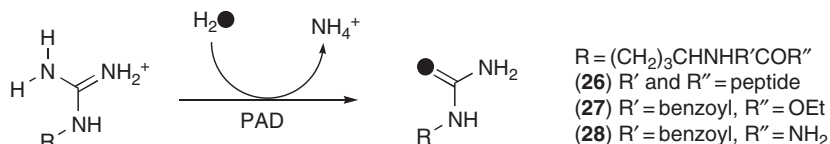
Figure 17 Peptidylarginine deiminase catalyzes the posttranslational modification of peptidylarginine to peptidylcitrulline (**26**). The filled circles represent ^{18}O labeling.

peptides to yield peptidyl citrulline (Cit) (**26**) (**Figure 17**). The only isoform of PAD isolated from prokaryotes is from *Porphyromonas gingivalis*, the organism associated with periodontitis.[96] However, five isoforms of PAD are known in humans: PAD1, PAD2, PAD3, PAD4 (also called PAD5), and PAD6. These isoforms have different tissue localizations and protein substrates (for reviews, see Thompson and Fast,[73] Gyorgy *et al.*,[97] and Vossenaar *et al.*[98]). Comprehensive lists of these proteins' substrates are yet to be determined, but some notable known substrates are listed here (**Table 1**). Small molecules with N^α-protecting groups, such as N^α-benzoyl-L-arginine, ethyl ester (**27**), are also accepted as substrates *in vitro*, and at least one isoform of PAD can use arginine derivatives bearing a free carboxylate.[99–101]

Modification of a protein's Arg residues to Cit has very interesting biophysical properties and can lead to protein unfolding, alterations of protein–protein interactions, and even changes in the oligomeric state.[102–104] This makes PAD enzymes well suited for their roles in cellular protein signaling networks, although much is unknown about their physiological roles. PAD enzymes are implicated in many inflammatory and autoimmune disorders. The presence of autoantibodies to citrullinated proteins, such as fibrin, is a key characteristic of rheumatoid arthritis. PAD2 and PAD4 are overexpressed in this disease state, suggesting that PAD enzymes may catalyze a key step in its progression.[105] PAD2 citrullinates CXC chemokine ligands 10 and 11, which may affect immune response.[106] PAD2 is also known to deiminate myelin basic protein (MBP), which leads to myelin instability, as seen in multiple sclerosis.[107] Deiminated keratin K1, another PAD substrate, is found in healthy epidermal tissue, but not in psoriasis-involved tissue.[108]

PAD4, the only known human PAD that localizes to the nucleus,[109] may also regulate gene expression. PAD4 can repress transcription of certain genes, such as p53 target genes p21 and OKL38 and the estrogen receptor, by citrullinating histones.[110–112] PAD-mediated suppression of p53 targets may be important in cancer progression by blocking apoptosis. This is supported by evidence that PAD4 is expressed in several tumors, notably adenocarcinoma,[113] and that PAD4 inactivation increases levels of p53 and other apoptotic proteins, presumably by citrullinating histones and preventing their methylation.[111] In contrast, experiments with hematopoietic cells show that overexpression of PAD4 also leads to p53- and p21-mediated apoptosis.[114] The nature of PAD4's relationship to apoptotic signaling remains to be investigated.

Much of the cellular PAD enzyme is normally inactive because of its regulation by calcium, complicating its study *in vivo*. However, Thompson's group has developed a fluorescently labeled inactivator that reacts only with the active form of the enzyme, allowing activity-based protein profiling of the PAD enzyme.[115] This approach holds promise for clarifying PAD's intricate involvement in biological processes and assisting in the diagnosis of rheumatoid arthritis.

Table 1 PAD isoform localizations and selected substrates

PAD isoform	Location	Protein substrates
PAD1	Epidermis, uterus	Keratins, filaggrin
PAD2	Brain, skeletal muscle, spleen, secretory glands	Myelin basic protein, vimentin
PAD3	Hair follicles	Trichohyalin
PAD4/PAD5	White blood cells	Histones, nucleophosmin/B23
PAD6/ePAD	Egg cells	–

8.05.3.3.2 Protein structure

PAD4 is the only human isoform that has been structurally well characterized. Crystal structures of PAD4 have been reported in the Ca^{2+}-bound and Ca^{2+}-free forms,[116] as well as complexed with N^{α}-benzoyl-L-arginine amide (**28**),[116] N^{α}-benzoyl-N^{5}-(2-fluoro-1-iminoethyl)-L-ornithine amide,[117] and N-terminal peptides of histones H3 and H4.[118] PAD4 crystallizes as a dimer mediated by interactions between the N-terminal domain of one monomer and the C-terminal domain of the other. Although the residues involved in PAD4 dimerization are generally conserved among the PAD isoforms, it is not clear whether dimerization is a common feature among PAD enzymes.[116] The shape of each monomeric unit has been described as a 'rubber boot', with the N-terminal domain as the heel and toe, and the C-terminal domain as the shaft (**Figure 18**).[116] The N-terminal domain contains two immunoglobulin-like subdomains that resemble those found in the Ca^{2+}-dependent proteins, protein kinase C and phospholipase C. Immunoglobulin-like domains are proposed to facilitate protein–protein interactions, suggesting a possible role in regulation or protein–protein interaction for the N-terminal domain of PAD4. The C-terminal domain, on the other hand, is catalytic and consists of a modified pentein fold.

The core active-site residues are largely conserved in the PAD subfamily, with one notable exception (**Figure 19**). Interestingly, PAD enzymes lack a carboxylate residue that is typically found within hydrogen
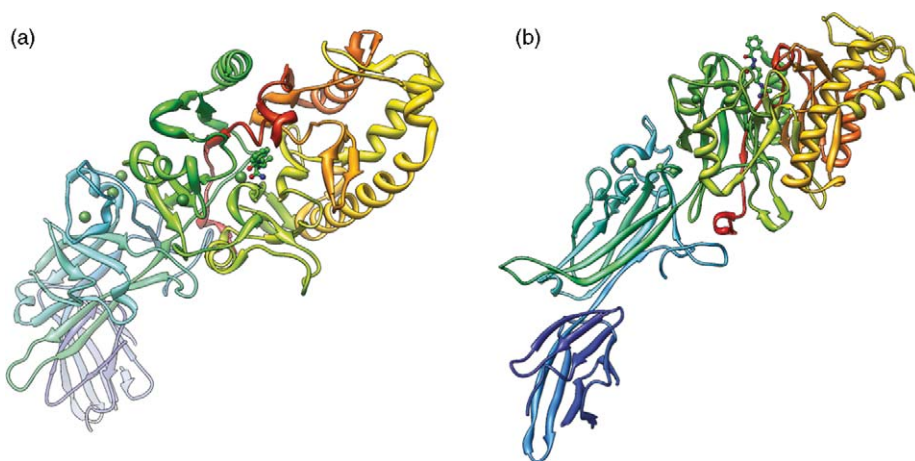
Figure 18 Structure of PAD. An elaborated pentein fold is observable in the 'top' (a) and 'side' (b) views of the C-terminal domain, with two immunoglobulin-like domains appended to the N-terminus. A ball-and-stick model of an active-site ligand (**28**) indicates that the substrate-binding site is located at the center of the propeller, near the face opposite the enzyme's N-domain and C-termini. Calcium ions are represented as spheres. This ribbon diagram is constructed using coordinates from a substrate-bound C645A mutant of human PAD4 (PDB accession code 1WDA).
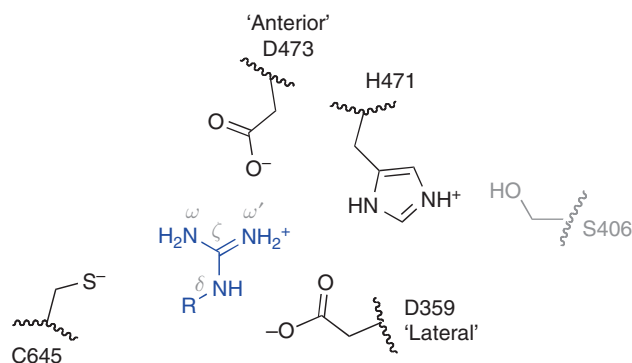
Figure 19 Active-site residues of PAD4 that interact with the guanidine of the peptidylarginine substrate (shown in blue). 'Core' residues are shown in black, and a neighboring Ser in gray. D359 is designated the 'lateral' Asp, and D473 the 'anterior' Asp, to denote their coordination of the substrate's guanidine. Amino acid numbering is taken from human PAD4.
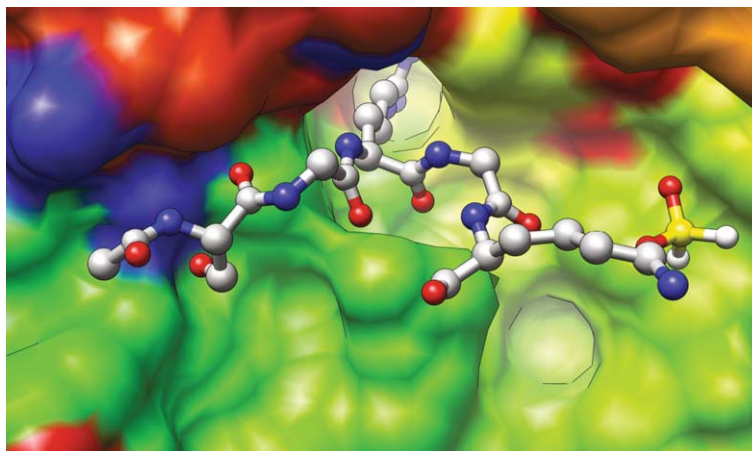
Figure 20   Structure of a histone H4-derived peptide complexed with PAD4. Only five of the peptide's 10 amino acids are ordered and adopt a β-turn-like conformation. Apart from the binding pocket for the substrate's arginine side chain, most interactions with the protein occur through hydrogen bonds to the peptide's backbone. The peptide and a molecule of dimethylsulfoxide (left) are displayed as ball-and-stick models, and the protein is surface coated.

bonding distance of the core His residue.[116] The functional purpose of this alternative sequence has not yet been determined, but it has been proposed that the loss of the acidic residue may reduce catalytic activity and serve to protect the cell from damagingly high levels of protein citrullination.[6]

The structural basis for recognition of protein substrates is not yet clear. However, structures of PAD4 in complex with N^{α}-benzoyl-L-arginine amide and peptides from histones H3 and H4 that are known to be citrullinated have provided clues as to how PAD4 binds and recognizes small molecule substrates. As with ADI and DDAH, the arginine residue of the substrate binds at the center of the propeller of the catalytic domain, and the conserved core residues interact with the guanidine side chain.[116]

The structure of PAD4 has also elucidated key differences in substrate recognition between PAD4 and other members of its superfamily. In ADI and DDAH, the active site is buried within the enzyme and covered by a 'lid' region that could prevent large peptides from entering. However, the PAD4 active site is located on the surface of the enzyme, which may facilitate the binding of sterically large peptides (**Figure 20**). Most of the contacts PAD4 makes with peptide substrates occur through hydrogen bonds to the substrate peptide's backbone. Because substantial interactions are not made with the substrate peptide's side chains, a consensus recognition sequence has been hard to define. Based on computational modeling and crystal structures with bound peptides, a very broad consensus sequence has been proposed, Φ-X-Arg-X-X, with the only real constraints being the presence of an arginine and an amino acid with a small side chain (Φ) two positions earlier.

This lack of sequence specificity has raised questions about how PAD4 recognizes substrates. As observed in crystal structures, five residues of the peptide substrate nearest to the PAD4 active site adopt an ordered β-turn-like bent conformation upon binding. However, all other residues in the peptide fragments were flexible and not visible in electron density maps.[118] This suggests that PAD4 recognizes disordered peptides and, upon binding, induces a β-turn-like bent conformation. This proposal is supported by evidence that PAD activity directly correlates with the structural order of the substrate.[102] The difficulty in predicting disordered regions in proteins has hindered identification of PAD substrates, and the problem of finding new substrates for PAD is consequently an active area of interest.

8.05.3.3.3   *Catalytic mechanism*

The catalytic mechanism of PAD4 has been extensively characterized[119–121] and is quite similar to that proposed for ADI. Similar to ADI and DDAH, PAD4 also shares the unusual feature that the core Cys is predominantly protonated in the resting state of the enzyme. Only approximately 15% of the resting enzyme is found in the proper protonation state to support catalysis.[120] One analysis of inactivation kinetics indicates that at least some of the substrate reacts with the enzyme without having to induce protonation upon binding.[120]
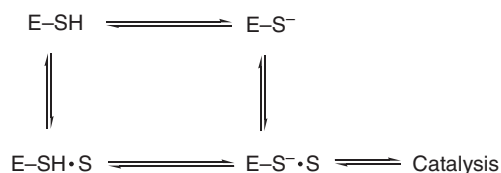
Figure 21 Thermodynamic box describing two possible routes for enzymes in the pentein superfamily to bind to the substrate. Enzyme (E), the core Cys residue (S⁻ or SH), and the cationic substrate (S) are abbreviated as noted.

In general, the interplay of binding a cationic substrate and the protonation state of the core Cys residue in the pentein superfamily enzymes can be described by a thermodynamic box (**Figure 21**, adapted from Knuckley *et al.*[119]); but the kinetically preferred reaction pathway for thiol deprotonation in any of the pentein enzymes has not been firmly established.

8.05.3.3.4 Regulation, inhibitors

PAD4 binds five calcium ions, two near the catalytic C-terminal domain and three in the N-terminal domain.[116] Calcium binding induces order into an unstructured region of the protein near the active site, suggesting that calcium binding is required to achieve a conformation that permits the substrate to bind.[116] Accordingly, the activity of PAD enzymes is strongly Ca^{2+} dependent and usually require millimolar concentrations of Ca^{2+} for activation.[120,122] Such Ca^{2+} concentrations are higher than typical micromolar cellular Ca^{2+} concentrations, so at the present time it is unclear how PAD is regulated by calcium signaling.

Arginine methylation and citrullination may be antagonistic posttranslational modifications that compete with one another. Methylation of arginine residues found in PAD substrates can block their citrullination. *In vitro* studies found that methylated arginines are only poor substrates of PAD and are unlikely to be significantly modified.[120,123] However, *in vivo* experiments indicate that PAD4 can hydrolyze monomethylated (but not dimethylated) arginine.[112] The interplay of these two posttranslational modifications is an active area of study.

The most potent inhibitors to date for PAD are affinity labels that work in a similar fashion as described for 2-chloroacetamidine (**25**) (**Figure 15(b)**). These inactivators incorporate an α-haloamidine moiety into a substrate analog and can achieve very potent and selective modification of the core Cys in PAD enzymes.[124] This strategy shows considerable therapeutic promise because only the active forms of the enzyme are targeted, possibly limiting toxicity.[117] Fluorescently tagged inactivators have been used to screen disease-modifying antirheumatic drugs used in multiple sclerosis therapies and discovered that an unexpected subset of these compounds – streptomycin, minocycline, and chlorotetracycline – are actually micromolar PAD inhibitors with molecular scaffolds quite unlike known peptide substrates.[125]

8.05.3.4 Agmatine Iminohydrolase

8.05.3.4.1 Biological function

Enzymes in the pentein superfamily are also involved in polyamine biosynthesis. Polyamines are small molecules essential for cell growth and differentiation. Their precursors are produced by the arginine decarboxylase (ADC) pathway in plants and some bacteria.[126] Elements of this pathway have also been recently identified in *Chlorella* viruses and may be important for virus replication.[127] In the ADC pathway, arginine is decarboxylated to form agmatine (**29**), which is hydrolyzed to give *N*-carbamoylputrescine (**30**) and then hydrolyzed again to form putrescine (**31**), an essential polyamine building block (**Figure 22**). Agmatine iminohydrolase (AIH), a member of the pentein superfamily, catalyzes the hydrolysis of agmatine to *N*-carbamoylputrescine and ammonia and is essential for polyamine biosynthesis in some organisms that use the ADC pathway. However, the essentiality of AIH activity to polyamine-related biological processes remains unclear, as many plants also have alternate pathways for putrescine biosynthesis.

AIH also plays a role in the AgDI pathway, which allows agmatine to be used as an energy source in *Enterococcus faecalis*. In this pathway, which is similar to the ADI pathway, agmatine is converted into
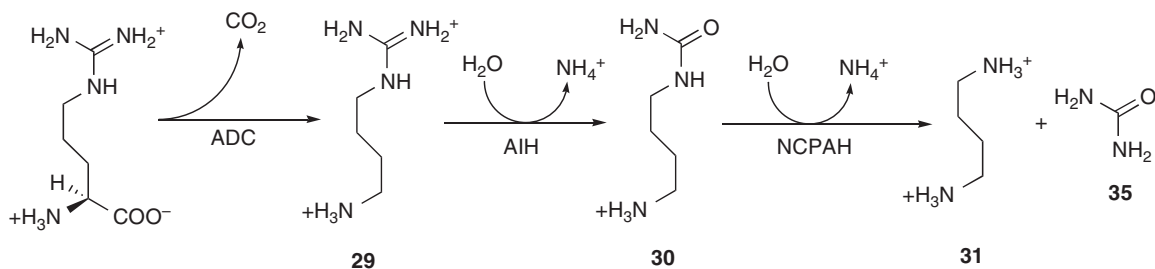
Figure 22 Biosynthesis of polyamines through the arginine decarboxylase pathway. Putrescine derived from L-arginine through the action of three successive enzymes: arginine decarboxylase (ADC), agmatine iminohydrolase (AIH), and *N*-carbamoylputrescine amidohydrolase (NCPAH).

N-carbamoylputrescine by AIH and subsequently phosphorylated to produce carbamoyl phosphate, which can be used to phosphorylate ADP, yielding ATP.[128]

8.05.3.4.2 *Protein structure*

The AIH enzymes from *E. faecalis* and *Arabidopsis thaliana* have been structurally characterized and show clear structural similarities with ADI, although the pentein folds contain a different set of loop insertions (**Figure 23**).[129] Oligomerization appears to vary by the enzyme's source: AIH from *E. faecalis* AIH crystallizes as a tetramer,[129] AIH from *A. thaliana*,[126] *P. aeruginosa*,[130] and maize shoots[131] are dimers, and soybean AIH is a monomer.[132]

The *E. faecalis* AIH active site is located at the center of the propeller and is covered by a three-pronged gate.[129] The core residues which interact with the guanidine of agmatine are conserved and are similar to those found in ADI (**Figure 24**). In *A. thaliana*, but not *E. faecalis* AIH, a second Cys residue is present in the active site, positioned close to agmatine's primary amino group.[133]

8.05.3.4.3 *Catalytic mechanism*

The catalytic mechanism for AIH has not been fully elucidated. In the structure of *E. faecalis* AIH complexed with agmatine, a covalent adduct to the core Cys is modeled to additional electron density found at the active site and identified as a thiouronium intermediate.[129] However, the geometry of the adduct does not appear to be planar, as would be expected for a group with sp^2 hybridization. The mechanism of the *A. thaliana* AIH is even
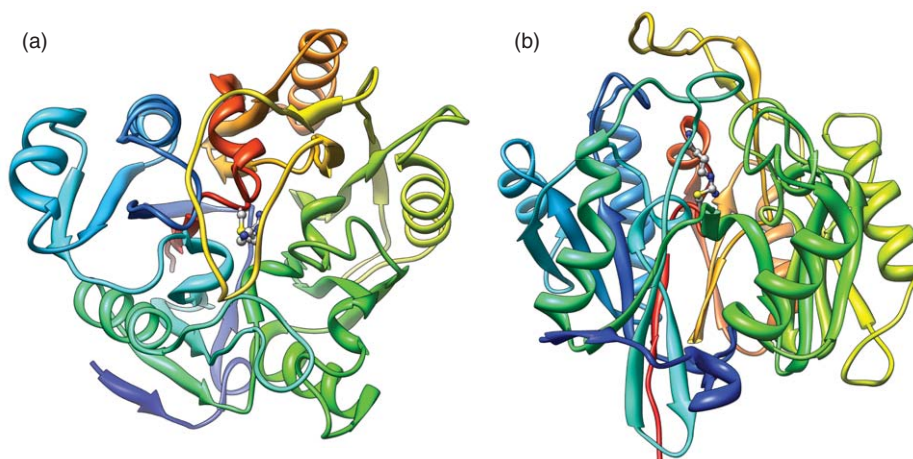
Figure 23 Structure of AIH. The pentein fold is visible in 'top' (a) and 'side' (b) views, but is elaborated by long loop inserts, especially in the fifth subdomain. A ball-and-stick model of a putative trapped intermediate is located at the center of the propeller, near the face opposite the C-terminus. The ribbon diagram was constructed using coordinates for *Enterococcus faecalis* AIH (PDB accession code 2JER).
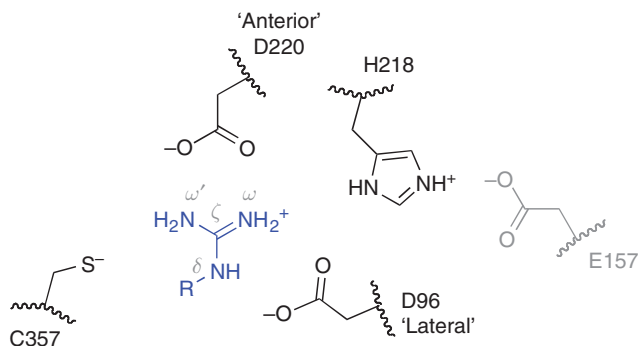
Figure 24 Active-site residues of AIH that interact with the guanidine of the substrate, agmatine (**29**) (shown in blue). 'Core' residues are in shown in black, and a neighboring Glu in gray. D96 is designated the 'lateral' Asp, and D220 the 'anterior' Asp, with respect to the substrate's N^δ–C^ζ bond. Amino acid numbering is taken from *Enterococcus faecalis* AIH.

less clear. Mutation of the core Cys residue reduces activity less than 10-fold from the wild type,[126] raising the possibility of an alternative reaction mechanism.

8.05.3.4.4 *Inhibitors, alternative substrates*

Little is known about *in vivo* regulation of AIH. Arcaine, a bisguanidine substrate analog, is reported to be a competitive inhibitor of AIH in maize[131] and cucumber seedlings,[134] and a substrate for *Chlorella* virus AIH.[127]

8.05.4 Dihydrolases

The dihydrolases catalyze sequential hydrolytic reactions on a guanidine, breaking each of the three C^ζ–N bonds of the substrate and liberating the guanidino carbon as carbon dioxide (or bicarbonate). Although the only characterized dihydrolase is the *N*-succinylarginine dihydrolase (SADH) found in some microorganisms, functional annotations of other suspected dihydrolases have been predicted based on sequence alignments. Proposed dihydrolase enzymes are found in various cyanobacteria as well as some pathogenic microorganisms.[6] The biological roles and catalytic functions of these hypothetical dihydrolases are not known. However, SADH is the best characterized example of this category and will be discussed in detail.

8.05.4.1 Succinylarginine Dihydrolase

8.05.4.1.1 *Biological function*

The arginine succinyl transferase (AST) pathway is a catabolic pathway used by some microorganisms as their primary route for arginine catabolism.[135] It consists of five enzymes found in a single operon: N^α-succinyl-transferase (AST); N^α-succinylarginine dihydrolase (SADH); N^α-succinylornithine 5-aminotransferase (SOAT), N^α-succinylglutamate 5-semialdehyde dehydrogenase (SGSD), and N^α-succinylglutamate desuccinylase (SGDS) (**Figure 25**). The second enzyme in the AST pathway, SADH, shows both sequence and structural homology to members of the pentein superfamily and catalyzes a related reaction.[136,137] However, instead of producing a ureido product (e.g., citrulline)-like related hydrolytic enzymes, SADH instead produces ammonia, N^α-succinylornithine (**32**), and CO_2 (or HCO_3^-) and so is designated a dihydrolase.

Unlike the ADI pathway, which generates carbamoyl phosphate, the AST pathway instead generates glutamate and ammonia, and is necessary in some organisms for arginine degradation when nitrogen sources are limited. For example, *E. coli* catabolizes approximately 97% of arginine through the AST pathway.[135] Nitrogen limitation can upregulate the expression of proteins encoded by in the AST operon, and these enzymes have been suggested as drug targets in pathogenic strains.[135–137]
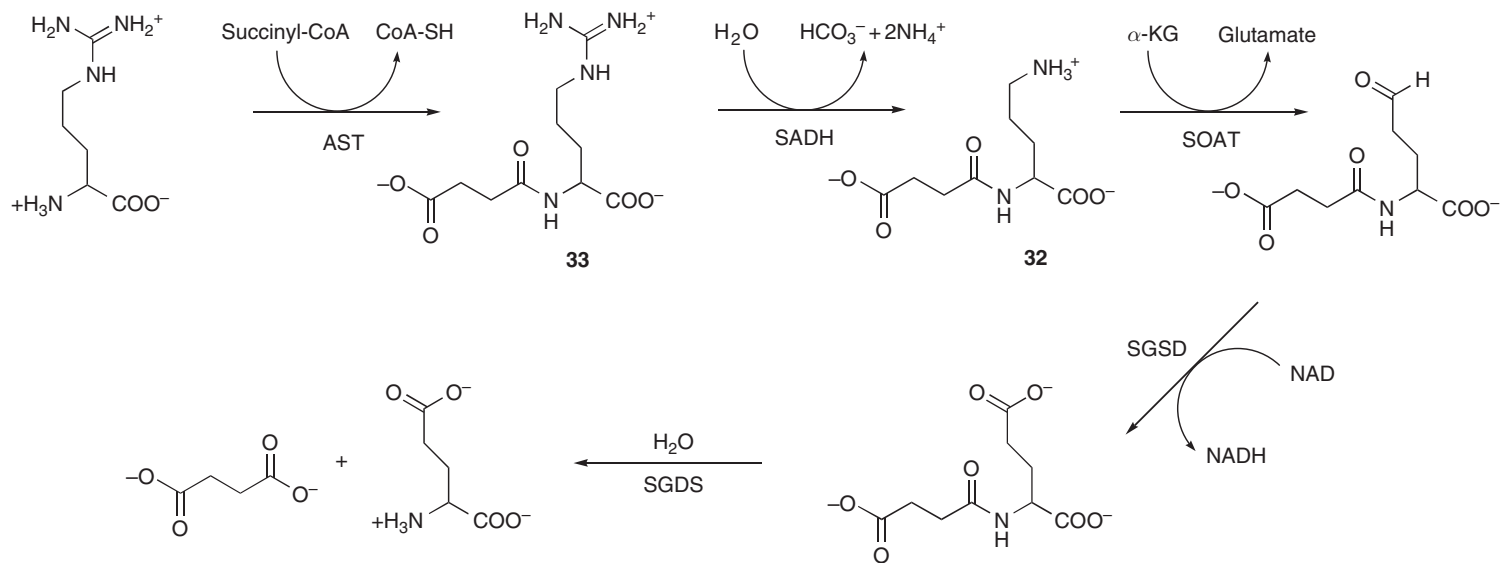
Figure 25 The arginine succinyl transferase (AST) pathway consists of five enzymes: N^{α}-succinyltransferase (AST), N^{α}-succinylarginine dihydrolase (SADH), N^{α}-succinylornithine 5-aminotransferase (SOAT), N^{α}-succinylglutamate 5-semialdehyde dehydrogenase (SGSD), and N^{α}-succinylglutamate desuccinylase (SGDS).

8.05.4.1.2 *Protein structure*

The best characterized SADH is the homodimeric AstB protein from *E. coli*, for which three structures have been reported: wild-type protein, wild-type protein in complex with N^α-succinylornithine (**32**), and a core Cys mutant in complex with N^α-succinylarginine (**33**).[137] The pentein fold is visible in each monomeric unit of the protein, with sequence inserts found in the first subdomain and the C-terminus (**Figure 26**). A disordered loop in the first insert becomes ordered in the complexed structures and makes numerous interactions with the N^α-succinyl moiety of bound ligands. In addition, a potassium ion is coordinated at a site distant from the active site, but the structural or functional importance of this ion has not been determined.

Most of the core residues observed in ADI are conserved in SADH, with at least one notable exception (**Figure 27**). The lateral aspartate is replaced in SADH with a conserved asparagine. This substitution has been proposed to enable the enzyme to catalyze a dihydrolase reaction as opposed to the single hydrolysis step catalyzed by ADI, DDAH, and PAD, although a functional switch involving mutation of this position has not been demonstrated. Nonetheless, functional annotations of dihydrolase activity have successfully been made based upon an asparagine in this position.[6,136]
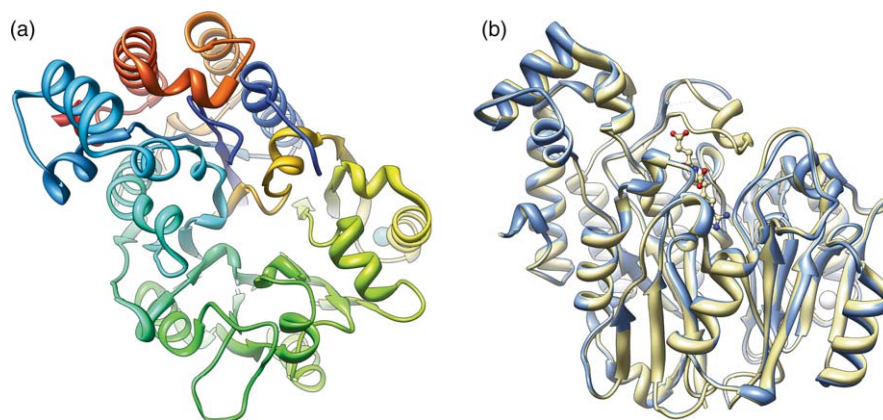
Figure 26 Structure of N^α-succinylarginine dihydrolase. The pseudosymmetry of the pentein fold is seen, with additional insertions in the first subdomain and at the C-terminus (a). An overlay of liganded and unliganded structures (b) shows that a disordered loop in the apo structure (blue) is ordered upon substrate binding (tan) and makes several hydrophobic and hydrogen-bonding interactions with the N^α-succinyl moiety of the bound ligand. In both structures, the sphere represents a bound potassium ion. This figure was prepared using PDB accession codes 1YNF and 1YNI, respectively.
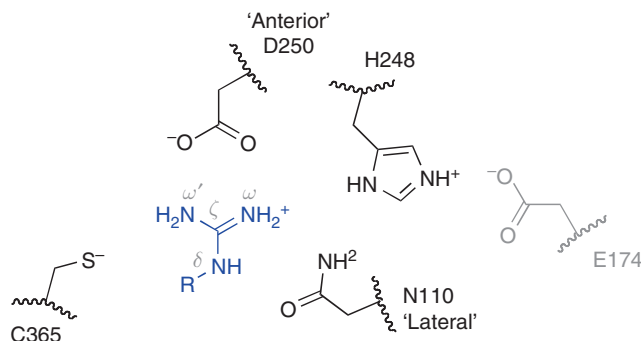
Figure 27 Active-site residues of SADH that interact with the guanidine of the substrate, N^α-succinylarginine (**33**) (shown in blue). 'Core' residues are shown in black and a neighboring Glu in gray. N110 is found in the same position as the 'lateral' Asp residues of ADI, and D250 serves as the 'anterior' Asp. Amino acid numbering is taken from the *Escherichia coli* SADH protein.

8.05.4.1.3 *Catalytic mechanism*

Only limited mechanistic data are available for SADH enzymes. Crude cell lysates of *Pseudomonas cepacia* and *P. aeruginosa* show consumption of N^α-succinylarginine with the concomitant production of N^α-succinylornithine and roughly two equivalents of ammonia.[138,139] Small amounts of N^α-succinylcitrulline (**34**) were also isolated, but appear to be a minor by-product or an intermediate that does not accumulate. However, N^α-succinylcitrulline does not serve as a substrate when added to crude mixtures under the same conditions as the reactions containing N^α-succinylarginine. Urea (**35**) formation was not detected, and urea added to crude lysates does not result in ammonia production, also disfavoring urea as an intermediate in the pathway. Disruption of the *E. coli* gene for SADH results in a loss of dihydrolase activity in crude cell lysates, as gauged by ammonia production.[135] Purified SADH produces ammonia from synthetic N^α-succinylarginine, and cocrystallization with substrate results in active-site electron density that can be fit to N^α-succinylornithine.[137] However, the stoichiometry of the reaction, and the identity and order of release of the reaction products have not been determined using purified enzyme.

The detailed catalytic mechanism of dihydrolases remains an open question. Considering the mechanistic constraints above, at least three possible reaction pathways can be proposed for SADH (**Figure 28**). The first hydrolysis step could form urea (path a), N^α-succinylcitrulline (path b), or N^α-succinylornithine (path c). The urea route (path a) is disfavored because no urease activity is detected. Based on structural similarities of SADH to ADI, the N^α-succinylcitrulline route (path b) was proposed.[137] This mechanism predicts that N^α-succinylcitrulline would be a good substrate for the enzyme, an expectation inconsistent with experiments performed on crude lysates. Finally, the N^α-succinylornithine route (path c) is consistent with all published experiments, but none of the proposed covalent intermediates have been isolated to date. Path c is distinguished from path a because no free urea is produced.

In future studies, the function of the substitution of Asn for the lateral Asp at the active site will be of particular interest. This change presumably alters charge distribution in the Michaelis complex and any enzyme-bound intermediates. It is tempting to propose that this single substitution is responsible for a change in mechanism,[137] presumably by favoring the N^δ as a leaving group over the N^ω atoms, but there are likely more subtle differences that account for the mechanistic divergence from the related hydrolases typified by ADI.

8.05.5 Amidinotransferases

In contrast to the hydrolyases described above, the amidinotransferases instead catalyze the transfer of a one-carbon amidine unit ($-C(=N)NH_2$) from the side chain of arginine to different acceptor amines. Various amidinotransferases have been found to participate in the biosynthesis of natural products such as streptomycin
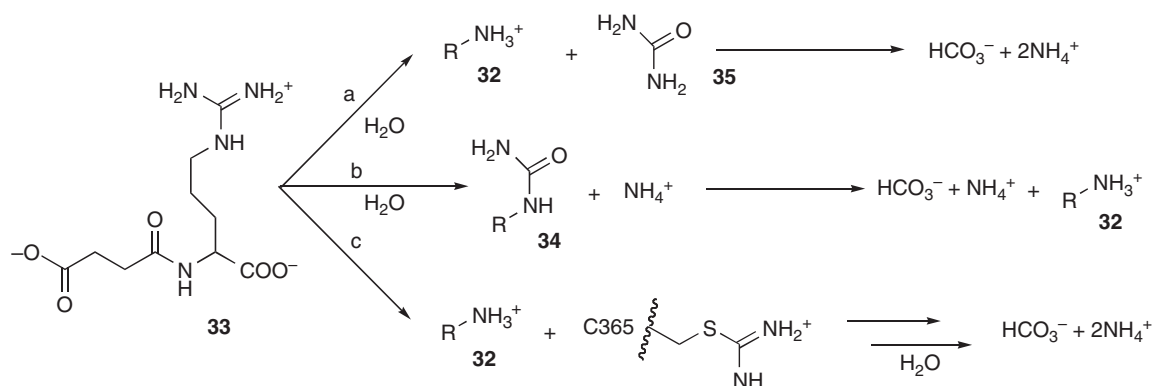
Figure 28 Three putative reaction pathways of N^α-succinylarginine dihydrolase. The urea (**35**)-forming pathway (a) is disfavored due to a lack of detectable urease activity. Based on structural similarities to ADI enzymes, the N^α-succinylcitrulline (**34**)-forming pathway (b) has been proposed, but N^α-succinylcitrulline does not serve as an alternative substrate. The N^α-succinylornithine (**32**)-producing pathway (c) is consistent with the current data, but no intermediates have been isolated. An alternative pathway not shown is elimination of ammonia and retention of a covalent Cys$_{365}$-S-(1-iminomethyl)-L-ornithine intermediate. In the final step of each pathway, the formation of bicarbonate or carbon dioxide has not been distinguished.
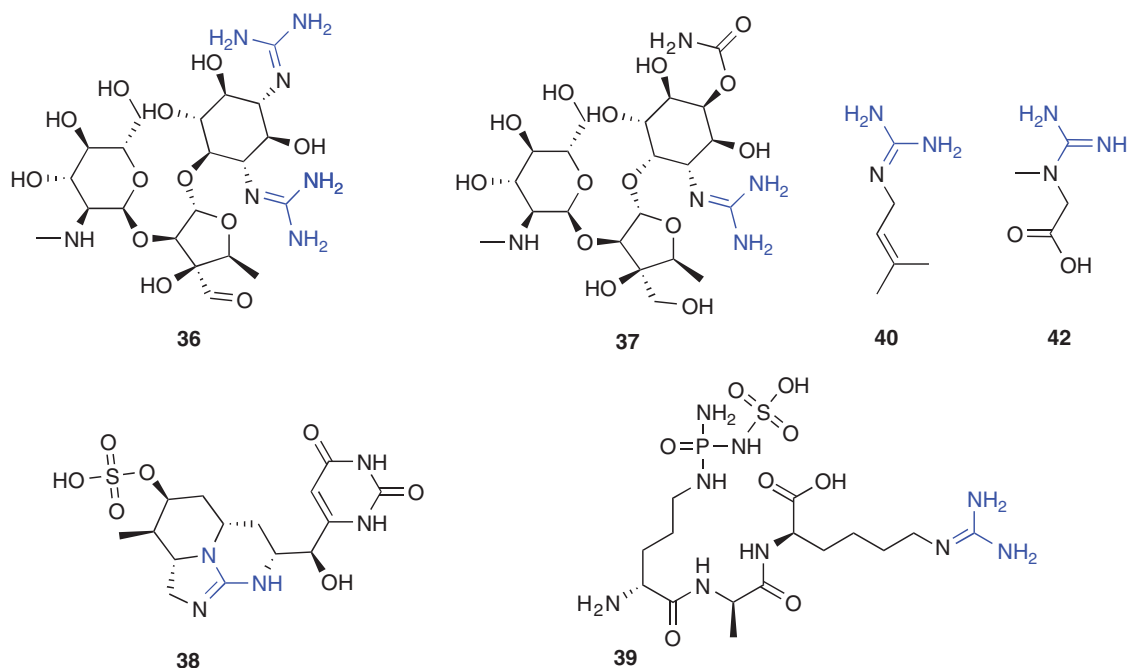
Figure 29 Natural products derived from pathways containing amidinotransferase enzymes from the pentein superfamily, including streptomycin (**36**), bluensomycin (**37**), cylindrospermopsin (**38**), phaseolotoxin (**39**), galegine (**40**), and creatine (**42**). In each case, the amidine moiety installed by the transfer reaction is shown in blue.

(**36**), bluensomycin (**37**), cylindrospermopsin (**38**), phaseolotoxin (**39**), and galegine (**40**) (**Figure 29**).[140–146] However, L-arginine:glycine amidinotransferase (AGAT) is the best characterized example and will be discussed here in detail.

8.05.5.1 L-Arginine:Glycine Amidinotransferase

8.05.5.1.1 *Biological function*

AGAT catalyzes the first and rate-limiting step in the biosynthesis of creatine (**Figure 30**).[147,148] AGAT catalyzes the transfer of an amidino group from arginine to the N^α-amine of glycine to yield L-ornithine (**3**) and guanidinoacetate (**41**). Guanidinoacetate methyltransferase (GAMT) then catalyzes the second and final step, using *S*-adenosylmethionine (AdoMet) as a methyl donor to yield creatine (**42**) and *S*-adenosylhomocysteine. Creatine and phosphocreatine (**43**) are important molecules in the energy metabolism of the nerve and muscle
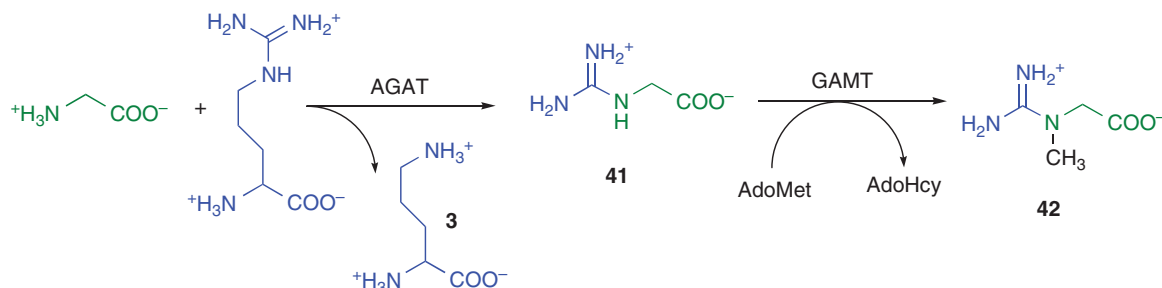
Figure 30 The creatine (**42**) biosynthesis pathway consists of two enzymes: L-arginine:glycine amidinotransferase (AGAT) and guanidinoacetate *N*-methyltransferase (GAMT). Atoms derived from glycine are shown in green, arginine in blue, and *S*-adenosylmethionine (AdoMet) in black.

tissues and can serve as a buffer against rapid changes in ATP/ADP ratios.[147] At least three inborn errors of creatine metabolism are known, and these can lead to epilepsy and mental retardation, among other effects. One of these errors occurs within the AGAT gene and consists of a homozygous G to A mutation that changes the W149 codon into a stop codon, resulting in a nonfunctional transcript that does not accumulate.[149]

8.05.5.1.2   *Protein structure*

Recombinant dimeric human AGAT is structurally well characterized by Huber and coworkers, with numerous mutant and ligand-bound X-ray crystal structures that provide mechanistic insight into how the transfer reaction occurs. Each monomer of AGAT is described as a 'basket with handles', with the fivefold pseudosymmetrical pentein fold serving as the basket, and various loop insertions on the 'top' face (opposite the N- and C-termini) as the handles (**Figure 31**).[150,151] Several of these loops change conformation upon ligand binding, with the apo- or glycine-bound structures showing a 'closed' conformation and the arginine- or ornithine-bound structures showing the 'open' form (**Figure 32**).[152,153] These conformational changes are thought to facilitate both ligand binding and product dissociation. It is interesting to note that a related enzyme, L-arginine:inosamine-phosphate amidinotransferase (StrB1) from *Streptomyces griseus*, is much slower and lacks some of these loop insertions. Homologous loop
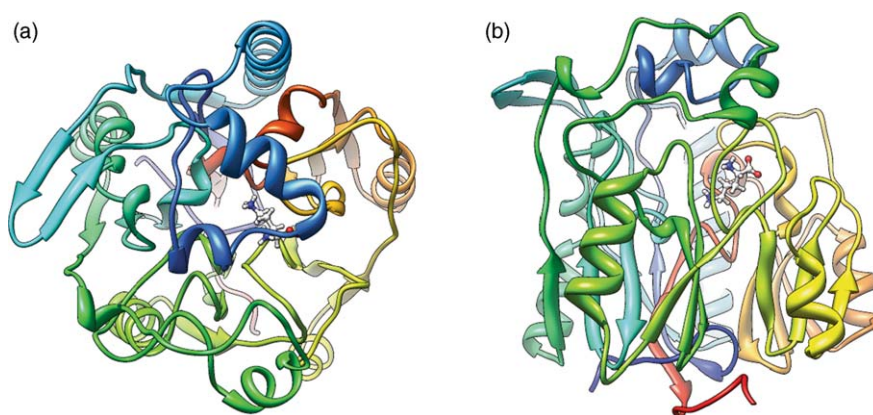
Figure 31   Structure of L-arginine:glycine amidinotransferase in complex with L-ornithine. A modified pentein fold can be seen in a 'top' view (a) and the 'side' view (b), with large sequence inserts found in the first and second subdomains. The bound L-ornithine inhibitor, shown as a ball-and-stick model, marks the substrate-binding site. This figure was prepared using coordinates from PDB accession code 3JDW.
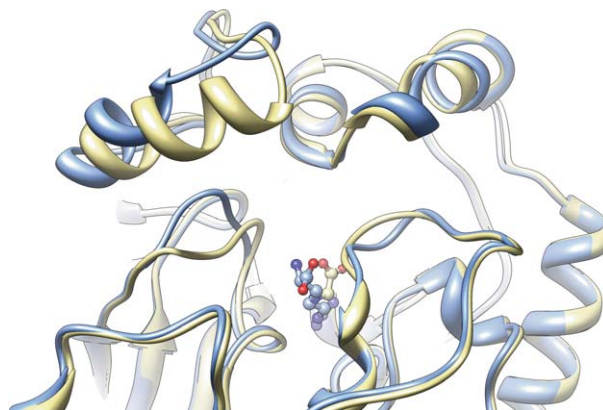
Figure 32   Closed (tan) and open (blue) conformations of L-arginine:glycine amidinotransferase complexed with L-arginine (blue) and glycine (tan), respectively. Ligands are shown in ball-and-stick model, and the proteins are shown using ribbon models. The figure was constructed using coordinates from PDB accession codes 4JDW and 5JDW, respectively.
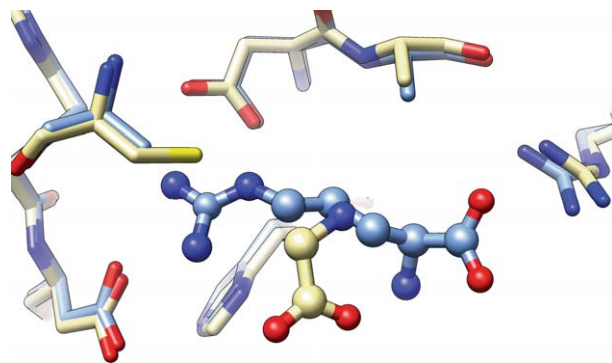
Figure 33 Structural overlay of L-arginine:glycine amidinotransferase complexed with L-arginine (blue) and glycine (tan). The arginine complex has a mutation of the core Cys residue to Ala. The glycine and L-arginine substrates appear to have overlapping binding sites. Ligands are shown using ball-and-stick models and protein residues by stick models.
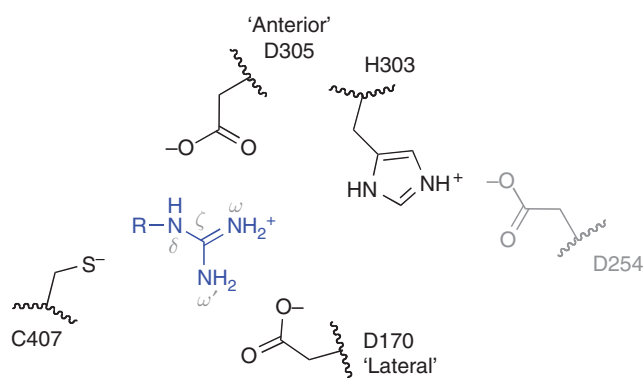
Figure 34 Active-site residues of AGAT that interact with the guanidine of the substrate, L-arginine (shown in blue). 'Core' residues are shown in black and a neighboring Asp in gray. Lateral and anterior designations are assigned to match those of the homologous ADI core residues. Core residues are conserved in AGAT, but the substrate orientation is now rotated to position the N^{δ}–C^{ζ} bond for cleavage.

deletions made in AGAT slow substrate turnover to similar rates, presumably by disrupting the ligand-induced conformational changes and their associated rate-enhancing features.[146,153]

Structures of a wild-type AGAT with bound glycine and of a core Cys mutant of AGAT with bound L-arginine help to identify the substrate-binding pockets.[151] The two substrates, glycine and L-arginine, have overlapping footprints in the same binding pocket, ruling out the formation of a ternary complex (**Figure 33**). The same core residues observed in the ADI structure are conserved in the AGAT structure, but notably the substrate is rotated approximately 120° around the C^{ζ} atom and now places the N^{δ} nitrogen where one of the N^{ω} nitrogens is placed in the ADI (**Figure 34**). This rotated placement of the substrate allows the same catalytic machinery that breaks the N^{ω}–C^{ζ} bond during ADI catalysis instead to break the N^{δ}–C^{ζ} bond during AGAT catalysis. Effectively, the anterior and lateral aspartates have swapped roles, except the Asp that coordinates the substrate's terminal nitrogens now appears to coordinate only one of the N^{ω} groups. This monodentate chelation is more reminiscent of the monodentate N^{ω} chelation observed in DDAH than the bidentate $N^{\omega},N^{\omega'}$ chelation seen in ADI. The core Cys and His residues of AGAT are structurally conserved with ADI, as is an additional Asp residue placed within hydrogen-bonding distance of the core His (**Figure 34**).

8.05.5.1.3 Catalytic mechanism

The amidinotransferases provide an interesting twist on the mechanism observed in the hydrolases above.[151,152] By conserving the core residues, but reorienting the substrate, the same catalytic machinery can be used, yet cleavage of a different N–C^{ζ} bond can be selected (**Figure 35**). Briefly, as was observed in ADI, PAD, and
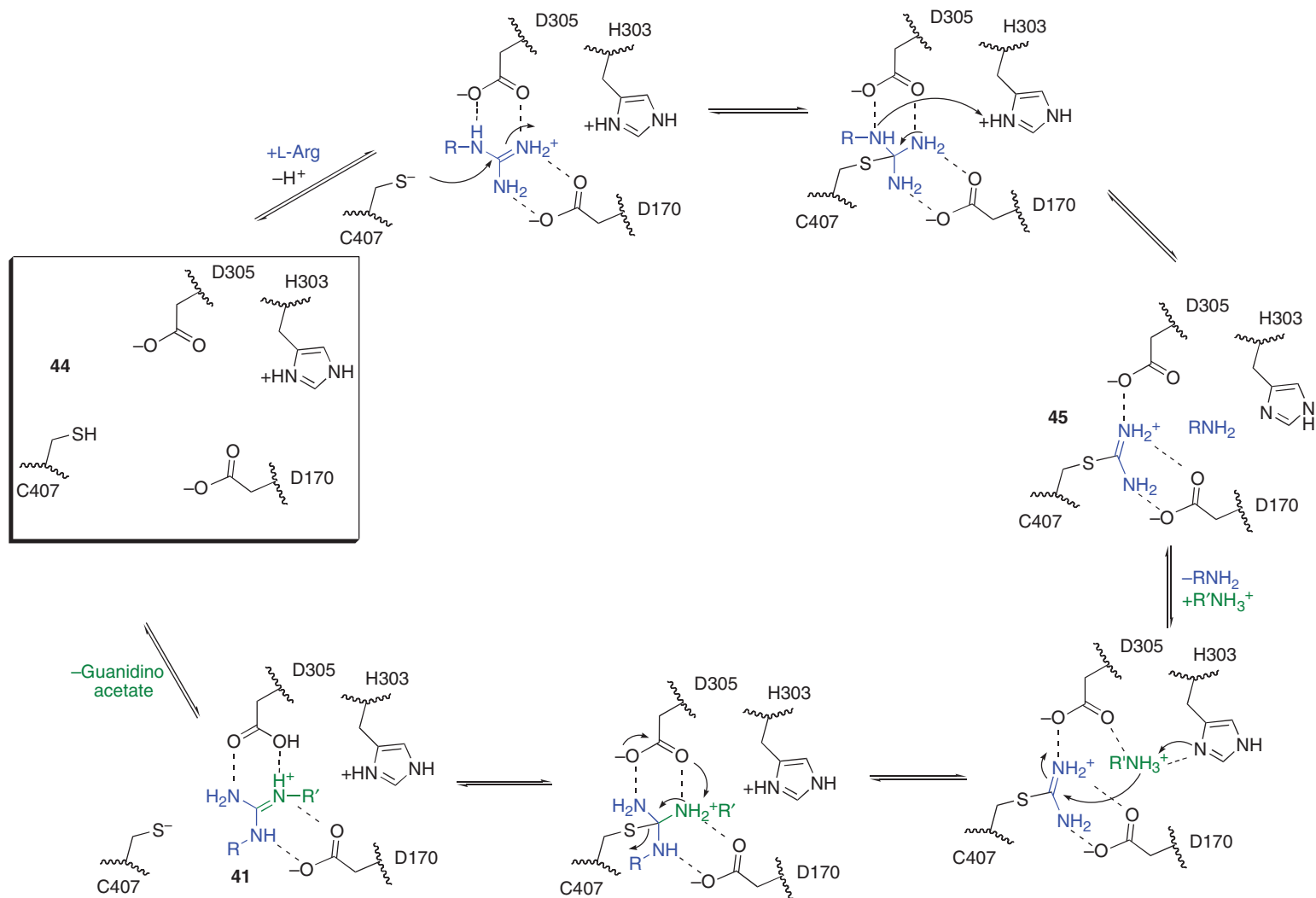
Figure 35 Proposed catalytic mechanism of AGAT. Substrate L-arginine is shown in blue and glycine in green, and 'core' residues are shown in black. Amino acid numbering is from human AGAT. The apo form of the enzyme is boxed.

DDAH, the core Cys residue of AGAT is proposed to be predominantly protonated in the resting state of the enzyme (**44**) and is possibly deprotonated by the neighboring Asp residue (D305), prior to its attack on the substrate's C^ζ carbon (not shown).[152] Formation of the resulting tetrahedral adduct is followed by collapse and breaking of the N^δ–C^ζ bond to produce L-ornithine (**3**) and an sp^2 hybridized thiouronium intermediate (**45**) bearing only one carbon derived from the L-arginine substrate. The core His residue presumably facilitates bond cleavage by donating a proton to the leaving group. The covalent enzyme intermediate (**45**) has been called an 'activated urea' and was characterized by trapping under acidic conditions.[154,155] After L-ornithine dissociates from the enzyme, glycine can enter and bind at a site adjacent to the covalent intermediate. Ping-pong kinetics, the ligand-bound structures described above, and trapping of the covalent intermediate all argue against a direct transfer mechanism and are consistent with a covalent enzyme intermediate.[151,152,154,156] The N^α group of glycine, assisted by the core His acting as a general base, can then attack the thiouronium intermediate, subsequently leading to the release of the final guanidinoacetate product (**41**).

It is not currently known what molecular features of AGAT work to stabilize the thiouronium intermediate against hydrolysis, in contrast to the labile intermediates observed in ADI. Arginine-bound structures of ADI and AGAT in which the active-site cysteines are replaced by alanines have been reported and give some clues.[23,151] An overlay of these active sites highlights the alternative rotated placement of the arginine substrates and shows that the core active-site residues are conserved (**Figure 36**). However, the relative orientation of the core residues to one another is slightly different between the two complexes, suggesting that subtle differences in orientation may contribute to the stability of the intermediate. These differences do not totally preclude a hydrolytic reaction in AGAT; hydrolysis of arginine by hog kidney AGAT is observed in the absence of the glycine cosubstrate, although this hydrolysis shunt is quite slower (<0.01-fold) than the corresponding transfer reaction.[157]

8.05.5.1.4 Regulation, inhibitors, alternative substrates

AGAT activity is highly regulated *in vivo* to maintain normal creatine metabolism. Expression of AGAT is enhanced by growth hormone and thyroxine, and feedback inhibited by creatine.[147] The gene for mouse AGAT is regulated in an epigenetic fashion, with expression occurring only from the maternal allele in mouse extraembyronic tissues.[158] In cases of hyperornithinemia, pathological ornithine accumulation (650 $\mu mol\, l^{-1}$
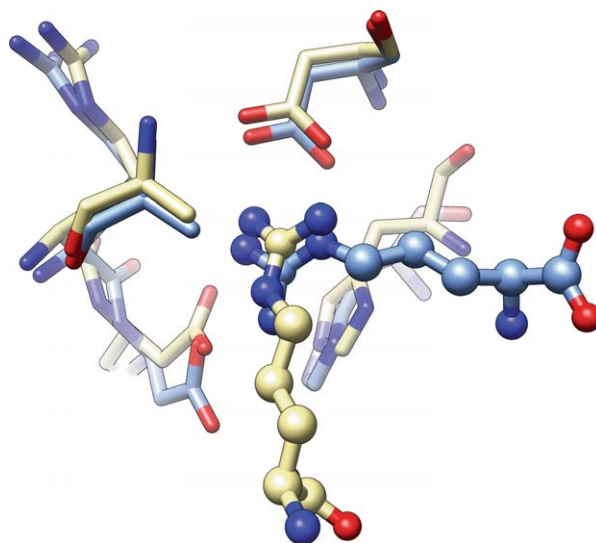
Figure 36 Relative orientations of substrates and core residues in AGAT (blue) and ADI (tan). In both complexes, the core Cys residue has been mutated to Ala (seen at the left of the figure). The anterior Asp residues are positioned at the top, and the core His and lateral Asp residues at the bottom right and left of the figure, respectively. Bound L-arginine substrates are shown as ball-and-stick models and protein residues as sticks. The figure was constructed using coordinates from PDB accession codes 2A9G and 4JDW, respectively.

to 1.3 mmol l^{-1}) inhibits AGAT ($K_i = 250$ µmol l^{-1}) and the subsequent block in creatine biosynthesis contributes to the atrophy of the muscle and eye tissues, which leads to blindness in this disorder.[159]

In addition to arginine and glycine, AGAT has been shown to tolerate alternative compounds for both the donor and acceptor substrates. Amidino-donating substrates include arginine, guanidinoacetic acid, and canavanine, and acceptor substrates include glycine, ornithine, canaline, ethanolamine, 4-aminobutyric acid, lysine, 5-aminovaleric acid, 3-aminopropionic acid, taurine, and hydroxylamine.[155,157,160,161] These results provide an interesting contrast to ADI, which is quite selective, and can be inhibited by canavanine.[34] This contrast is easily understood by a comparison of mechanisms. Whereas ADI forms an alternative oxythiouronium intermediate (**10**) that is not well hydrolyzed, AGAT catalyzes the cleavage of a different C$^\zeta$–N bond and forms a covalent intermediate identical to that of the normal reaction (**45**).

8.05.6   Conclusion

The pentein superfamily includes a diverse set of proteins, from binding proteins to enzymes that display hydrolase, dihydrolase, and amidinotransferase activities.[5,6] Although these enzymatic activities catalyze different bond cleavage and transfer reactions, they appear to share several common structural and mechanistic themes. Each shows a common β/α propeller structure with fivefold pseudosymmetry and is elaborated by different loops to accommodate various substrates. Each has a very polar active site and uses a core of similar residues to catalyze a reaction on a guanidinium group. Although the core Cys is used as a nucleophile, it is predominantly protonated and hence not very reactive in the resting enzyme. Through a combination of reverse protonation or substrate-assisted mechanisms, attack of this Cys residue on the substrate's guanidinium carbon results in the formation of a covalent thiouronium adduct. Substrate orientations differ between the hydrolase and amidinotransferase reactions, allowing the use of the same catalytic machinery to break different C$^\zeta$–N bonds. Although many mechanistic features remain unexplored, it is interesting to note that most of the hydrolytic enzymes retain an exit channel for their leaving groups, and that this channel overlays with the substrate-binding channel of the amidinotransferases (**Figure 37**). Perhaps the exit channel is actually a vestigial substrate-binding channel derived from a common ancestor with both hydrolytic and transferase activities. This hypothesis is consistent with the observation of a low-level hydrolysis shunt in AT and the artificial recapitulation of an AT reaction in DDAH.[70,157] Therefore, based on the structural and mechanistic
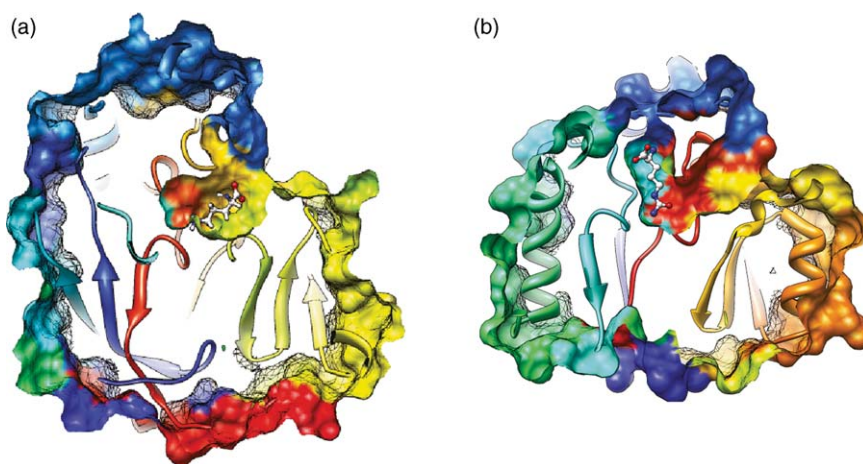
Figure 37   Cut-away surface-coated ribbon diagrams of AGAT (a) and DDAH (b) with active-site ligands bound. Ornithine is bound to AGAT, and citrulline is bound to a core Cys mutant of DDAH. Proteins are shown in approximately the same orientation with respect to their homologous core residues. The exit channel for dimethylamine in DDAH (b) is located in the same position as the substrate-binding channel of AGAT (a), suggesting that the exit channel may be a vestigial binding site derived from an ancestral enzyme with both hydrolytic and amidiotransfer activities. The figure was constructed using coordinates from PDB accession codes 3JDW and 1H70, respectively.

commonalities of this diverse set of proteins, the pentein superfamily can be classified as a mechanistically diverse superfamily.[162] Understanding how these different enzymatic activities have diverged and why the common features of the mechanisms have been retained will help drug discovery efforts and will provide insight into the physiology and evolutionary history of this protein superfamily.

Acknowledgments

References

1. C. M. Groft; R. Beckmann; A. Sali; S. K. Burley, *Nat. Struct. Biol.* **2000**, *7*, 1156–1164.
2. S. A. Teichmann; A. G. Murzin; C. Chothia, *Curr. Opin. Struct. Biol.* **2001**, *11*, 354–363.
3. M. Paoli, *Nat. Struct. Biol.* **2001**, *8*, 744–745.
4. E. F. Pettersen; T. D. Goddard; C. C. Huang; G. S. Couch; D. M. Greenblatt; E. C. Meng; T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
5. H. Shirai; T. L. Blundell; K. Mizuguchi, *Trends Biochem. Sci.* **2001**, *26*, 465–468.
6. H. Shirai; Y. Mokrab; K. Mizuguchi, *Proteins* **2006**, *64*, 1010–1023.
7. K. Si; U. Maitra, *Mol. Cell Biol.* **1999**, *19*, 1416–1426.
8. M. Ceci; C. Gaviraghi; C. Gorrini; L. A. Sala; N. Offenhauser; P. C. Marchisio; S. Biffo, *Nature* **2003**, *426*, 579–584.
9. V. Gandin; A. Miluzio; A. M. Barbieri; A. Beugnet; H. Kiyokawa; P. C. Marchisio; S. Biffo, *Nature* **2008**, *455*, 684–688.
10. P. Rosso; G. Cortesina; F. Sanvito; A. Donadini; B. Di Benedetto; S. Biffo; P. C. Marchisio, *Head Neck* **2004**, *26*, 408–417.
11. F. Sanvito; F. Vivoli; S. Gambini; G. Santambrogio; M. Catena; E. Viale; F. Veglia; A. Donadini; S. Biffo; P. C. Marchisio, *Cancer Res.* **2000**, *60*, 510–516.
12. T. P. Chendrimada; K. J. Finn; X. Ji; D. Baillat; R. I. Gregory; S. A. Liebhaber; A. E. Pasquinelli; R. Shiekhattar, *Nature* **2007**, *447*, 823–828.
13. P. Ray; U. Basu; A. Ray; R. Majumdar; H. Deng; U. Maitra, *J. Biol. Chem.* **2008**, *283*, 9681–9691.
14. M. Zuniga; G. Perez; F. Gonzalez-Candelas, *Mol. Phylogenet. Evol.* **2002**, *25*, 429–444.
15. R. Cunin; N. Glansdorff; A. Pierard; V. Stalon, *Microbiol. Rev.* **1986**, *50*, 314–352.
16. A. Casiano-Colon; R. E. Marquis, *Appl. Environ. Microbiol.* **1988**, *54*, 1318–1324.
17. P. J. Schofield; M. R. Edwards; J. Matthews; J. R. Wilson, *Mol. Biochem. Parasitol.* **1992**, *51*, 29–36.
18. L. A. Knodler; E. O. Sekyere; T. S. Stewart; P. J. Schofield; M. R. Edwards, *J. Biol. Chem.* **1998**, *273*, 4470–4477.
19. J. L. Weickmann; D. E. Fahrney, *J. Biol. Chem.* **1977**, *252*, 2615–2620.
20. H. Takaku; M. Takase; S. Abe; H. Hayashi; K. Miyazaki, *Int. J. Cancer* **1992**, *51*, 244–249.
21. H. Baur; E. Luethi; V. Stalon; A. Mercenier; D. Haas, *Eur. J. Biochem.* **1989**, *179*, 53–60.
22. A. Galkin; L. Kulakova; E. Sarikaya; K. Lim; A. Howard; O. Herzberg, *J. Biol. Chem.* **2004**, *279*, 14001–14008.
23. A. Galkin; X. Lu; D. Dunaway-Mariano; O. Herzberg, *J. Biol. Chem.* **2005**, *280*, 34080–34087.
24. K. Das; G. H. Butler; V. Kwiatkowski; A. D. Clark, Jr.; P. Yadav; E. Arnold, *Structure (Camb.)* **2004**, *12*, 657–667.
25. X. Lu; L. Li; R. Wu; X. Feng; Z. Li; H. Yang; C. Wang; H. Guo; A. Galkin; O. Herzberg; P. S. Mariano; B. M. Martin; D. Dunaway-Mariano, *Biochemistry* **2006**, *45*, 1162–1172.
26. Y. Wei; H. Zhou; Y. Sun; Y. He; Y. Luo, *Proteins* **2007**, *66*, 740–750.
27. C. Wang; D. Xu; L. Zhang; D. Xie; H. Guo, *J. Phys. Chem. B* **2007**, *111*, 3267–3273.
28. X. Lu; A. Galkin; O. Herzberg; D. Dunaway-Mariano, *J. Am. Chem. Soc.* **2004**, *126*, 5374–5375.
29. D. W. Smith; D. E. Fahrney, *Biochem. Biophys. Res. Commun.* **1978**, *83*, 101–106.
30. D. W. Smith; R. L. Ganaway; D. E. Fahrney, *J. Biol. Chem.* **1978**, *253*, 6016–6020.
31. J. L. Weickmann; M. E. Himmel; D. W. Smith; D. E. Fahrney, *Biochem. Biophys. Res. Commun.* **1978**, *83*, 107–113.
32. L. Li; Z. Li; C. Wang; D. Xu; P. S. Mariano; H. Guo; D. Dunaway-Mariano, *Biochemistry* **2008**, *47*, 4721–4732.
33. R. E. Marquis; G. R. Bender; D. R. Murray; A. Wong, *Appl. Environ. Microbiol.* **1987**, *53*, 198–200.
34. L. Li; Z. Li; D. Chen; X. Lu; X. Feng; E. C. Wright; N. O. Solberg; D. Dunaway-Mariano; P. S. Mariano; A. Galkin; L. Kulakova; O. Herzberg; K. B. Green-Church; L. Zhang, *J. Am. Chem. Soc.* **2008**, *130*, 1918–1931.
35. G. A. Rosenthal, *Amino Acids* **2001**, *21*, 319–330.
36. A. Boyar; R. E. Marsh, *J. Am. Chem. Soc.* **1982**, *104*, 1995–1998.
37. X. Lu; L. Li; X. Feng; Y. Wu; D. Dunaway-Mariano; J. R. Engen; P. S. Mariano, *J. Am. Chem. Soc.* **2005**, *127*, 16412–16413.
38. H. Gong; F. Zolzer; G. von Recklinghausen; J. Rossler; S. Breit; W. Havers; T. Fotsis; L. Schweigerer, *Biochem. Biophys. Res. Commun.* **1999**, *261*, 10–14.
39. C. M. Ensor; F. W. Holtsberg; J. S. Bomalaski; M. A. Clark, *Cancer Res.* **2002**, *62*, 5443–5450.
40. C. Y. Yoon; Y. J. Shim; E. H. Kim; J. H. Lee; N. H. Won; J. H. Kim; I. S. Park; D. K. Yoon; B. H. Min, *Int. J. Cancer* **2007**, *120*, 897–905.

41. F. Izzo; P. Marra; G. Beneduce; G. Castello; P. Vallone; V. De Rosa; F. Cremona; C. M. Ensor; F. W. Holtsberg; J. S. Bomalaski; M. A. Clark; C. Ng; S. A. Curley, *J. Clin. Oncol.* **2004**, *22*, 1815–1822.
42. P. A. Ascierto; S. Scala; G. Castello; A. Daponte; E. Simeone; A. Ottaiano; G. Beneduce; V. De Rosa; F. Izzo; M. T. Melucci; C. M. Ensor; A. W. Prestayko; F. W. Holtsberg; J. S. Bomalaski; M. A. Clark; N. Savaraj; L. G. Feun; T. F. Logan, *J. Clin. Oncol.* **2005**, *23*, 7660–7668.
43. I. S. Park; S. W. Kang; Y. J. Shin; K. Y. Chae; M. O. Park; M. Y. Kim; D. N. Wheatley; B. H. Min, *Br. J. Cancer* **2003**, *89*, 907–914.
44. T. Ogawa; M. Kimoto; K. Sasaoka, *J. Biol. Chem.* **1989**, *264*, 10205–10209.
45. J. M. Leiper; J. Santa Maria; A. Chubb; R. J. MacAllister; I. G. Charles; G. S. Whitley; P. Vallance, *Biochem. J.* **1999**, *343* (Pt. 1), 209–214.
46. C. T. Tran; M. F. Fox; P. Vallance; J. M. Leiper, *Genomics* **2000**, *68*, 101–105.
47. D. Wang; P. S. Gill; T. Chabrashvili; M. L. Onozato; J. Raggio; M. Mendonca; K. Dennehy; M. Li; P. Modlinger; J. Leiper; P. Vallance; O. Adler; A. Leone; A. Tojo; W. J. Welch; C. S. Wilcox, *Circ. Res.* **2007**, *101*, 627–635.
48. M. Knipp, *Chembiochem* **2006**, *7*, 879–889.
49. A. J. Cardounel; J. L. Zweier, *J. Biol. Chem.* **2002**, *277*, 33995–34002.
50. P. Vallance; J. Leiper, *Nat. Rev. Drug Discov.* **2002**, *1*, 939–950.
51. J. Leiper; M. Nandi; B. Torondel; J. Murray-Rust; M. Malaki; B. O'Hara; S. Rossiter; S. Anthony; M. Madhani; D. Selwood; C. Smith; B. Wojciak-Stothard; A. Rudiger; R. Stidwill; N. Q. McDonald; P. Vallance, *Nat. Med.* **2007**, *13*, 198–203.
52. N. M. Olken; Y. Osawa; M. A. Marletta, *Biochemistry* **1994**, *33*, 14784–14791.
53. D. Tsikas; R. H. Boger; J. Sandmann; S. M. Bode-Boger; J. C. Frolich, *FEBS Lett.* **2000**, *478*, 1–3.
54. P. Vallance; A. Leone; A. Calver; J. Collier; S. Moncada, *Lancet* **1992**, *339*, 572–575.
55. R. H. Boger; S. M. Bode-Boger; A. Szuba; P. S. Tsao; J. R. Chan; O. Tangphao; T. F. Blaschke; J. P. Cooke, *Circulation* **1998**, *98*, 1842–1847.
56. K. Y. Lin; A. Ito; T. Asagami; P. S. Tsao; S. Adimoolam; M. Kimoto; H. Tsuji; G. M. Reaven; J. P. Cooke, *Circulation* **2002**, *106*, 987–992.
57. M. C. Stuhlinger; R. K. Oka; E. E. Graf; I. Schmolzer; B. M. Upson; O. Kapoor; A. Szuba; M. R. Malinow; T. C. Wascher; O. Pachinger; J. P. Cooke, *Circulation* **2003**, *108*, 933–938.
58. J. P. Cooke, *Circulation* **2004**, *109*, 1813–1818.
59. V. Kostourou; S. P. Robinson; G. S. Whitley; J. R. Griffiths, *Cancer Res.* **2003**, *63*, 4960–4966.
60. S. Rossiter; C. L. Smith; M. Malaki; M. Nandi; H. Gill; J. M. Leiper; P. Vallance; D. L. Selwood, *J. Med. Chem.* **2005**, *48*, 4670–4678.
61. M. Tanaka; K. Sydow; F. Gunawan; J. Jacobi; P. S. Tsao; R. C. Robbins; J. P. Cooke, *Circulation* **2005**, *112*, 1549–1556.
62. J. Santa Maria; P. Vallance; I. G. Charles; J. M. Leiper, *Mol. Microbiol.* **1999**, *33*, 1278–1279.
63. M. Galimand; M. Gamper; A. Zimmermann; D. Haas, *J. Bacteriol.* **1991**, *173*, 1598–1606.
64. T. Jahns; H. Kaltwasser, *FEMS Microbiol. Lett.* **1990**, *60*, 131–135.
65. K. Kakishima; A. Shiratsuchi; A. Taoka; Y. Nakanishi; Y. Fukumori, *Biochem. Biophys. Res. Commun.* **2007**, *355*, 587–591.
66. R. B. Dowling; R. Newton; A. Robichaud; P. J. Cole; P. J. Barnes; R. Wilson, *Am. J. Respir. Cell Mol. Biol.* **1998**, *19*, 950–958.
67. C. D. Sohaskey, *J. Bacteriol.* **2008**, *190*, 2981–2986.
68. D. Frey; O. Braun; C. Briand; M. Vasak; M. G. Grutter, *Structure* **2006**, *14*, 901–911.
69. J. Murray-Rust; J. Leiper; M. McAlister; J. Phelan; S. Tilley; J. Santa Maria; P. Vallance; N. McDonald, *Nat. Struct. Biol.* **2001**, *8*, 679–683.
70. T. W. Linsky; A. F. Monzingo; E. M. Stone; J. D. Robertus; W. Fast, *Chem. Biol.* **2008**, *15*, 467–475.
71. M. J. Plevin; B. S. Magalhaes; R. Harris; A. Sankar; S. J. Perkins; P. C. Driscoll, *J. Mol. Biol.* **2004**, *341*, 171–184.
72. S. M. Fundel; D. L. Pountney; P. M. Gehrig; R. Bogumil; P. Faller; M. Vasak, *FEBS Lett.* **1996**, *395*, 33–38.
73. P. R. Thompson; W. Fast, *ACS Chem. Biol.* **2006**, *1*, 433–441.
74. E. M. Stone; A. L. Costello; D. L. Tierney; W. Fast, *Biochemistry* **2006**, *45*, 5618–5630.
75. E. M. Stone; M. D. Person; N. J. Costello; W. Fast, *Biochemistry* **2005**, *44*, 7069–7078.
76. J. Leiper; J. Murray-Rust; N. McDonald; P. Vallance, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 13527–13532.
77. M. Knipp; O. Braun; P. M. Gehrig; R. Sack; M. Vasak, *J. Biol. Chem.* **2003**, *278*, 3410–3416.
78. L. Hong; W. Fast, *J. Biol. Chem.* **2007**, *282*, 34684–34692.
79. S. P. Forbes; L. J. Druhan; J. E. Guzman; N. Parinandi; L. Zhang; K. B. Green-Church; A. J. Cardounel, *Biochemistry* **2008**, *47*, 1819–1826.
80. M. B. Hampton; I. Stamenkovic; C. C. Winterbourn, *FEBS Lett.* **2002**, *517*, 229–232.
81. J. Sohn; J. Rudolph, *Biochemistry* **2003**, *42*, 10060–10070.
82. W. C. Barrett; J. P. DeGnore; S. Konig; H. M. Fales; Y. F. Keng; Z. Y. Zhang; M. B. Yim; P. B. Chock, *Biochemistry* **1999**, *38*, 6699–6705.
83. C. C. Winterbourn; D. Metodiewa, *Free Radic. Biol. Med.* **1999**, *27*, 322–328.
84. D. A. Mitchell; M. A. Marletta, *Nat. Chem. Biol.* **2005**, *1*, 154–158.
85. M. Xian; K. Wang; X. Chen; Y. Hou; A. McGill; B. Zhou; Z. Y. Zhang; J. P. Cheng; P. G. Wang, *Biochem. Biophys. Res. Commun.* **2000**, *268*, 310–314.
86. O. Braun; M. Knipp; S. Chesnov; M. Vasak, *Protein Sci.* **2007**, *16*, 1522–1534.
87. M. Knipp; O. Braun; M. Vasak, *J. Am. Chem. Soc.* **2005**, *127*, 2372–2373.
88. M. Knipp; J. M. Charnock; C. D. Garner; M. Vasak, *J. Biol. Chem.* **2001**, *276*, 40449–40456.
89. P. Vallance; H. D. Bush; B. J. Mok; R. Hurtado-Guerrero; H. Gill; S. Rossiter; J. D. Wilden; S. Caddick, *Chem. Commun. (Camb.)* **2005**, 5563–5565.
90. B. Hartzoulakis; S. Rossiter; H. Gill; B. O'Hara; E. Steinke; P. J. Gane; R. Hurtado-Guerrero; J. M. Leiper; P. Vallance; J. M. Rust; D. L. Selwood, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3953–3956.
91. E. M. Stone; T. H. Schaller; H. Bianchi; M. D. Person; W. Fast, *Biochemistry* **2005**, *44*, 13744–13752.
92. K. Sydow; C. E. Mondon; J. Schrader; H. Konishi; J. P. Cooke, *Arterioscler. Thromb. Vasc. Biol.* **2008**, *28*, 692–697.

93. S. Ueda; S. I. Yamagishi; Y. Matsumoto; Y. Kaida; A. Fujimi-Hayashida; K. Koike; H. Tanaka; K. Fukami; S. Okuda, *Life Sci.* **2009**, *84*, 853–856.
94. R. Shibata; S. Ueda; S. Yamagishi; Y. Kaida; Y. Matsumoto; K. Fukami; A. Hayashida; H. Matsuoka; S. Kato; M. Kimoto; S. Okuda, *Nephrol. Dial. Transplant.* **2009**, *24*, 1162–1169.
95. D. Zakrzewicz; O. Eickelberg, *BMC Pulm. Med.* **2009**, *9*, 5.
96. W. T. McGraw; J. Potempa; D. Farley; J. Travis, *Infect. Immun.* **1999**, *67*, 3248–3256.
97. B. Gyorgy; E. Toth; E. Tarcsa; A. Falus; E. I. Buzas, *Int. J. Biochem. Cell Biol.* **2006**, *38*, 1662–1677.
98. E. R. Vossenaar; A. J. Zendman; W. J. van Venrooij; G. J. Pruijn, *Bioessays* **2003**, *25*, 1106–1118.
99. H. Takahara; Y. Oikawa; K. Sugawara, *J. Biochem.* **1983**, *94*, 1945–1953.
100. K. Nomura, *Arch. Biochem. Biophys.* **1992**, *293*, 362–369.
101. K. Nakashima; T. Hagiwara; A. Ishigami; S. Nagata; H. Asaga; M. Kuramoto; T. Senshu; M. Yamada, *J. Biol. Chem.* **1999**, *274*, 27786–27792.
102. E. Tarcsa; L. N. Marekov; G. Mei; G. Melino; S. C. Lee; P. M. Steinert, *J. Biol. Chem.* **1996**, *271*, 30709–30716.
103. P. Proost; T. Loos; A. Mortier; E. Schutyser; M. Gouwy; S. Noppen; C. Dillen; I. Ronsse; R. Conings; S. Struyf; G. Opdenakker; P. C. Maudgal; J. Van Damme, *J. Exp. Med.* **2008**, *205*, 2085–2097.
104. K. Kizawa; H. Takahara; H. Troxler; P. Kleinert; U. Mochida; C. W. Heizmann, *J. Biol. Chem.* **2008**, *283*, 5004–5013.
105. C. Foulquier; M. Sebbag; C. Clavel; S. Chapuy-Regaud; R. Al Badine; M. C. Mechin; C. Vincent; R. Nachat; M. Yamada; H. Takahara; M. Simon; M. Guerrin; G. Serre, *Arthritis Rheum.* **2007**, *56*, 3541–3553.
106. T. Loos; A. Mortier; M. Gouwy; I. Ronsse; W. Put; J. P. Lenaerts; J. Van Damme; P. Proost, *Blood* **2008**, *112*, 2648–2656.
107. M. A. Moscarello; F. G. Mastronardi; D. D. Wood, *Neurochem. Res.* **2007**, *32*, 251–256.
108. A. Ishida-Yamamoto; T. Senshu; H. Takahashi; K. Akiyama; K. Nomura; H. Iizuka, *J. Invest. Dermatol.* **2000**, *114*, 701–705.
109. K. Nakashima; T. Hagiwara; M. Yamada, *J. Biol. Chem.* **2002**, *277*, 49562–49568.
110. H. Yao; P. Li; B. J. Venters; S. Zheng; P. R. Thompson; B. F. Pugh; Y. Wang, *J. Biol. Chem.* **2008**, *283*, 20060–20068.
111. P. Li; H. Yao; Z. Zhang; M. Li; Y. Luo; P. R. Thompson; D. S. Gilmour; Y. Wang, *Mol. Cell. Biol.* **2008**, *28*, 4745–4758.
112. Y. Wang; J. Wysocka; J. Sayegh; Y. H. Lee; J. R. Perlin; L. Leonelli; L. S. Sonbuchner; C. H. McDonald; R. G. Cook; Y. Dou; R. G. Roeder; S. Clarke; M. R. Stallcup; C. D. Allis; S. A. Coonrod, *Science* **2004**, *306*, 279–283.
113. X. Chang; J. Han, *Mol. Carcinog.* **2006**, *45*, 183–196.
114. G. Y. Liu; Y. F. Liao; W. H. Chang; C. C. Liu; M. C. Hsieh; P. C. Hsu; G. J. Tsay; H. C. Hung, *Apoptosis* **2006**, *11*, 183–196.
115. Y. Luo; B. Knuckley; M. Bhatia; P. J. Pellechia; P. R. Thompson, *J. Am. Chem. Soc.* **2006**, *128*, 14468–14469.
116. K. Arita; H. Hashimoto; T. Shimizu; K. Nakashima; M. Yamada; M. Sato, *Nat. Struct. Mol. Biol.* **2004**, *11*, 777–783.
117. Y. Luo; K. Arita; M. Bhatia; B. Knuckley; Y. H. Lee; M. R. Stallcup; M. Sato; P. R. Thompson, *Biochemistry* **2006**, *45*, 11727–11736.
118. K. Arita; T. Shimizu; H. Hashimoto; Y. Hidaka; M. Yamada; M. Sato, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5291–5296.
119. B. Knuckley; M. Bhatia; P. R. Thompson, *Biochemistry* **2007**, *46*, 6578–6587.
120. P. L. Kearney; M. Bhatia; N. G. Jones; L. Yuan; M. C. Glascock; K. L. Catchings; M. Yamada; P. R. Thompson, *Biochemistry* **2005**, *44*, 10570–10582.
121. M. Leopoldini; T. Marino; M. Toscano, *Theor. Chem. Acc.* **2008**, *120*, 459–466.
122. M. Nakayama-Hamada; A. Suzuki; K. Kubota; T. Takazawa; M. Ohsaka; R. Kawaida; M. Ono; A. Kasuya; H. Furukawa; R. Yamada; K. Yamamoto, *Biochem. Biophys. Res. Commun.* **2005**, *327*, 192–200.
123. R. Raijmakers; A. J. Zendman; W. V. Egberts; E. R. Vossenaar; J. Raats; C. Soede-Huijbregts; F. P. Rutjes; P. A. van Veelen; J. W. Drijfhout; G. J. Pruijn, *J. Mol. Biol.* **2007**, *367*, 1118–1129.
124. Y. Luo; B. Knuckley; Y. H. Lee; M. R. Stallcup; P. R. Thompson, *J. Am. Chem. Soc.* **2006**, *128*, 1092–1093.
125. B. Knuckley; Y. Luo; P. R. Thompson, *Bioorg. Med. Chem.* **2008**, *16*, 739–745.
126. T. Janowitz; H. Kneifel; M. Piotrowski, *FEBS Lett.* **2003**, *544*, 258–261.
127. S. Baumann; A. Sander; J. R. Gurnon; G. M. Yanai-Balser; J. L. Van Etten; M. Piotrowski, *Virology* **2007**, *360*, 209–217.
128. J. P. Simon; V. Stalon, *J. Bacteriol.* **1982**, *152*, 676–681.
129. J. L. Llacer; L. M. Polo; S. Tavarez; B. Alarcon; R. Hilario; V. Rubio, *J. Bacteriol.* **2007**, *189*, 1254–1265.
130. Y. Nakada; Y. Itoh, *Microbiology* **2003**, *149*, 707–714.
131. H. Yanagisawa, *Phytochemistry* **2001**, *56*, 643–647.
132. K. H. Park; Y. D. Cho, *Biochem. Biophys. Res. Commun.* **1991**, *174*, 32–36.
133. G. E. Wesenberg; D. W. Smith; G. N. Philips, Jr.; C. A. Bingman; S. T. M. Allard, Protein Data Bank accession code 1vkp, 2004, to be published.
134. Y. Sakakibara; H. Yanagisawa, *Protein Expr. Purif.* **2003**, *30*, 88–93.
135. B. L. Schneider; A. K. Kiupakis; L. J. Reitzer, *J. Bacteriol.* **1998**, *180*, 4278–4286.
136. H. Shirai; K. Mizuguchi, *FEBS Lett.* **2003**, *555*, 505–510.
137. A. Tocilj; J. D. Schrag; Y. Li; B. L. Schneider; L. Reitzer; A. Matte; M. Cygler, *J. Biol. Chem.* **2005**, *280*, 15800–15808.
138. A. Jann; V. Stalon; C. V. Wauven; T. Leisinger; D. Haas, *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 4937–4941.
139. C. Vander Wauven; V. Stalon, *J. Bacteriol.* **1985**, *164*, 882–886.
140. U. Markisch; G. Reuter, *J. Basic Microbiol.* **1990**, *30*, 425–433.
141. G. Shalev-Alon; A. Sukenik; O. Livnah; R. Schwarz; A. Kaplan, *FEMS Microbiol. Lett.* **2002**, *209*, 87–91.
142. G. Hernandez-Guzman; A. Alvarez-Morales, *Mol. Plant Microbe Interact.* **2001**, *14*, 545–554.
143. T. K. Mihali; R. Kellmann; J. Muenchhoff; K. D. Barrow; B. A. Neilan, *Appl. Environ. Microbiol.* **2008**, *74*, 716–722.
144. J. B. Walker, *Methods Enzymol.* **1975**, *43*, 451–458.
145. L. Costa Pla, *Biochim. Biophys. Acta* **1971**, *242*, 541–548.
146. E. Fritsche; A. Bergner; A. Humm; W. Piepersberg; R. Huber, *Biochemistry* **1998**, *37*, 17664–17672.
147. J. B. Walker, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1979**, *50*, 177–242.
148. A. Humm; E. Fritsche; S. Steinbacher, *Biol. Chem.* **1997**, *378*, 193–197.
149. C. B. Item; S. Stockler-Ipsiroglu; C. Stromberger; A. Muhl; M. G. Alessandri; M. C. Bianchi; M. Tosetti; F. Fornai; G. Cioni, *Am. J. Hum. Genet.* **2001**, *69*, 1127–1133.

150. A. Humm; E. Fritsche; K. Mann; M. Gohl; R. Huber, *Biochem. J.* **1997**, *322* (Pt. 3), 771–776.
151. A. Humm; E. Fritsche; S. Steinbacher; R. Huber, *EMBO J.* **1997**, *16*, 3373–3385.
152. E. Fritsche; A. Humm; R. Huber, *Eur. J. Biochem.* **1997**, *247*, 483–490.
153. E. Fritsche; A. Humm; R. Huber, *J. Biol. Chem.* **1999**, *274*, 3026–3032.
154. E. Grazi; N. Rossi, *J. Biol. Chem.* **1968**, *243*, 538–542.
155. J. B. Walker, *J. Biol. Chem.* **1956**, *218*, 549–556.
156. J. B. Walker, *J. Biol. Chem.* **1956**, *221*, 771–776.
157. S. Ratner; O. Rochovansky, *Arch. Biochem. Biophys.* **1956**, *63*, 277–295.
158. L. L. Sandell; X. J. Guan; R. Ingram; S. M. Tilghman, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4622–4627.
159. I. Sipila, *Biochim. Biophys. Acta* **1980**, *613*, 79–84.
160. S. Ratner; O. Rochovansky, *Arch. Biochem. Biophys.* **1956**, *63*, 296–315.
161. Y. Watanabe; J. F. Van Pilsum; I. Yokoi; A. Mori, *Life Sci.* **1994**, *55*, 351–358.
162. J. A. Gerlt; P. C. Babbitt, *Annu. Rev. Biochem.* **2001**, *70*, 209–246.

Biographical Sketches
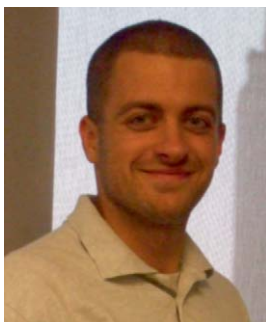
Thomas W. Linsky received his B.A. and M.S. degrees in 2003 and 2004, respectively, from Case Western Reserve University in Cleveland, OH, where he studied computer science and communication sciences. He worked at NASA Glenn Research Center designing communication network systems for future exploration missions to the moon and Mars. However, his interests soon turned to the network of systems that govern biological and chemical processes, and he is currently a Ph.D. candidate in biochemistry working with Professor Walter Fast at the University of Texas at Austin. He has published work on mechanistic partitioning in the pentein superfamily, and his research interests include enzyme mechanisms, enzyme evolution, discovery, and design of lead compounds for enzyme inhibition, and fusion of computational modeling with experimentally derived information.

Walter Fast received his B.S. degree in chemistry from Wheaton College, Wheaton, IL, where he is shown with his daughter Lillian. After a brief but enlightening period of graduate studies in biochemistry at Brandeis University, Waltham, MA, with Professor Lizbeth

Hedstrom, he completed his Ph.D. degree at Northwestern University, Evanston, IL, in 1998 with Professor Richard B. Silverman. At Northwestern, he studied mechanism-based inactivators of NO synthase. He then worked as a postdoctoral fellow with Professor Stephen J. Benkovic at Penn State University, University Park, PA, where he studied the kinetics and mechanism of a dinuclear zinc beta-lactamase involved in the evolution of antibiotic resistance. He was appointed as an assistant professor in the Division of Medicinal Chemistry, College of Pharmacy at the University of Texas at Austin in 2002 and was promoted with tenure to associate professor in 2008. His research interests include mechanism and inhibition of enzymes in the pentein superfamily as well as enzymatic methods of disrupting interbacterial communication pathways.

8.06   Tunnels and Intermediates in the Glutamine-Dependent Amidotransferases

Nigel G. J. Richards, Robert N. Humkey, Kai Li, and Megan E. Meyer, University of Florida, Gainesville, FL, USA

Tania C. Córdova de Sintjago, Universidad Central de Venezuela, Caracas, Venezuela

8.06.1 Introduction

Controlling the utilization of potentially toxic small molecules, such as carbon monoxide,[1] indole,[2] and formaldehyde,[3] or reactive intermediates, such as 5′-phosphoribosylamine (PRA),[4] is an important problem in cellular metabolism.[5–7] In the case of amidotransferases,[8,9] which transfer nitrogen from the side chain of L-glutamine to various electrophilic acceptors, structural studies over the past decade have revealed tunnels that connect two, or more, active sites within the enzyme through which ammonia is thought to be transported during the catalytic cycle.[10–13] Although the existence of such tunnels prevents, in principle, the release of ammonia into the cellular milieu, many questions remain to be addressed concerning the coordination of catalytic activity in the different sites, the 'driving force' for ammonia transfer, and the molecular events by which these tunnels were formed during evolution. In this chapter, we seek to present a comprehensive overview of structure-based knowledge concerning the translocation of ammonia between active sites located in a single amidotransferase,[11] to summarize current thinking about the evolution of molecular tunnels and the efficiency with which ammonia transfer takes place, the molecular mechanisms that coordinate catalysis in multiple active sites, and the nature of the motion through, and selectivity of, the tunnel structures. We will therefore omit discussion of several interesting enzymes, such as glutamine-dependent NAD$^+$ synthetase,[14,15] which seems to possess an unusual glutaminase domain,[16] cobyric acid synthetase,[17–19] and 4-amino-4-deoxychorismate synthase (an interesting target for drug discovery),[20–22] because high-resolution X-ray crystal structures of the complete amidotransferase remain to be reported. In addition, there will be no description of guanosine-5′-monophosphate synthetase (GMPS),[23,24] even though a structure has been determined for the enzyme because it adopts an inactive conformation lacking an ammonia tunnel.[25] Several high-resolution crystal structures have also been reported for anthranilate synthase,[26–28] the amidotransferase catalyzing the initial two reactions in tryptophan biosynthesis (**Scheme 1**),[29] which likely shares a common evolutionary ancestor[30] with 4-amino-4-deoxychorismate synthase[21] and isochorismate synthase.[31,32] Although all these three enzymes play important roles in aromatic amino acid biosynthesis,[33] almost nothing is known about residues defining an ammonia tunnel linking the glutaminase and synthase sites in anthranilate synthase, and the small amount of kinetic information that has been reported provides no information on the efficiency and regulation of ammonia translocation in this enzyme.
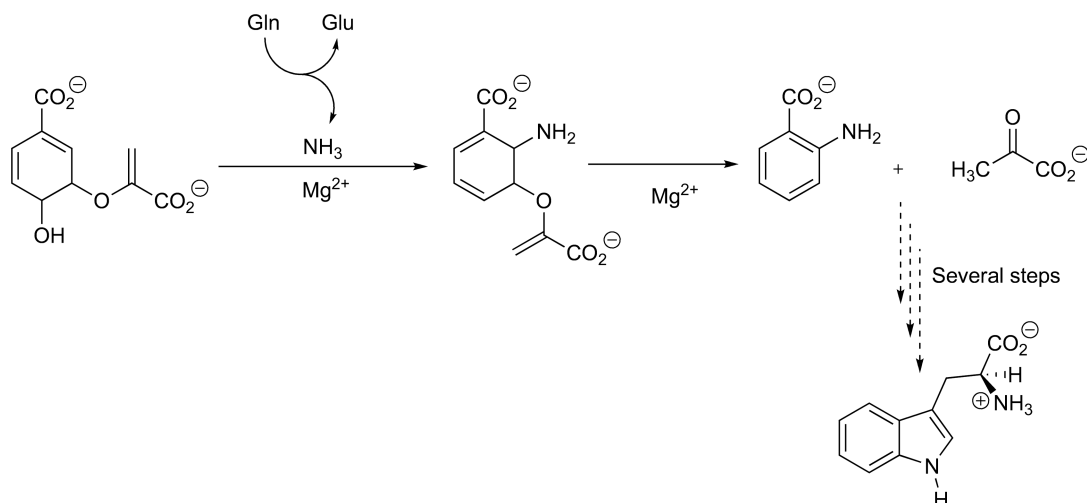
Scheme 1

8.06.2   Glutamine-Dependent Amidotransferases and Nitrogen Metabolism

8.06.2.1   Importance of L-Glutamine in Cellular Metabolism

For many organisms, especially plants, access to nitrogen is a key problem in cellular metabolism.[34,35] The ultimate source of this element is ammonia obtained by reduction of atmospheric nitrogen.[36,37] The cellular concentration of free ammonia must be closely regulated, however, by the action of glutamine synthetase (GS) in eukaryotic cells due to the toxicity of this reactive molecule.[38] An additional chemical problem in using ammonia as a nitrogen source, even in prokaryotes, is protonation at cellular pH to yield an unreactive ammonium cation. The ATP-dependent incorporation of ammonia into the primary amide of glutamine, which is catalyzed by GS,[39–41] not only eliminates unwanted nitrogen reactivity but also prevents protonation of the nitrogen atom. The central importance of glutamine synthesis in metabolism is also illustrated by (1) the large number of mechanisms used in the regulation of GS[38,42] and (2) the biosynthesis of L-methionine sulfoximine, a natural product that is a potent GS inhibitor.[43,44] On the other hand, storing nitrogen in the form of an unreactive amide clearly requires a mechanism for releasing ammonia from its amino acid carrier in a controlled manner and catalyzing its reaction with a suitable electrophile. This is the function of glutamine-dependent amidotransferases, which play key roles in the biosynthesis of many important metabolites, including purines,[45] pyrimidines,[46] amino acids,[47–49] amino sugars,[50] cofactors,[14,51,52] folate,[53] and correctly charged tRNA^Gln and tRNA^Asn in many classes of microorganisms.[54,55] Indeed, their central role in purine and histidine biosynthesis is consistent with the view that these enzymes arose very early in biochemical evolution.[56,57]

8.06.2.2   Classification of Glutamine-Dependent Amidotransferases

Glutamine-dependent amidotransferases possess a modular structure in which one domain catalyzes the hydrolysis of glutamine to glutamate and ammonia, and the other is responsible for mediating the reaction of ammonia with a suitable electrophilic intermediate. To date, three glutaminase 'modules' have been identified, each of which is found in multiple enzymes thereby forming the basis for their classification.[58] Class I amidotransferases, which have also been termed Triad or G-type enzymes,[8] possess a triad of conserved cysteine, histidine, and glutamic acid residues reminiscent of that observed in a number of thiol proteinases, such as papain and the caspases.[59,60] Enzymes in this family play key roles in primary metabolism, mediating several steps in purine biosynthesis, the *de novo* construction of histidine, and the creation of folate and tryptophan precursors (**Scheme 2**). All Class II amidotransferases, which have also been classified as Ntn or F-type enzymes,[58] possess an N-terminal cysteine that seems to be the only residue essential for catalytic activity.[61] Members of this second amidotransferase family play critical roles in the biosynthesis of amino acids
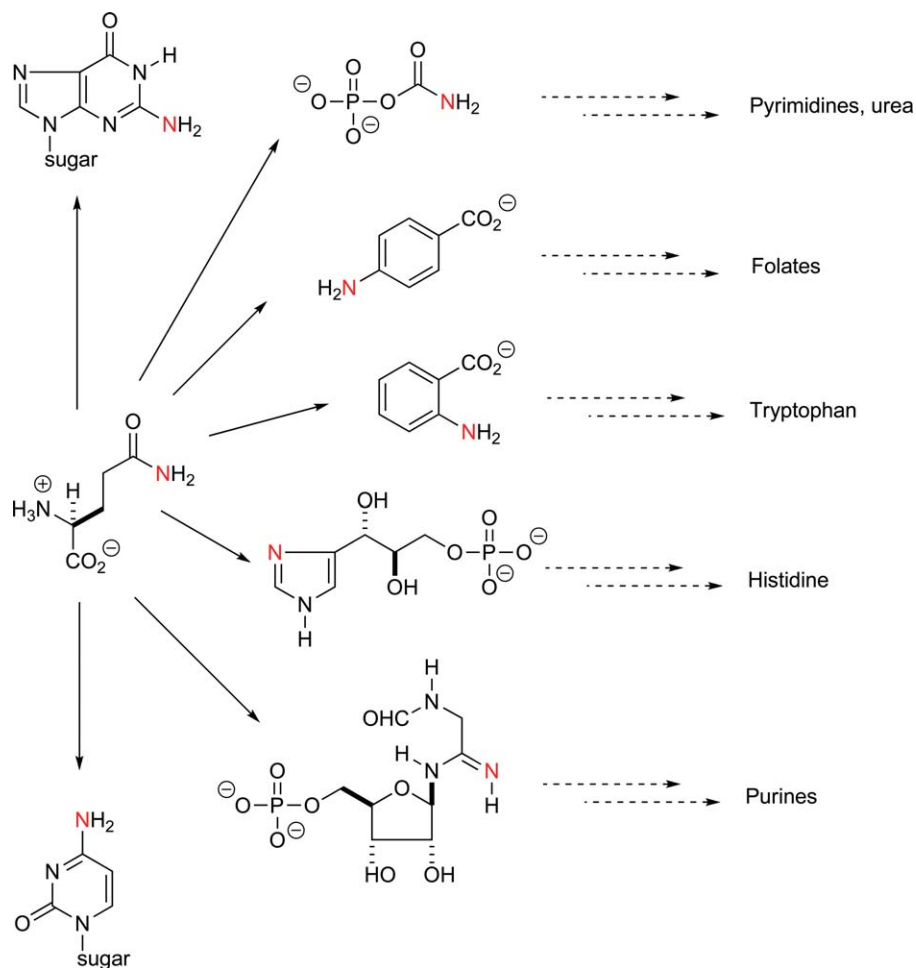
Scheme 2

and cofactors, and may also mediate nitrogen transfer steps in the construction of a number of secondary metabolites (**Scheme 3**). Finally, the Class III amidotransferases exhibit a conserved 'amidase' motif in their glutaminase domain. Although few members of this enzyme family have yet been identified, Glu-tRNA(Gln) amidotransferase plays a critical role in the synthesis of Gln-tRNA(Gln) in many prokaryotes,[55] and is therefore a potential target for drug discovery (**Scheme 4**).[62]

8.06.2.3   Overview of Chemical Mechanism in Glutamine-Dependent Enzymes

Given that all amidotransferases have one of the three types of glutaminase domain, differences in the overall reactions that are catalyzed by these enzymes result from the second 'synthase' or 'synthetase' domain, which contains the machinery necessary to catalyze the nucleophilic addition of ammonia to an electrophilic acceptor thereby forming a new C–N bond (**Scheme 5**). Nitrogen acceptors may already contain reactive functional groups, as in the case of 5′-phosphoribosylpyrophosphate (PRPP) or fructose-6-phosphate, or the domain must activate the substrate for reaction with ammonia. The latter class of substrates includes compounds such as aspartic acid, uracil-5-phosphate, or glutamyl-tRNA(Gln). As discussed below, many of these domains are themselves members of other enzyme superfamilies derived from a common evolutionary ancestor.[63,64] For example, the synthetase domains in asparagine synthetase (ASNS)[65] and GMPS,[25] which both catalyze the
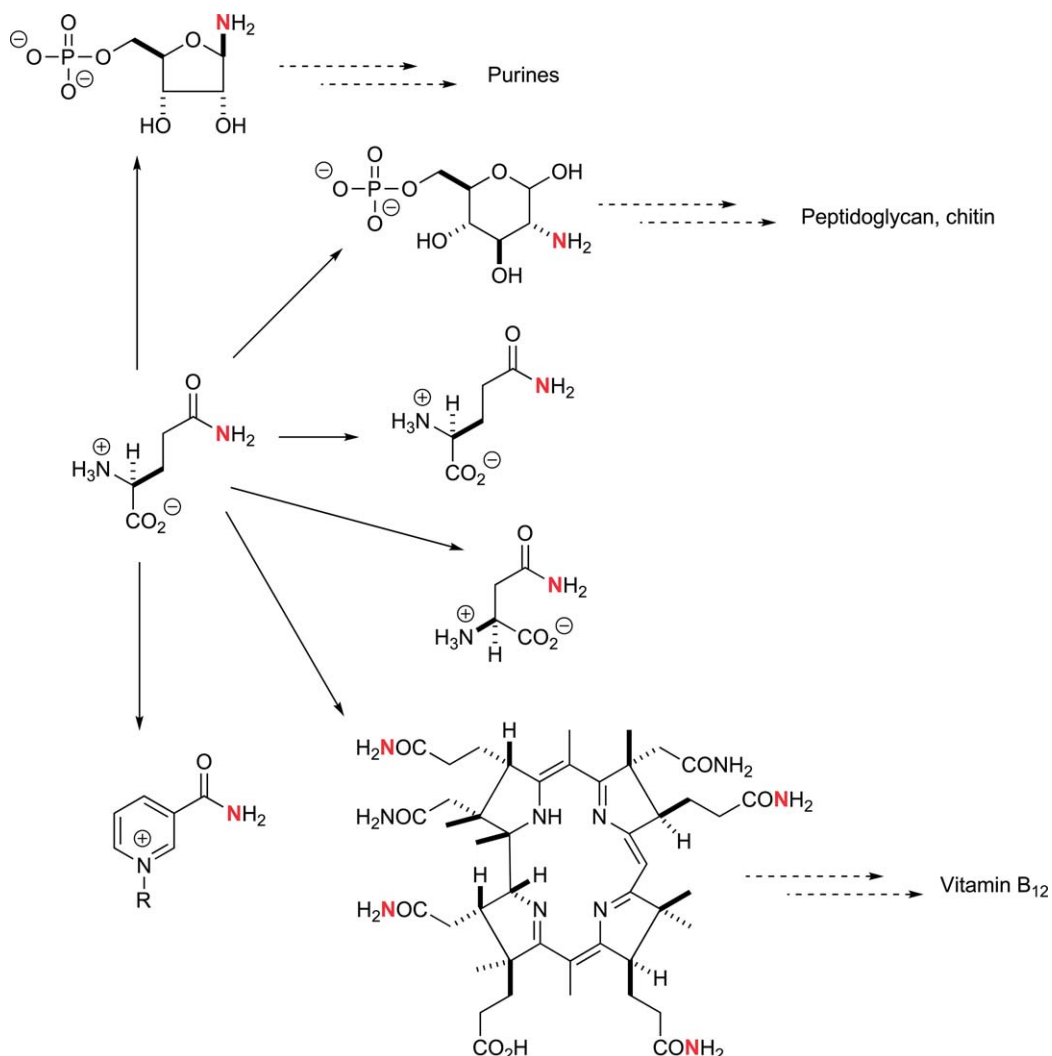
Scheme 3

formation of electrophilic, adenylated intermediates, are homologous with other pyrophosphatases that can be identified by a conserved primary structural element termed the PP-motif.[66]

It is generally agreed that the catalytic mechanism of glutamine hydrolysis in Class I amidotransferases is essentially identical to that used by thiol proteinases, that is, that the amide moiety undergoes nucleophilic attack by the thiolate to yield a tetrahedral intermediate (**Scheme 6**).[67] Protonation of the nitrogen by the active site histidine residue then initiates collapse to a thioester, with C–N bond cleavage releasing ammonia into a tunnel leading away from the active site. Subsequent hydrolysis of the thioester then gives glutamate and regenerates the Cys/His dyad. Evidence for this proposal has been provided by studies of *Escherichia coli* carbamoyl phosphate synthetase (CPS) in which the covalent adduct between the active site cysteine and an aldehyde analog of glutamate was characterized by high-resolution X-ray crystallography.[68] This structure also revealed that the tetrahedral intermediate is stabilized by hydrogen bonding to the backbone amide nitrogens of Gly241 and Leu270 (*E. coli* numbering), which therefore form the equivalent of the 'oxyanion hole' that is seen in both thiol and serine proteinases.[59,69] Evidence for the thioester intermediate has been obtained from kinetic studies on a number of Class I amidotransferases,[53,70–73] and replacement of His353 by asparagine permitted direct observation of the thioester in the glutaminase subunit of *E. coli* CPS.[74] Somewhat unexpectedly, the conserved active site glutamate in Class I glutaminase domains does not seem to be important for hydrolytic
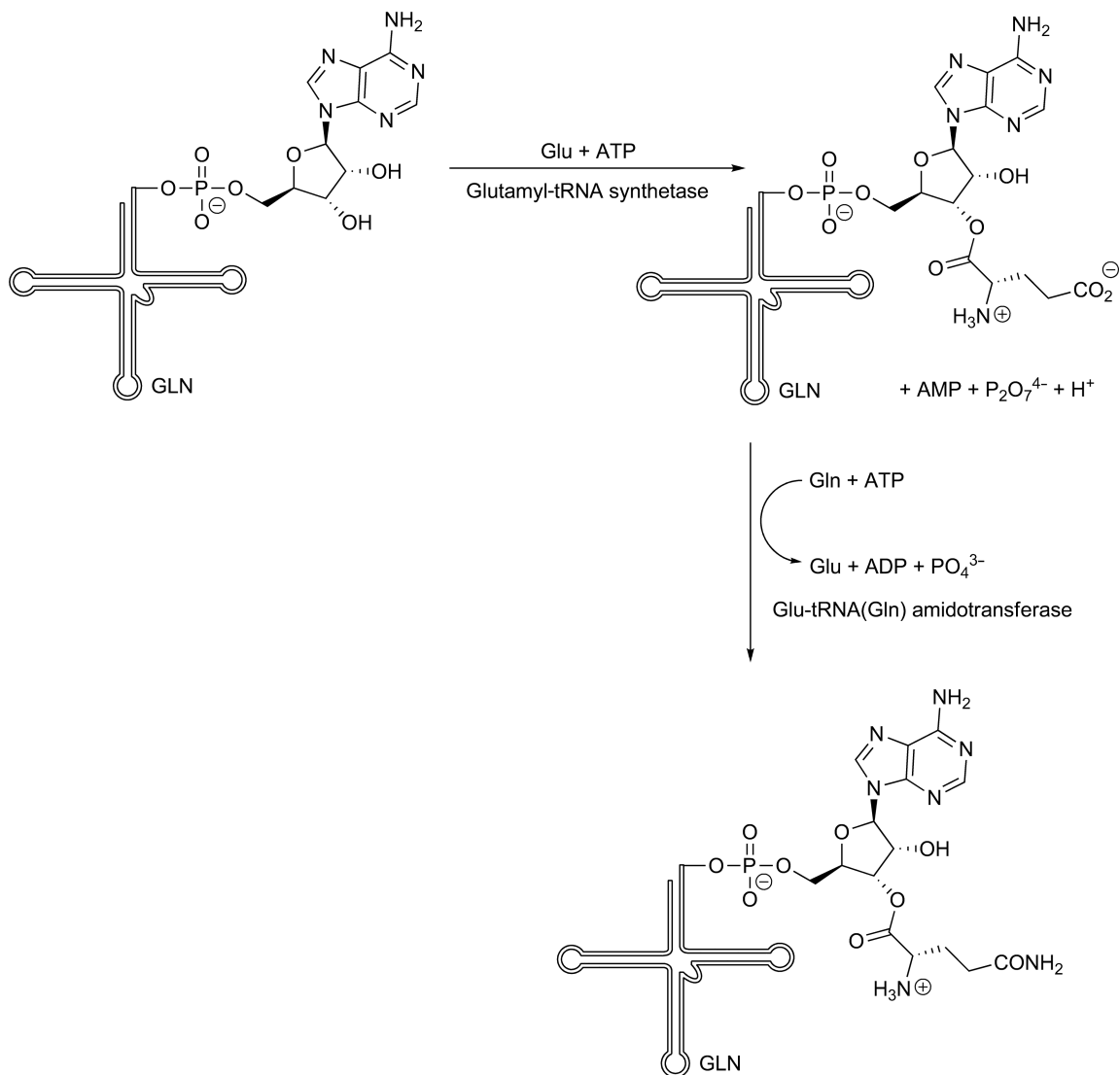
Scheme 4

function,[75] thereby contradicting earlier mechanistic proposals.[8,76] Kinetic experiments have, however, implicated this residue in the molecular mechanisms that couple the partial reactions catalyzed by this class of enzymes.

The chemical mechanism used by the Class II amidotransferases to hydrolyze glutamine has been less well characterized, although the replacement of the conserved N-terminal cysteine (Cys1) in the glutaminase domain by either alanine or serine abolishes catalytic activity.[47,77,78] In addition, structural studies on both glutamine 5′-phosphoribosylpyrophosphate amidotransferase (GPATase) and glutamine fructose-6-phosphate amidotransferase (GFAT) have demonstrated that this residue is covalently modified by the irreversible inhibitor 6-diazo-5-oxonorleucine (DON),[79,80] which is a reactive analog of glutamine.[81] The functional importance of the conserved asparagine residue in stabilizing an oxyanion intermediate, and possibly the thioester intermediate, has also been established.[82] Despite the presence of several highly conserved histidines in the glutaminase active site of the Class II enzymes, none of these residues appear to be important in catalysis.[67,83] As a result, it is believed that the amino group of Cys1 acts as the general acid/base for the hydrolysis reaction (**Scheme 7**).[84] Although this hypothesis is consistent with the reduced pK_a values that have
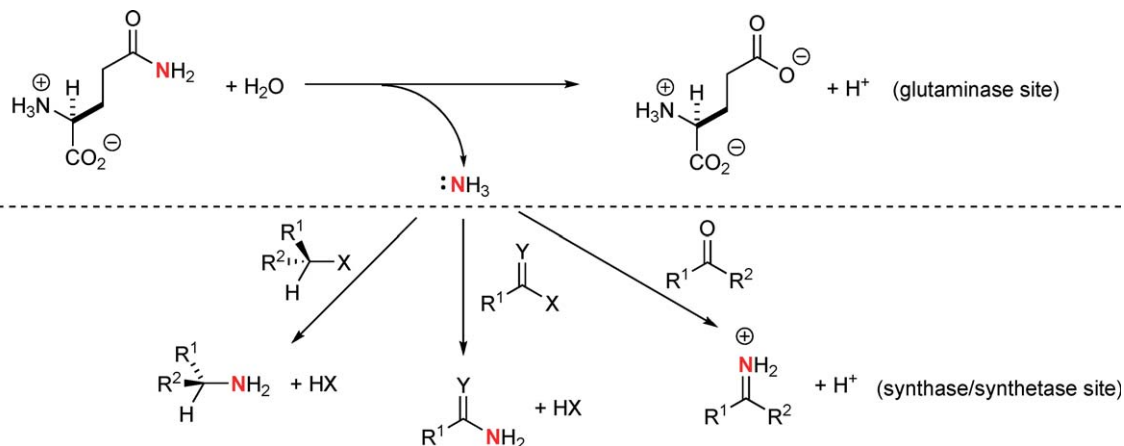
Scheme 5
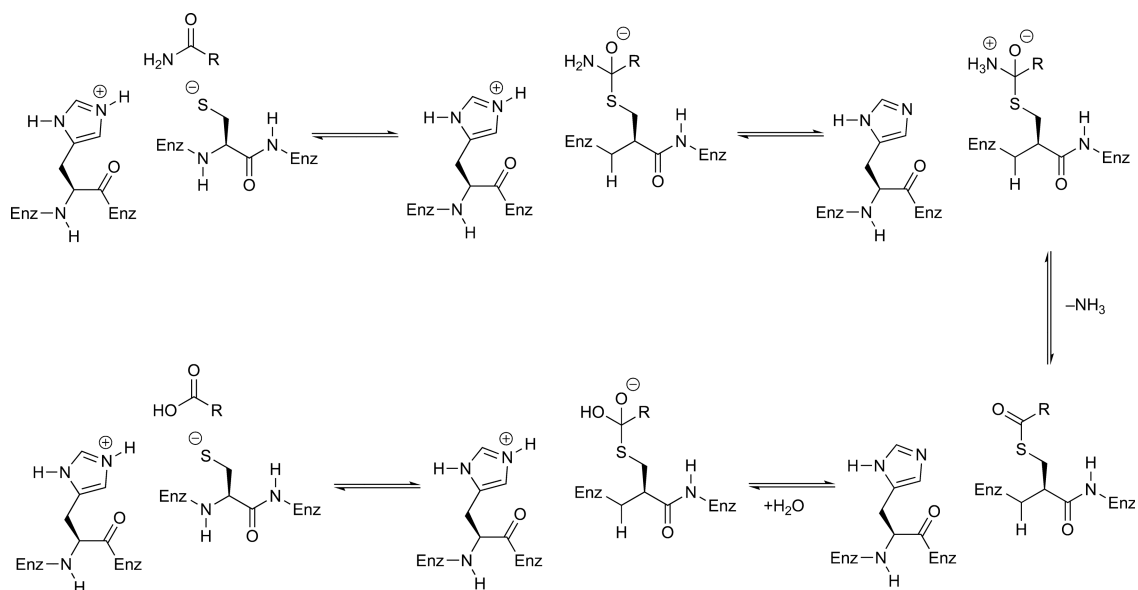
Scheme 6

been observed for the N-terminal amine in 4-oxalocrotonate isomerase,[85] it has proven difficult to obtain unambiguous, direct evidence for the participation of this moiety in catalyzing glutamine hydrolysis. It is interesting that structural comparisons show that the Class II amidotransferase glutaminase domain is evolutionarily related to the family of Ntn amidohydrolases,[86,87] which use N-terminal serine and threonine residues to catalyze peptide bond cleavage.[88–91]

In contrast to both of these amidotransferase families, the Class III glutaminase domain, observed only in glutamine-tRNA^{Gln} amidotransferase to date,[92] contains a conserved serine residue that is essential to catalyzing the conversion of glutamine to glutamate and ammonia.[93] This residue is located within a contiguous segment of amino acids that is a marker for amidases,[94] a class of bacterial enzymes that mediate the industrial conversion of a variety of amides into carboxylic acids.[95] As glutamine and asparagine are not substrates for well-characterized amidases, such as that from *Rhodococcus rhodochrous*,[96] the Class III glutaminase domain may have evolved from amidases that acquired new substrate specificity. Remarkably, structural studies and site-directed mutagenesis experiments on fatty acid amide hydrolase[97] and malonidase E2[98] suggest that amide
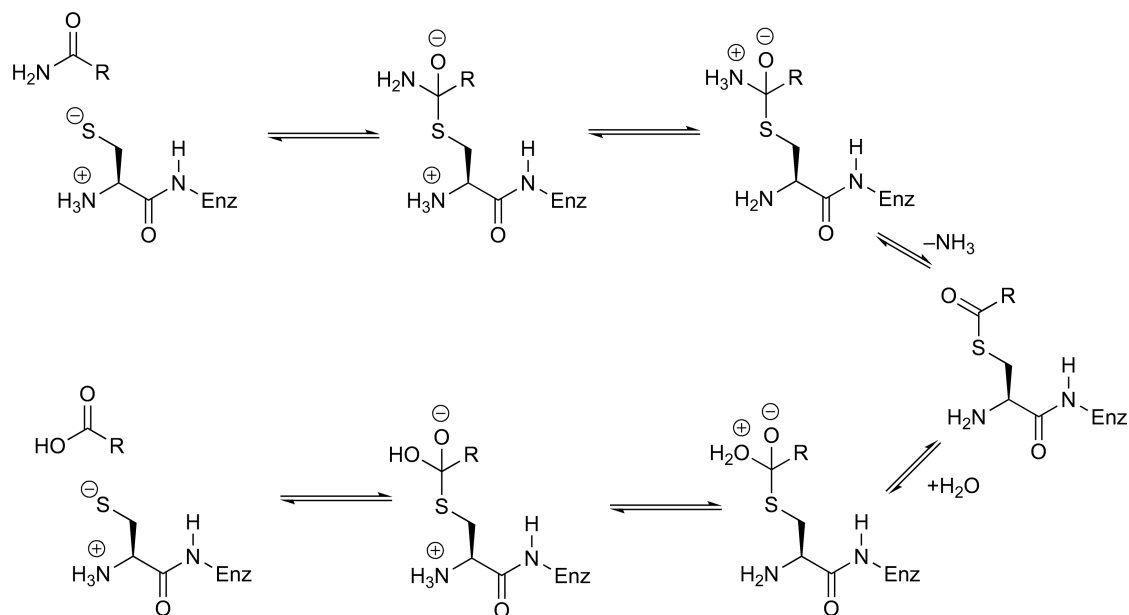
Scheme 7

bond hydrolysis involves activation of the serine nucleophile by a conserved lysine residue in a novel Ser-*cis*-Ser–Lys catalytic triad (**Scheme 8**).[99–101] Similar experiments for the putative catalytic residues in glutamine-tRNA[Gln] amidotransferase remain to be reported.

8.06.3 Class I Amidotransferases

8.06.3.1 Carbamoyl Phosphate Synthetase

8.06.3.1.1 Catalytic mechanism

CPS catalyzes the synthesis of carbamoyl phosphate (CP) from ATP and bicarbonate, using either glutamine or ammonia as a nitrogen source depending on the type of the enzyme.[102] Thus, three isoforms of CPS are present in eukaryotes, which are used to generate CP for use in different metabolic pathways. In contrast, prokaryotes possess a single CPS, classified as Type II, for which glutamine is the sole nitrogen source. The complicated, CPS-catalyzed transformation proceeds via four separate chemical steps that are catalyzed in three spatially distinct active sites.[103,104] *Escherichia coli* CPS is the best characterized form of the enzyme, and has proved a useful model system for exploring how intramolecular tunnels linking the active sites mediate the translocation of various reaction intermediates between the active sites.[105–108] The active enzyme is a heterodimer composed of a small subunit, which mediates the release of ammonia from glutamine using the Cys/His dyad that is present in all Class I amidotransferases. The large subunit of the CPS heterodimer is composed of four distinct domains (carboxyphosphate synthetic, oligomerization, CP synthetic, and allosteric), and contains two other catalytic sites together with binding sites for allosteric regulators, such as ornithine and uridine 5′-monophosphate (UMP) (**Figure 1**).[111] Ammonia is therefore generated in the small subunit and travels to the large subunit where it undergoes reaction with carboxyphosphate,[112,113] (itself produced from bicarbonate anion and ATP) to give carbamate as the first unstable intermediate). The carbamate product is then translocated to a third active site where it undergoes phosphorylation to yield the final product (**Scheme 9**). As might be expected from the fact that both active sites in the large subunit catalyze similar chemistry, that is, phosphorylation of a carboxylate moiety, the large subunit is composed of two halves that are almost superimposable, presumably as a result of a gene duplication event.[114] The modular nature of CPS is also evident from the fact that both catalytic domains in the large subunit are homologous to those of other enzymes known to activate carboxylate groups by
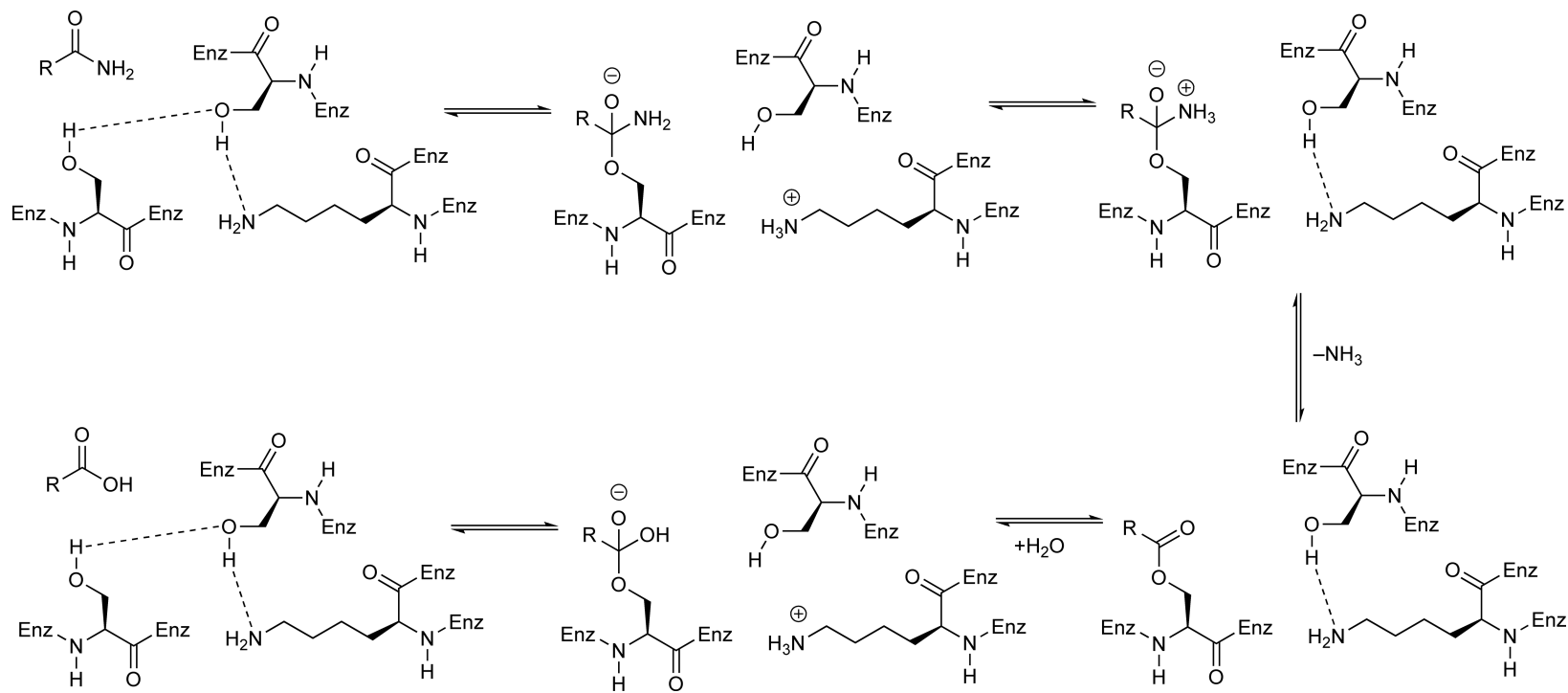
Scheme 8

Figure 1   Cartoon representation of the *Escherichia coli* CPS heterodimer (1JDB) complexed to ADP and ornithine, an allosteric inhibitor of the enzyme. The glutaminase subunit is colored in orange. In the synthetase subunit, the carboxyphosphate-forming, oligomerization, carbamoyl phosphate-forming, and the allosteric domains are rendered in green, yellow, blue, and red, respectively. Bound ADP and ornithine molecules are shown as Corey–Pauling–Koltun (CPK) models: C – gray, O – red, N – blue, and P – orange. Image rendered in PYMOL[109] using coordinate data taken from the Protein Data Bank.[110]

phosphorylation, including biotin carboxylase.[115] Indeed, the functional similarity of cognate residues in CPS and biotin carboxylase that mediate phosphorylation has been demonstrated by site-directed mutagenesis experiments.[116]

8.06.3.1.2   *Enzyme structure*

In addition to providing the first clear insights into active site structure and domain organization, the 2.8 Å resolution X-ray structure of CPS complexed to ADP, inorganic phosphate (P_i), Mn^{2+}, and ornithine revealed a series of intramolecular tunnels connecting the three active sites in the small and large subunits of the enzyme (**Figure 2**).[103] This structure therefore provided the first direct evidence for ammonia transfer between glutaminase and synthetase (or synthase) sites in glutamine-dependent amidotransferases. As discussed in detail elsewhere,[11–13,67] translocating ammonia through an intramolecular tunnel permits control of protonation state and precludes the futile conversion of glutamine to glutamate by preventing diffusion of this reactive
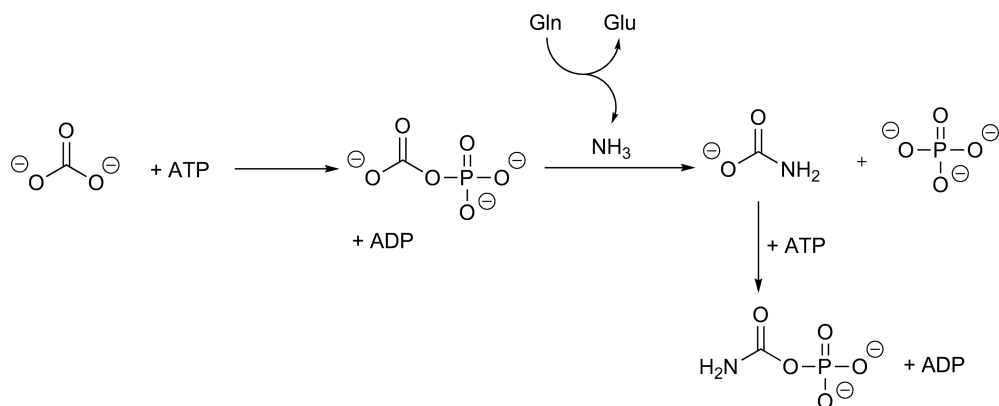
Scheme 9

Figure 2 Cartoon representation of the *Escherichia coli* CPS heterodimer (1JDB) showing residues (gray dotted spheres) that define the tunnels linking the three active sites in the enzyme. Subunits, domains, and ligands are colored as described in **Figure 1**. Image rendered in PYMOL.
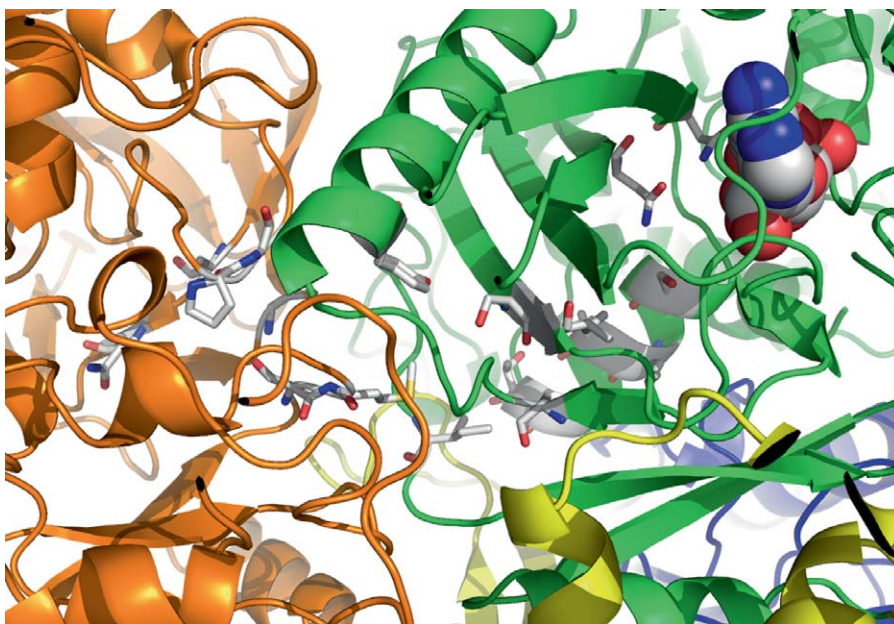
Figure 3   Residues lining the ammonia tunnel of the *Escherichia coli* CPS heterodimer (1JDB). The glutaminase site of the small unit is at the left and the carboxyphosphate-forming domain, containing bound ADP (CPK representation), is at the right. Side chains pointing into the tunnel are rendered as cylinders: C – gray, O – red, and N – blue. Image rendered in PYMOL.

intermediate into solution. The first intramolecular tunnel in CPS, connecting the glutaminase (small subunit) and carboxyphosphate-forming site (large subunit), is predominantly lined with small or nonionizable, polar side chains (Ser35, Met36, Gly293, Ala309, Asn311, Pro358, and Gly359 of the small subunit; Ser233, Ile234, Ala251, Tyr261, Asn283, Asn301, Ser307, Leu310, Ala311, Ala314, Thr315, and Ile352 of the large subunit) (**Figure 3**). This lack of ionizable groups supports the notion that ammonia is transported between the active sites in its neutral, reactive form, which can form hydrogen bonding interactions as it moves through the protein.[117] On the other hand, the average minimum radius of the approximately cylindrical tunnel was observed to be only 3.2 Å, close to the diameter of an ammonia molecule, with a constriction of 2.1 Å just prior to the carboxyphosphate active site. It is therefore likely that dynamical motions in the protein, perhaps associated with the formation of specific reaction intermediates and the presence of ammonia in the tunnel, act to permit translocation between the two sites.

A second intramolecular tunnel linking the carboxyphosphate- and CP-forming sites was also evident in the CPS crystal structure. This tunnel presumably mediates the translocation of carbamate that is unstable and undergoes decarboxylation with a half-life of 70 ms at neutral pH.[118] This intermediate is therefore prevented from leaving the enzyme prior to phosphorylation with ATP, and the 'one-dimensional' nature of its diffusion along the tunnel presumably optimizes the time taken for it to reach the final active site. In addition to residues possessing hydrophobic side chains, the 'carbamate' tunnel is lined with polar, nonionizable groups capable of hydrogen bonding to the charged intermediate (Ile18, Ala23, Met174, Gly175, Met378, Val381, Gly575, Gln829, Asn843, Thr849, Met911, Ser913, and Thr914) and has a larger radius than that of the ammonia tunnel. The entrance to the third active site, which mediates ATP-dependent phosphorylation, is lined with charged side chains (Glu577, Arg848, Lys891, and Glu916) (**Figure 4**).

8.06.3.1.3   *Investigating intramolecular ammonia translocation*

Two alternate models may be envisaged for ammonia transfer between the small and large subunits of CPS (**Scheme 10**).[107] In the first, ammonia released in the glutaminase site might diffuse to the carboxyphosphate-forming active site along the observed tunnel. On the other hand, ammonia might be released into solution and then reenter the large subunit to react with the sequestered carboxyphosphate intermediate. In order to delineate
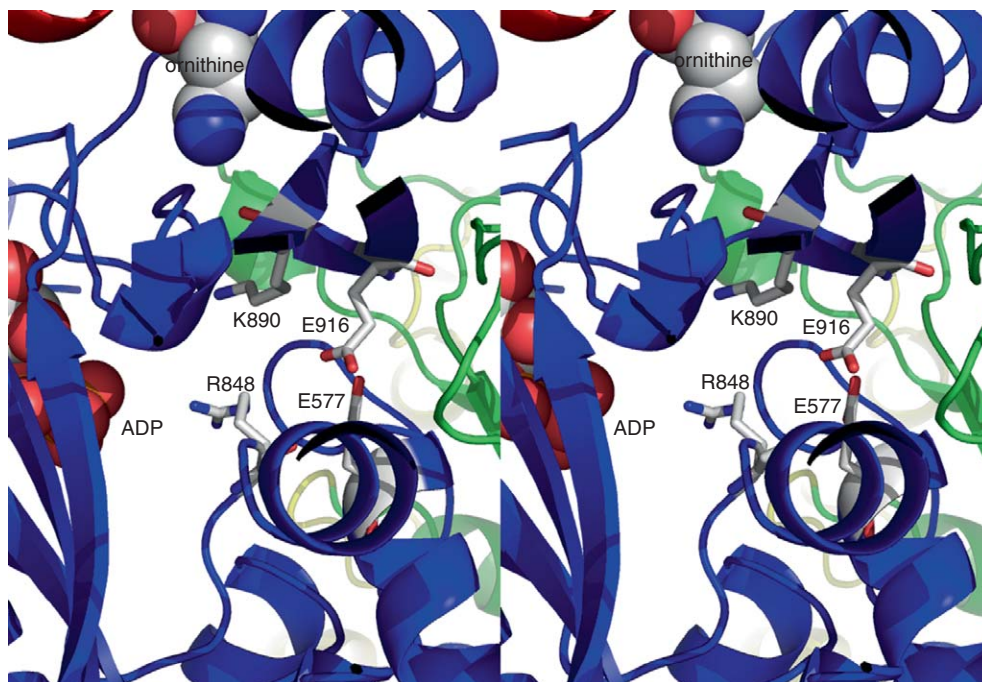
Figure 4   Stereoview of the polar residues (Glu577, Arg848, Lys890, and Glu916) defining a 'gateway' for carbamate ion in the large subunit (blue) of the *Escherichia coli* CPS heterodimer (1JDB). The location of the allosteric site is indicated by the ornithine molecule (CPK representation). The backbone of the glutaminase subunit is also shown (green). Side chains of the 'gateway' residues are rendered as cylinders: C – gray, O – red, and N – blue. Image rendered in PYMOL.

which model was operative for CPS, a competition experiment was performed in which ^{15}N-labeled ammonia was present in the normal assay mixtures for glutamine-dependent activity,[107] and the product CP assayed for the incorporation of ^{15}N-label using ^{15}N nuclear magnetic resonance (NMR) spectroscopy. The results showed that although CPS could use exogenous ^{15}N-labeled ammonia in place of glutamine, ^{15}N incorporation was essentially completely suppressed in the presence of 'unlabeled' glutamine (**Figure 5**). This simple experiment eliminates the possibility that ^{14}N-ammonia is released from the active form of the heterodimer because mixing with the exogenous ^{15}N-labeled material would lead to ~80% of product molecules containing the heavy atom isotope. Unfortunately, using ^{15}N NMR spectroscopy to measure the extent of isotopic incorporation demands significant amounts of acquisition time, because of the magnetic properties of the ^{15}N nucleus,[119] precluding the routine application of this methodology to large numbers of samples.

Having established the importance of the CPS tunnel in ammonia transfer, efforts were made to explore its molecular properties using site-directed mutagenesis experiments aimed at 'blocking' or 'perforating' the putative tunnel. Early work focused on Gly359 of the small subunit, which is located on the interior wall of the ammonia tunnel, because computer-based modeling suggested that the introduction of a side chain at this position would block the tunnel thereby affecting the rate and/or efficiency of glutamine-dependent CP synthesis.[120,121] The G359Y, G359F, G359S, and G359L CPS mutants, in which the glycine was replaced by tyrosine, phenylalanine, serine, and leucine, respectively, exhibited the kinetic properties anticipated for a 'blocked' ammonia tunnel. For example, in the absence of exogenous ammonia, up to 48 equivalents of glutamine were consumed by these CPS mutant enzymes per CP formed in the overall reaction. This was accompanied by a substantial decrease in the catalytic efficiency (k_{cat}/K_M) for glutamine-dependent CP formation, and a 'lag' between the time of glutamate and CP production was observed.[121] No effect of the mutations was observed, however, on the steady-state kinetic parameters when ammonia was used as the nitrogen source. All of these observations are consistent with those expected if ammonia failed to pass along the intramolecular tunnel into the carboxyphosphate-forming active site. Subsequent structural characterization of the G359F CPS mutant by high-resolution X-ray crystallography, however, revealed that the presence of the large phenylalanine side chain had not created a tunnel blockage but
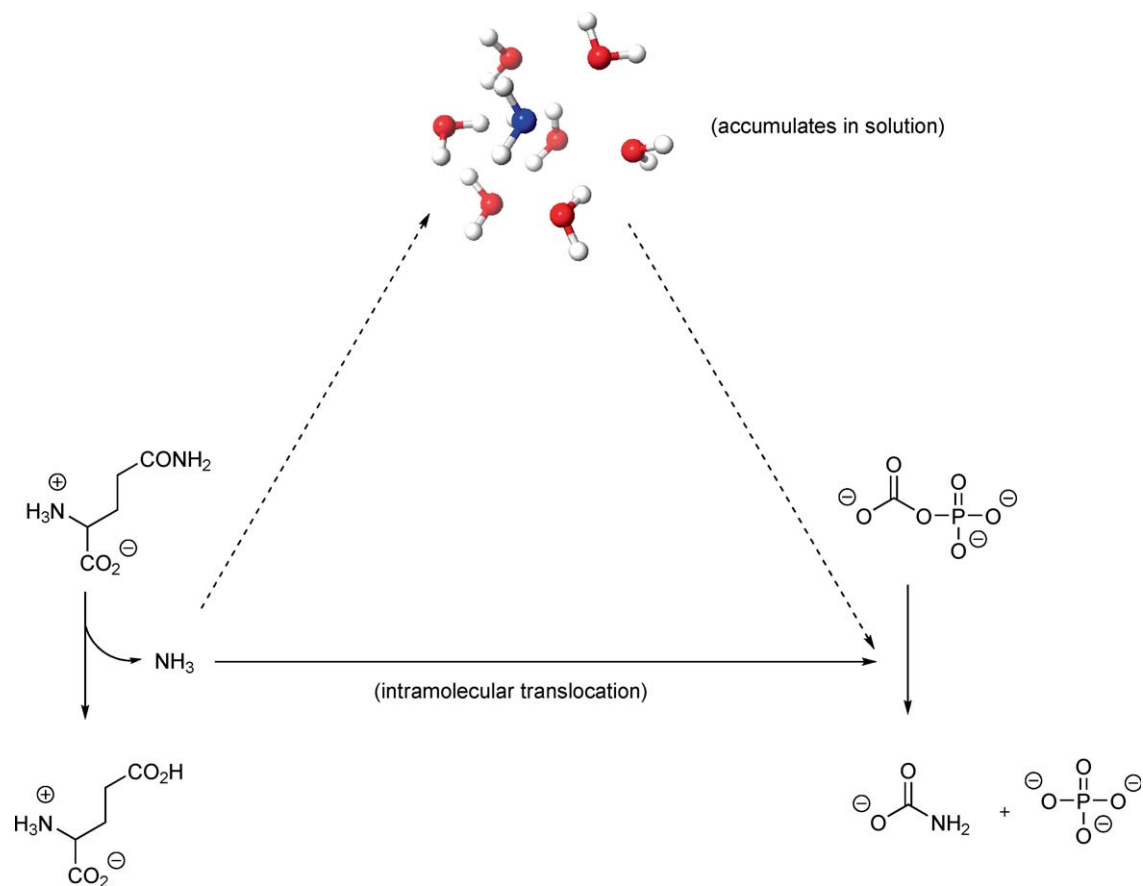
Scheme 10

had caused the conformational rearrangement of a critical loop segment, resulting in a perforation of the tunnel wall (**Figure 6**).[106] The observed kinetic properties of this CPS mutant were therefore caused by the leakage of ammonia from the enzyme during glutamine-dependent catalytic turnover, illustrating the difficulty of determining the structure–function properties of residues within ammonia tunnels using only steady-state kinetics and product stoichiometry measurements.

An alternate strategy in which large side chains were replaced by smaller ones in an effort to create holes in the tunnel, and therefore facilitate ammonia leakage, was also explored using CPS.[105] Hence, residues lining the ammonia tunnel at the interface of the two CPS subunits (αPro360 and αHis361 in the small subunit; βGln262 and βArg265 in the large subunit) were systematically replaced by alanine, and the resulting CPS variants characterized using steady-state kinetics. Perhaps surprisingly, the properties of the βR265A single and βQ262A/βR265A double CPS variants were essentially unchanged relative to those of the wild-type enzyme, illustrating the 'plasticity' of the interface in accommodating small perturbations. Similarly, substituting alanine residues in place of αPro360 and αHis361 gave a CPS variant that exhibited similar kinetic behavior to wild-type enzyme. It was only when three residues at the interface, αPro360, αHis361, and βArg265, were simultaneously replaced by alanine that the rate of glutamine-dependent CP synthesis was reduced by three orders of magnitude. Given that the glutaminase activities of wild-type CPS and the αP360A/αH361A/βR265A variant were similar, and both enzymes could use ammonia as a nitrogen source, it is likely that these changes do lead to tunnel perforation, although this remains to be verified by X-ray crystallography. Even with the loss of several hydrogen bonds, however, the two CPS subunits retained a tight association in the αP360A/αH361A/βR265A variant.

Indirect evidence supporting this idea has been provided by the absence of ^{18}O incorporation into bicarbonate when the CPS-catalyzed reaction is performed in ^{18}O-containing water demonstrates that
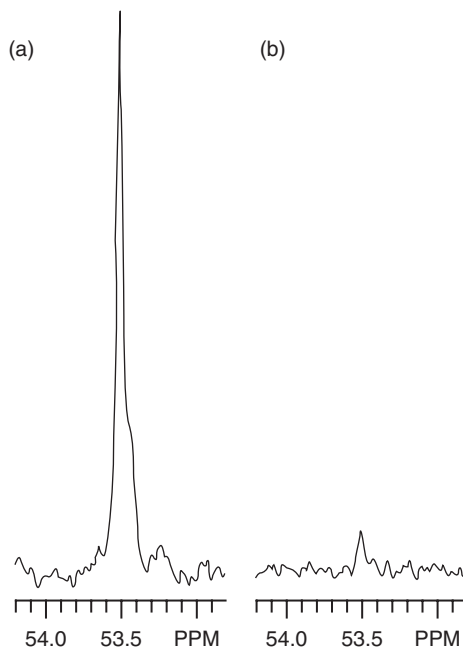
Figure 5 (a) ^{15}N NMR spectrum of ^{15}N-labeled citrulline formed from ^{15}N-carbamoyl phosphate produced in the CPS-catalyzed reaction when [^{15}N]-NH$_4$Cl is used as the sole nitrogen source (pH 7.6). (b) ^{15}N-labeled citrulline produced when unlabeled glutamine (25 mmol l^{-1}) is present as a source of nitrogen in addition to [^{15}N]-NH$_4$Cl (100 mmol l^{-1}) (pH 7.6). The ability of CPS to use ammonia is suppressed by the presence of glutamine, showing that intramolecular ammonia translocation occurs with high efficiency. Reproduced from L. S. Mullins; F. M. Raushel, *J. Am. Chem. Soc.* **1999**, *121*, 3803–3804 with permission from the American Chemical Society.

carbamate is not released into solution during catalytic turnover.[122] This is consistent with a model in which carbamate formed in the carboxyphosphate active site is translocated through the second tunnel to the third active site where it is phosphorylated to yield the final product. Mutagenesis experiments were therefore performed to introduce structural defects into the carbamate tunnel observed in the CPS crystal structure, and the CPS variants characterized by steady-state kinetic measurements.[123] In contrast to the ammonia tunnel, the carbamate tunnel is lined with conserved glutamate (βGlu25, βGlu383, βGlu577, βGlu604, and βGlu916) and arginine (βArg306 and βArg848) side chains. The molecular mechanisms by which these residues might prevent the decomposition of carbamate by preventing its protonation remain poorly defined, although it has been suggested that these residues function as 'gatekeepers' controlling access to the tunnel. Once again, it proved to be remarkably difficult to identify mutations that resulted in significantly altered kinetic parameters for the CPS variant relative to the wild-type enzyme. The replacement of Gly575 by leucine, however, yielded an enzyme for which CP synthesis was substantially decreased even though glutamine-dependent ATPase activity was increased fivefold. This observation is consistent with impaired translocation of carbamate to the third active site due to steric blockage of the tunnel by the larger side chains. Similar results were obtained for the βA23L/βG575L double mutant. Unfortunately, no crystal structures have been reported for these CPS variants, and the interpretation of these steady-state data is complicated by the fact that diffusion of intermediates through the tunnels is a relatively fast process (3×10^4 s^{-1} for carbamate translocation on the basis of a simple one-dimensional diffusion model).[13]

8.06.3.1.4   *Coordinating the catalytic activities of CPS*

As observed for many other Class I glutamine-dependent amidotransferases, the glutaminase activity of CPS only becomes significant in the presence of other substrates, suggesting the existence of molecular interactions that coordinate the reactions taking place in at least two of the three active sites.[124] Such regulation is, of course, in addition to that effected allosterically by binding small molecules, such as UMP and ornithine, into noncatalytic sites elsewhere in the enzyme.[125–128] The presence of ATP and bicarbonate has significant effects on the kinetics of
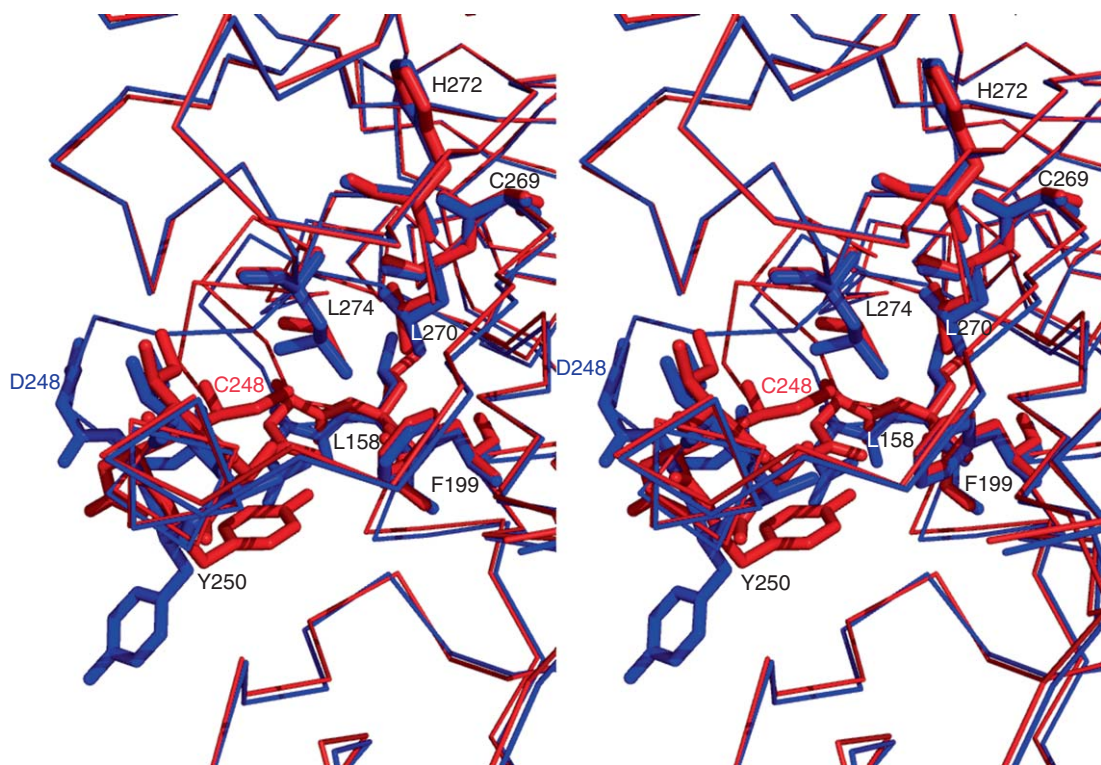
Figure 6 Superimposition of wild-type CPS (red) (1JDB) and the C248D CPS mutant (blue) (1T36) showing the structural changes when Cys248 is replaced by an aspartate residue. Image rendered in PYMOL.
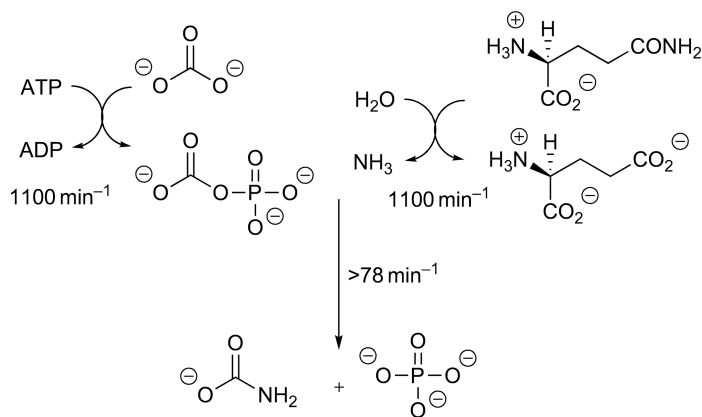
Scheme 11

the glutaminase reaction,[129] presumably as the result of conformational changes that 'couple' the activities of the large and small subunits.[130] A series of pre-steady-state kinetic measurements was therefore undertaken to obtain the intrinsic rates of the reactions catalyzed by each of the three CPS active sites (**Scheme 11**).[108] These experiments showed that the glutaminase- and carboxyphosphate-forming activities proceed with almost identical intrinsic rates when all substrates are present, meaning that ammonia release is strictly coordinated with the synthesis of carboxyphosphate, the electrophilic nitrogen acceptor.[108] An important consequence of this kinetic coupling is that only a single ammonia molecule is likely present in the tunnel during each catalytic turnover of the enzyme. A similar activation of glutaminase activity could be obtained by mutating Cys248 on the small subunit to aspartate. Somewhat unexpectedly, the crystal structure of the αC248D CPS variant showed only small
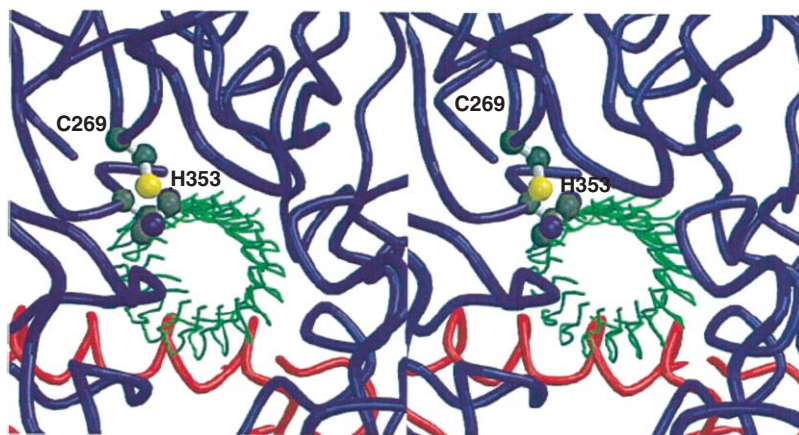
Figure 7 Stereoview showing the escape route for ammonia in the G359F CPS mutant (green 'chicken wire') as calculated from the crystal structure (1M6V) using the ALIMENTARY software package (103). In wild-type CPS, ammonia release is blocked because this region is occupied by the loop composed of residues Gly359 to Asp363, which becomes disordered in the G359F CPS mutant. Cys259 and His353, which catalyze the glutaminase reaction are shown as ball-and-stick representations, and the main chains of the large and small subunits are colored red and blue, respectively. Reproduced from J. B. Thoden; X. Huang; F. M. Raushel; H. M. Holden, *J. Biol. Chem.* **2002**, *277*, 39722–39727, with permission from the American Society for Biochemistry & Molecular Biology.

conformational changes in a loop segment relative to that of the wild-type enzyme (**Figure 7**).[131] The complexity of these structural modifications was further illustrated, however, by the fact that UMP was bound to the αC248D CPS variant in the crystal even though this allosteric inhibitor was not included in either the purification or crystallization buffers. The details of the structural interactions that stimulate glutaminase activity in the CPS small subunit therefore remain to be elucidated. Recent advances in monitoring conformational changes in the enzyme by fluorescence may provide a tool for resolving this problem.[132]

8.06.3.2 Imidazole Glycerol-3-Phosphate Synthase

8.06.3.2.1 Catalytic mechanism

Histidine biosynthesis in *E. coli* proceeds in 10 steps that are catalyzed by eight enzymes, three of which are bifunctional.[133] The protein encoded by the *HisH* gene shares considerable sequence similarity to the glutaminase subunit of CPS, but this enzyme exhibits glutaminase activity only in the presence of a second protein encoded by *HisF*.[134] Biochemical experiments have shown that the HisH/HisF heterodimer (IGP synthase) catalyzes the intriguing transformation of N^1-[(5′-phosphoribulosyl)-formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (PRFAR) into imidazole glycerol-3-phosphate (IGP) and 5′-(5-aminoimidazole-4-carboxamide) ribonucleotide (AICAR) (**Scheme 12**).[135] The latter product is also an intermediate formed during *de novo* purine biosynthesis,[136] and this link between amino acid and nucleotide metabolism has been adduced as evidence for an early origin of histidine in the evolution of life.[137,138] In contrast to bacteria, the IGP synthases present in fungi and plants are single polypeptides in which the glutaminase (HisH) and synthase (HisF) moieties are fused together as N- and C-terminal catalytic domains, respectively.[139] Despite kinetic studies of IGP synthase from a number of sources, many details of the catalytic mechanism remain ambiguous.[135,140,141] It is known, however, that PRFAR binding stimulates glutaminase activity and ammonia is translocated from the glutaminase site through the HisF subunit, or C-terminal synthase domain depending on the source of the enzyme, to the synthase active site where it reacts with a ketone to give an imine. Hydrolysis is then thought to take place giving AICAR and an intermediate that cyclizes to IGP in a thermodynamically favorable reaction (**Scheme 13**). The timing of these molecular events remains to be verified, however, and although it is believed that the rate of the cyclization is accelerated by general acid/base catalysis, the catalytic residues responsible have yet to be identified.
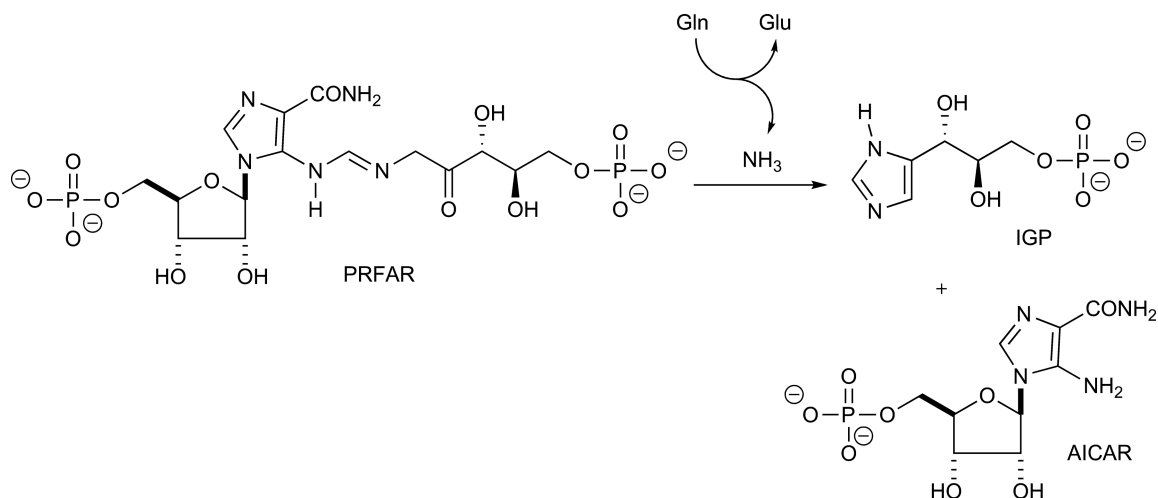
Scheme 12

8.06.3.2.2   Enzyme structure

Several structures have been solved for the IGP synthase HisH/HisF heterodimer and the fused IGP synthase present in *Saccharomyces cerevisiae*, both as the free enzyme and as a complex with various inhibitors and substrate analogs.[139,142–145] In the bacterial form of the enzyme,[143–145] both subunits are $(\beta/\alpha)_8$ barrels, a fold that is adopted by a large number of metabolically important enzymes that catalyze a tremendous variety of transformations.[146,147] The two subunits pack together forming an interface composed of highly conserved residues that is $\sim$1700 $\mathring{A}^2$ in area (**Figure 8**). A similar structure was observed for the single-chain, yeast IGP synthase,[139] and the location of the glutaminase site in the N-terminal domain was unambiguously identified by using acivicin, a reactive analog of glutamine,[148] to covalently modify the thiolate of the catalytically important cysteine residue.[149] Electron density for bound PRFAR in this acivicin-IGP synthase structure could not be observed, however, and the location of the cyclase site was initially assigned from two phosphate ions that were proposed to mimic the phosphates at both ends of this substrate. As a result, the two active sites in the yeast enzyme are separated by a distance of 30 $\mathring{A}$, implying that ammonia released in the glutaminase site must diffuse through the center of the C-terminal $(\beta/\alpha)_8$ barrel to reach the reactive carbonyl group of PRFAR (**Figure 9**). This finding was unexpected because it was anticipated that the two active sites would be located at the same interface of the two catalytic domains. This assignment was later confirmed from the structure of a ternary complex between PRFAR and the acivicin-modified yeast IGP synthase,[142] which showed PRFAR to be bound in an extended conformation within a deep cleft across the face of the HisF $(\beta/\alpha)_8$ barrel causing a series of conformational changes that reorganized the oxyanion hole of the glutaminase site. The interior surface of the proposed ammonia tunnel is best regarded as being defined by four layers of side chains from residues located on the β-strands of the C-terminal $(\beta/\alpha)_8$ barrel (**Figure 10**). A particularly intriguing feature is the presence of a 'gate' consisting of four conserved, charged residues (Arg239, Glu293, Lys360, and Glu465 in the yeast enzyme), which divides the tunnel into two 'chambers' (**Figure 11**). In the yeast IGP synthase,[139,142] the chamber closest to the PRFAR-binding site is composed of small, hydrophobic residues (Ile241, Val400, Ile497, and Ala519 in yeast) rather than the large side chains that are observed in the interior of other $(\beta/\alpha)_8$ barrel proteins, thereby creating a cavity through which ammonia can be translocated. There are also a small number of polar side chains that are capable of hydrogen bonding to neutral ammonia, such as Thr328, Thr295, and Ser362 (yeast numbering). These may also interact with buried water molecules during the catalytic turnover of the enzyme. Similar conclusions were reached based on the crystal structures of the heterodimeric IGP synthases from *Thermotoga maritima*[143,144] and *Thermus thermophilus*.[145]

8.06.3.2.3   Investigating intramolecular ammonia translocation

The primary experimental evidence for intramolecular ammonia translocation between the glutaminase- and PRFAR-binding sites is the observation that mutation of Arg5 to a histidine residue in the HisF subunit
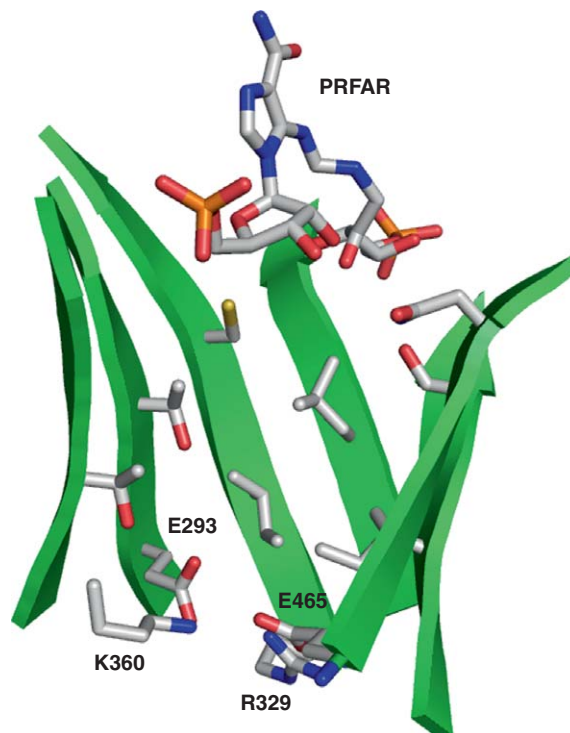
PRFAR

Gln

NH₃

Glu

−H₂O
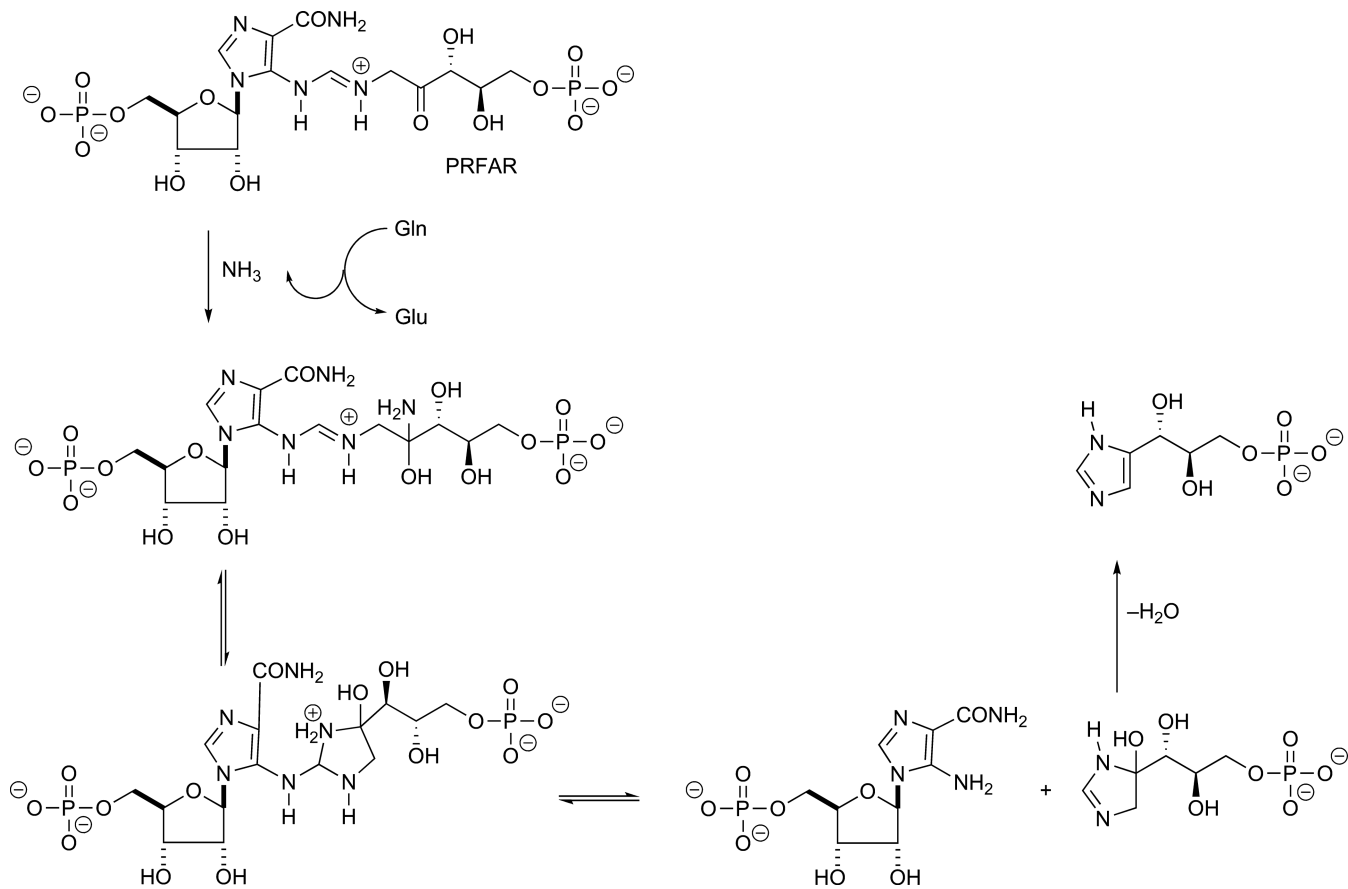
Scheme 13
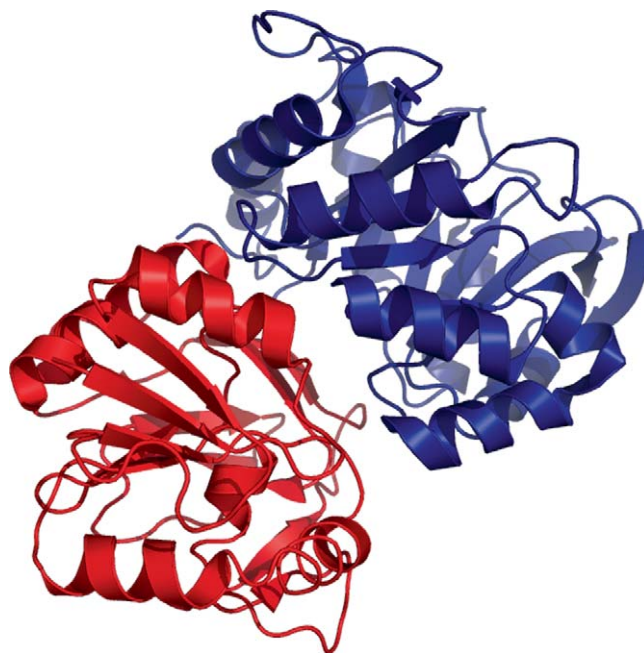
Figure 8 Cartoon representation of the HisF/HisH heterodimer from *Thermus thermophilus* (1KA9). The HisF (cyclase) and the HisH (glutaminase) subunits are colored blue and red, respectively. Image rendered in PYMOL.
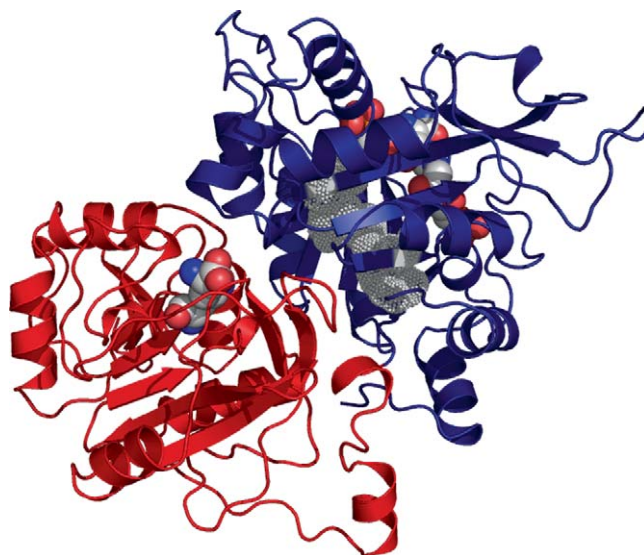

Figure 9 Cartoon representation of the *Saccharomyces cerevisiae* IGPS/PRFAR complex in which Cys83 is covalently modified with acivicin (1OX5) showing a cavity that is proposed to be part of the ammonia tunnel (gray dotted spheres). The glutaminase and cyclase domains are colored red and blue, respectively, and the acivicin and PRFAR moieties are rendered as CPK models: C – gray, O – red, N – blue, and P – orange. Image rendered in PYMOL.

decouples the glutaminase and synthase activities of the bacterial enzyme, presumably as a result of ammonia loss from the enzyme.[134] In addition, the presence of PRFAR stimulates the glutaminase activity of IGP synthase by a factor of 4900.[150] Perhaps the most striking evidence for the intramolecular translocation of ammonia through the proposed tunnel, however, has been provided by a series of computational studies using 'steered' molecular dynamics (MD) simulations.[151–153] In this method, which takes advantage of Jarzynski's

Figure 10 Close-up of the four layers of side chains located on β-strands of the C-terminal $(\beta/\alpha)_8$ barrel, which define the interior surface of the ammonia tunnel in *Saccharomyces cerevisiae* IGP synthase (1OX5). The PRFAR binding site is at the top of the tunnel, where the PRFAR ligand is shown as a CPK model: C – gray, O – red, N – blue, and P – orange. Image rendered in PYMOL.
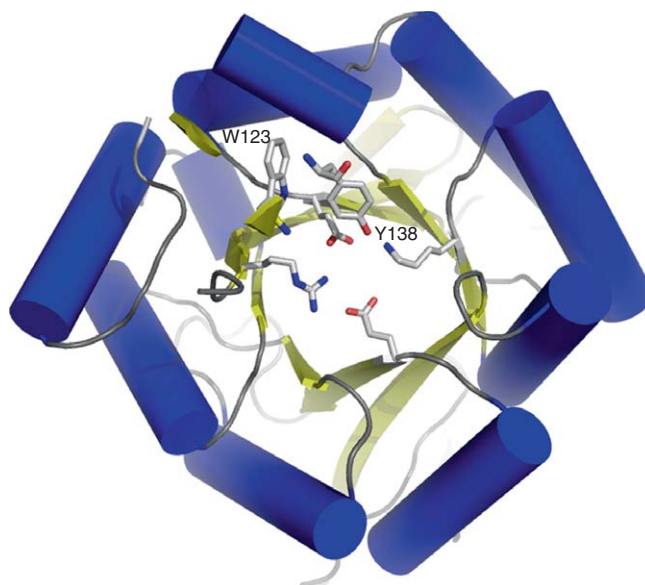
Figure 11 View of the *Thermotoga maritima* HisF subunit (1GPW) showing the location of the 'gate' into the synthase active site defined by the side chains of Arg5, Glu46, Lys99, and Glu167 (rendered as cylinders). The entry of bulk water is thought to be precluded by the presence of the Trp123 and Tyr138 side chains (cylinders). Coloring: C – gray, O – red, and N – blue. Helices and β-strands are rendered as blue cylinders and yellow arrows, respectively. Image rendered in PYMOL.

identity,[154] an 'external' force is applied to the ammonia molecule to 'push' it through the tunnel, while the protein structure undergoes dynamical motions.[155,156] These simulations therefore allow an assessment of the free energy barriers to ammonia movement along an intramolecular tunnel. In early work on the free *T. maritima* enzyme, it was found that ammonia can indeed pass through the center of the HisF $(\beta/\alpha)_8$ barrel,[152,153] and that the passage of ammonium ion is precluded by hydrophobic side chains defining the walls of the tunnel. The calculations also suggested that a water molecule remains within hydrogen bonding distance to the ammonia as this reactive intermediate moves through the tunnel to the synthase active site. Similar conclusions were reached in simulations on an enzyme model in which PRFAR was modeled into its HisF-binding site and the active site cysteine in HisH was modified to the γ-glutamyl thioester that would be present after ammonia generation.[153] A more interesting observation from the latter set of computational studies, however, was that the tunnel structure could discriminate between bulk water molecules and ammonia released by glutamine hydrolysis. This seemingly counterintuitive finding was explained on the basis of the differential electrostatic properties of water and ammonia.[151] Furthermore, the computational modification of Arg5 in the charged 'gate' to an alanine residue gave an IGP synthase variant in which bulk water could access chamber II of the tunnel during the MD simulation in contrast to the wild-type enzyme. The presence of a hydrogen-bonded water network in this region then appears to occlude the passage of ammonia through the protein.[152] Experimental support for this proposal was obtained by characterization of a series of IGP synthase variants in which residues involved in mediating this discrimination were changed using site-directed muta-genesis methods. In particular, Arg5, which is present in the charge gate region of the tunnel, was replaced by alanine. The coupling of glutaminase and amidotransferase activities in wild-type enzyme was lost in the resulting R5A IGP synthase variant under steady-state reaction conditions.

8.06.3.2.4   *Coordinating the catalytic activities of IGP synthase*

The glutaminase activity of IGP synthase is highly regulated by the presence of PRFAR with k_{cat}/K_M being negligible in the absence of the electrophilic substrate.[150] X-ray crystallography of the yeast enzyme showed that this is a consequence of conformational changes that take place on PRFAR binding,[142,150] which orient residues defining the catalytic triad in the glutaminase site correctly. These studies also showed that the relative orientation of the two domains in *S. cerevisiae* is altered by PRFAR binding. Thus, substrate binding causes the domains to move together thereby permitting the formation of new interdomain interactions, such as a salt bridge between Lys258, which is functionally important in the bacterial form of the enzyme, and Asp474 (**Figure 12**). Moreover, the position of a conserved glutamine (Gln397) is altered such that it can interact with bound glutamine in the IGP synthase/PRFAR complex. In the free enzyme, the domains move apart to give an 'open' structure and the Gln397 side chain cannot reach the glutaminase active site. Site-directed mutagenesis experiments were then used to explore which PRFAR/protein interactions might be important in causing these structural changes, revealing the importance of Lys258, which not only interacts with the substrate but also forms a salt bridge with the side chain of a conserved aspartate (Asp474).

The dissection of the molecular mechanisms mediating the coordination of glutaminase and synthase activities in the yeast IGP synthase has also been aided by computational simulations, which have provided information on the function and dynamical response of tunnel residues when ammonia is present in the active enzyme. By combining equilibrium[155] and steered MD simulations,[156] with site-directed mutagenesis and kinetic methods, the importance of a network of conserved residues,[157] including an interdomain salt bridge between Lys196 and Asp359,[158] in mediating the allosteric communication between the catalytic domains of yeast IGP synthase has been demonstrated. For example, the interaction of Lys196 and Asp359 appears to be critical in correctly orienting the side chain of His193, which mediates general acid/base catalysis during glutamine hydrolysis.[158] These studies represent the most detailed molecular analysis of catalytic coordination in a glutamine-dependent amidotransferase reported to date.

8.06.3.3   **Formylglycinamide Ribonucleotide Synthetase**

8.06.3.3.1   *Catalytic mechanism*

The conversion of formylglycinamide ribonucleotide (FGAR) into formylglycinamidine (FGAM) is the fourth step in the *de novo* biosynthesis of purines.[4,136] This reaction is catalyzed by formylglycinamide ribonucleotide
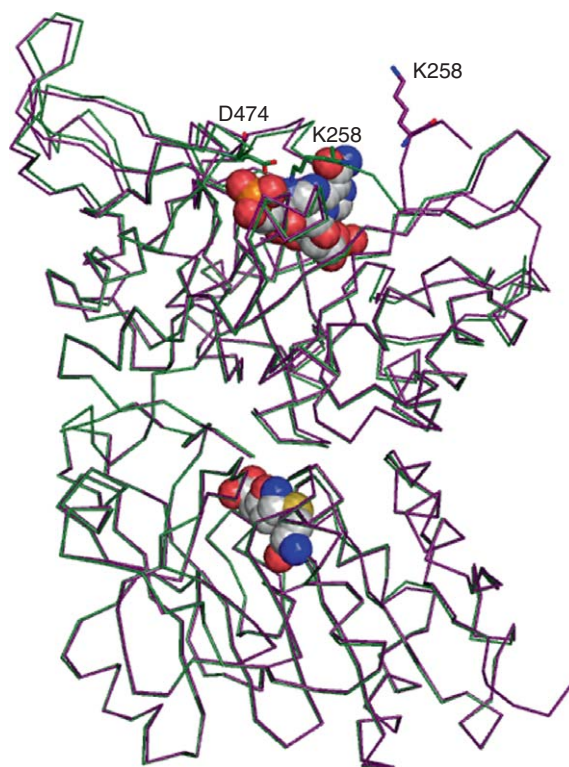
Figure 12 Conformational changes in *Saccharomyces cerevisiae* IGP synthase resulting from PRFAR binding in the C-terminal domain. The movement of the Lys258 side chain to form a salt bridge with Arg474 upon PRFAR binding is evident from superimposition of the C_α backbones of the IGPS/acivicin (green) (1OX4) and the IGPS/acivicin/PRFAR complex (violet) (1OX5). PRFAR and the acivicin moiety are rendered as CPK models, and the Lys258 and Arg474 side chains are shown as cylinders. Coloring: C – gray, O – red, N – blue, S – yellow, and P – orange. Image rendered in PYMOL.
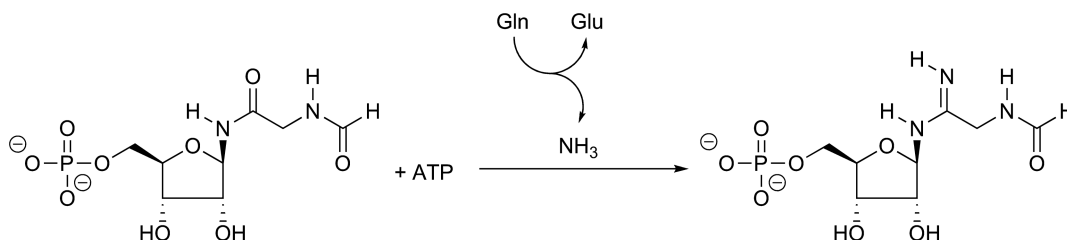
Scheme 14

synthetase (FGAR-AT) using ATP as an energy source and L-glutamine as the nitrogen donor (**Scheme 14**).[159,160] Two forms of the enzyme have been identified. In eukaryotes and most Gram-negative bacteria, FGAR-AT is a multidomain protein that is encoded by the *purL* gene as a single polypeptide chain of ~1300 amino acids (lgPurL).[159] In Gram-positive bacteria, such as *Bacillus subtilis*, and archaebacteria, however, the active enzyme exists as a complex of three separate subunits (PurS, smPurL, and PurQ).[161] The genes for these three proteins in *B. subtilis* have been cloned,[162] and used to obtain the recombinant proteins in recent studies of the function of each protein in FGAM synthesis. As expected from its sequence homology to other Class I amidotransferase glutaminase domains,[67] PurQ catalyzes the release of ammonia from glutamine and can be covalently modified by incubation with DON.[161] ATP and FGAR have been shown to bind to smPurL,[161] and PurS (the expression of which is required for the production of active enzyme[163]) seems to be implicated in orienting the PurQ and smPurL subunits correctly, presumably so that ammonia can diffuse through a solvent-inaccessible tunnel connecting the
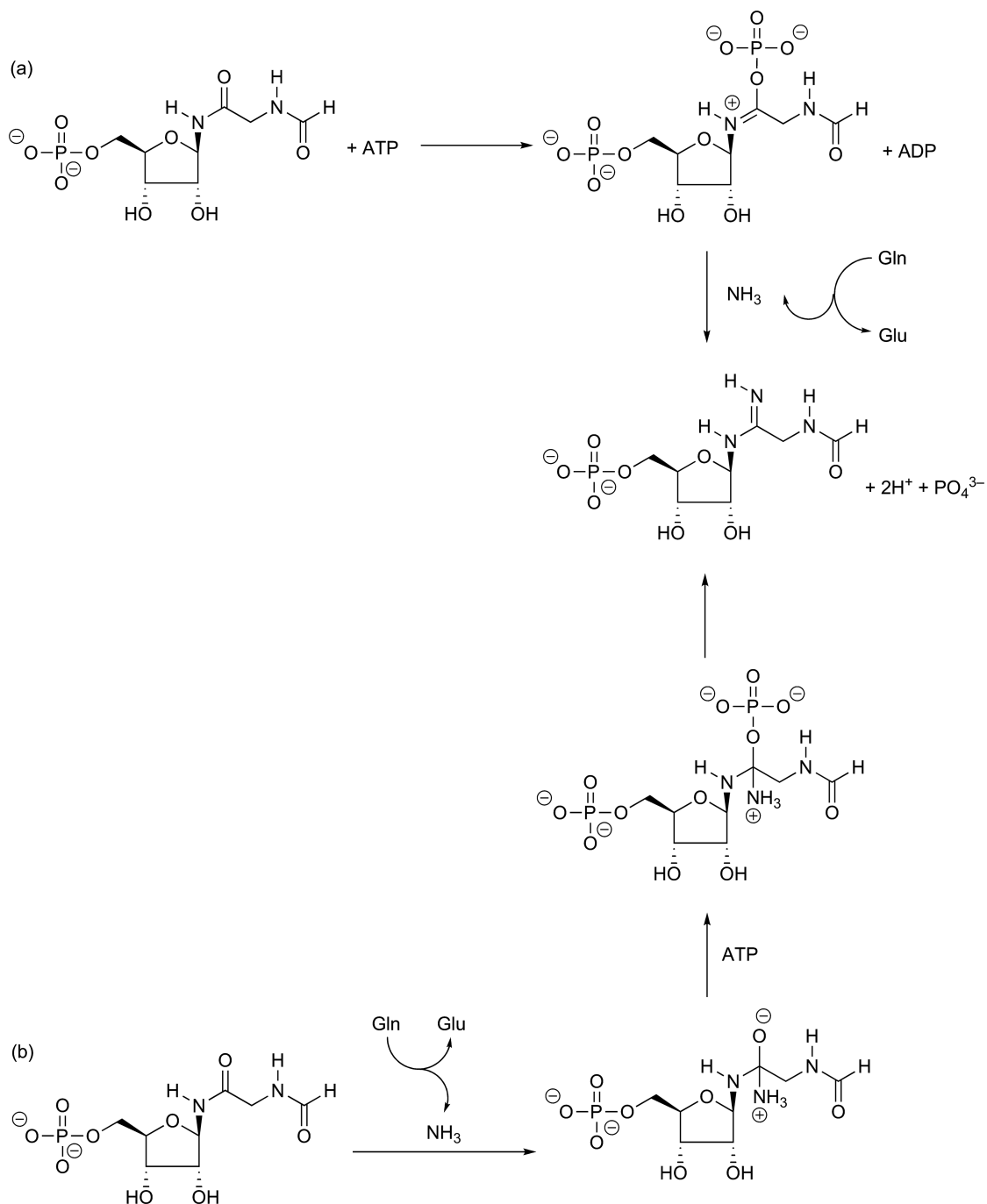
Scheme 15

active sites of the enzyme.[164] In the case of lgPurL isolated from chicken liver, it has been shown that the glutaminase activity of the enzyme is negligible in the absence of FGAR and ATP, and that the oxygen atom of the aminated carbonyl group is transferred to P_i during catalytic turnover.[165] This is consistent with a mechanism in which ATP and FGAR react to yield an iminophosphate intermediate that can undergo subsequent reaction with ammonia (**Scheme 15(a)**). An alternate mechanism can be envisaged in which ammonia reacts with the amide to yield a tetrahedral intermediate that subsequently undergoes phosphorylation and elimination (**Scheme 15(b)**).[166]
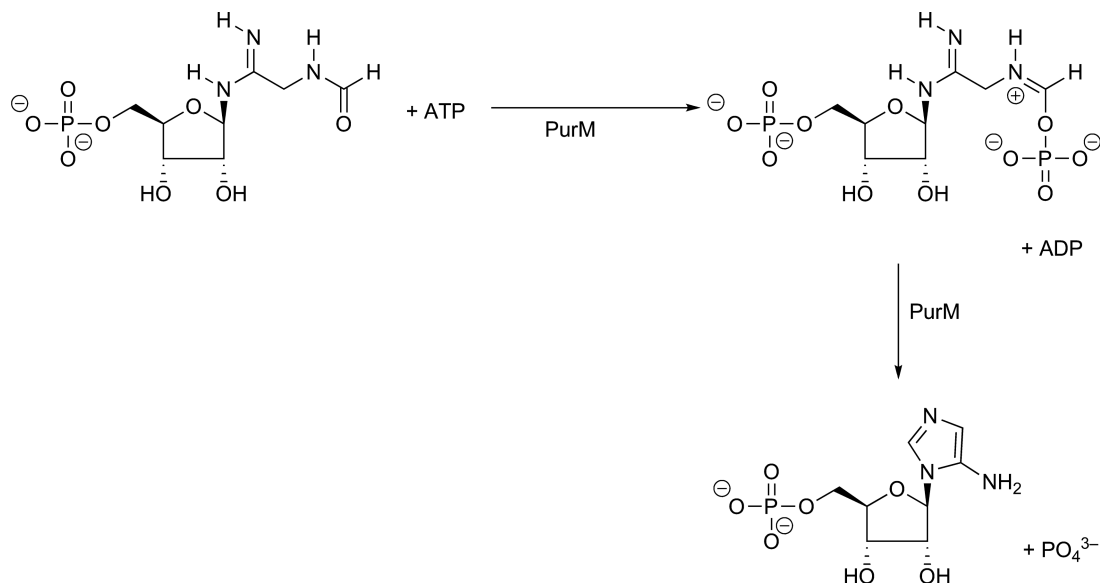
Scheme 16

This chemistry seems less probable, however, given the structural similarity of the smPurL protein[167,168] and aminoimidazole ribonucleotide (AIR) synthetase (PurM), another enzyme in the purine biosynthetic pathway that is known to form an iminophosphate intermediate (**Scheme 16**).[169] Under *in vitro* conditions, FGAR-AT can use ammonia as an alternate nitrogen source when glutamine is absent.

8.06.3.3.2 Enzyme structure

The X-ray crystal structure of *Salmonella typhimurium* lgPurL shows that the protein is composed of four domains (**Figure 13**).[170] The glutaminase active site is located within the fourth, C-terminal domain, as
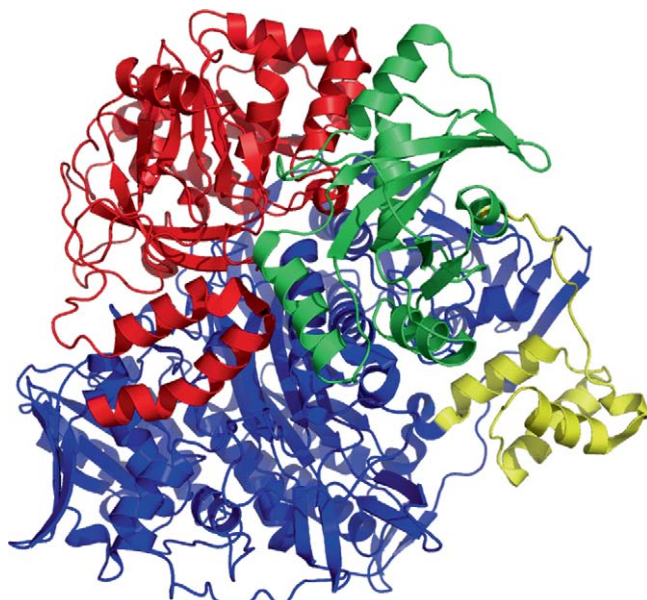
Figure 13 Cartoon representation of *Salmonella typhimurium* lgPurL (1T3T). The N-terminal domain (1–140) is colored green, the linker domain (141–214) is colored yellow, the FGAM synthetase domain (215–979) is colored blue, and the glutaminase domain (980–1295) is colored red. Image rendered in PYMOL.

evidenced by (1) electron density consistent with a γ-glutamyl thioester intermediate formed from the side chain of Cys1135[170] and (2) the structural similarity of this domain to the glutaminase subunits present in CPS,[103] HisH,[144] anthranilate synthase,[27] and GMPS.[25] Two sulfate ions are also observed bound to the third domain of the enzyme, the fold of which resembles that seen in the dimeric structures of the PurS proteins from and *B. subtilis*. The separation of the sulfates is equivalent to that between the phosphate groups of FGAR, and the synthetase site was therefore assigned to this domain of the protein. It is interesting to note that this domain is composed of two similar subdomains, and resembles the PurM dimer from *E. coli*.[167] Indeed, this fold has also been observed in NiFe-hydrogenase maturation protein,[171] and sequence comparisons suggest it may also be adopted by selenophosphate synthetase[172] and thiamine phosphate kinase.[173] Finally, the N-terminal domain of *S. typhimurium* lgPurL, which is linked to the synthetase domain by a domain composed predominantly of hydrophobic residues, is structurally homologous to the PurS dimers present in *B. subtilis*[164] and *Methanobacterium thermoautotrophicum*.[174] As discussed above, the most likely function for this PurS-like domain is to position the synthetase and glutaminase domains of the enzyme so that the active sites are connected by an ammonia tunnel. The three-dimensional domain organization of the protein also seems to depend on an MgADP molecule, which is bound within a hydrophobic cavity formed by Val333, Phe335, Met366, Pro370, Phe389, Phe651, and Leu652 (**Figure 14**). In an interesting finding, this ADP could not be exchanged with its tritiated isotopomer using dialysis, being removed only by denaturation of the enzyme. Several residues that interact with this ligand are conserved in the lgPurL and smPurL enzyme families, and its general structural importance may be illustrated by the observation that ADP is required for assembly of the functional *B. subtilis* smPurL/PurS/PurQ complex.[161] Two possible tunnels could be identified in the *S. typhimurium* lgPurL structure through which ammonia might be translocated from the glutaminase to the synthetase sites, which are separated by a distance of ∼30 Å (**Figure 15**).[170] Of these, one tunnel is formed from a number of conserved residues (Gly313, Ile316, Arg317, and Phe380) and therefore seems more likely to mediate ammonia transloca- tion. It seems likely that conformational changes in the N-terminal domain of lgPurL as a result of ammonia synthesis give rise to a transient, well-defined tunnel linking the active sites.

A series of titration experiments together with the structure of *S. typhimurium* lgPurL has provided some insight into the likely organization of the enzyme complex composed of the *B. subtilis* PurS, smPurL, and PurQ subunits. Thus, maximal enzyme activity is observed when these three proteins are mixed in a 2:1:1 ratio, respectively.[161] The formation of this complex, however, also requires the presence of the three substrates (FGAR, MgATP, and glutamine), complicating efforts to obtain a high-resolution crystal structure for the active form of *B. subtilis* FGAR-AT. The finding that smPurL is more stable when coexpressed with the PurS and PurQ components, however, has been interpreted as evidence that these three proteins may be weakly associated within the cell. A model for such a complex in *B. subtilis* has been constructed based on the domain
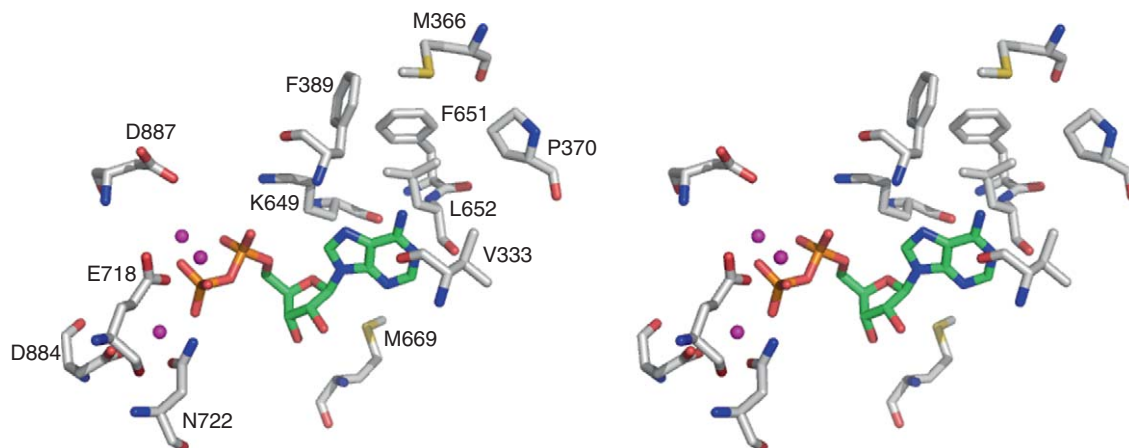

Figure 14   Stereoimage showing the highly hydrophobic, auxiliary MgADP-binding site in *Salmonella typhimurium* lgPurL (1T3T). Mg^{2+} ions are represented by magenta spheres. Coloring: C – green (ADP) and gray (protein), O – red, N – blue, P – orange, and S – yellow. Image rendered in PYMOL.
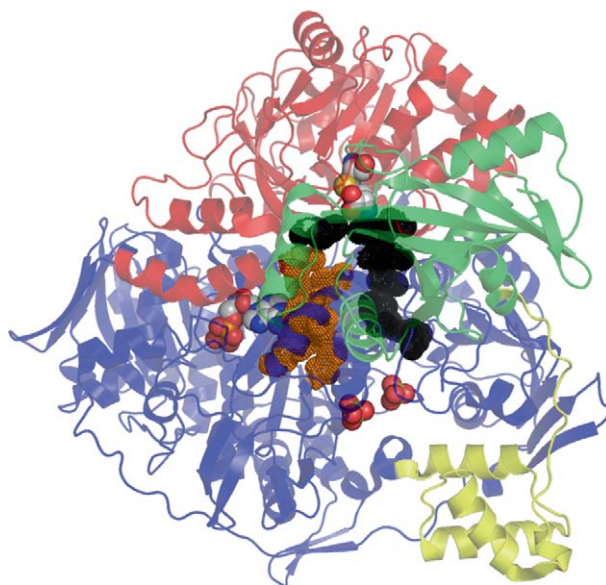
Figure 15 Cartoon representation of the two putative ammonia tunnels observed in the *Salmonella typhimurium* lgPurL (1T3T). The favored pathway (black dotted spheres) is formed from the side chains of conserved residues in the enzyme. The other putative tunnel (orange dotted spheres) passes close to the auxiliary ADP-binding site. Protein domain coloring is identical to that used in **Figure 13**, and atoms in the γ-glutamylthioester intermediate, bound ADP, and the sulfate ions are shown as CPK models. Coloring: C – gray, O – red, N – blue, S – yellow, and P – orange. Image rendered in PYMOL.

structure of the *S. typhimurium* FGAR-AT,[170] permitting a preliminary assignment of residues defining the ammonia tunnel.[164] For example, residues Arg121, Ile131, and Glu252 in the *Bacillus* smPurL subunit seem to define the entrance into the FGAR-binding site, and are also present (Arg317, Ile316, and Glu478) in the *Salmonella* enzyme. Similarly, Phe1094 in lgPurL has a counterpart in *B. subtilis* PurQ, and it is possible that this residue controls the entry of ammonia into the tunnel after breakdown of the initial tetrahedral intermediate.

Details of the molecular interactions of FGAR and ATP with the synthetase site have been delineated from a series of high-resolution crystal structures of *T. maritima* smPurL complexed with substrates, a substrate analog and the reaction product.[175] These structures not only revealed a novel ATP-binding motif (**Figure 16**) but also suggested that the secondary, catalytically nonactive ADP-binding site that is also present in the *S. typhimurium* lgPurL, actually binds to ATP. Perhaps most importantly for efforts to resolve the catalytic mechanism, the structure of a ternary complex in which FGAR-AT was bound to AMPCPP (an unreactive ATP analog) and
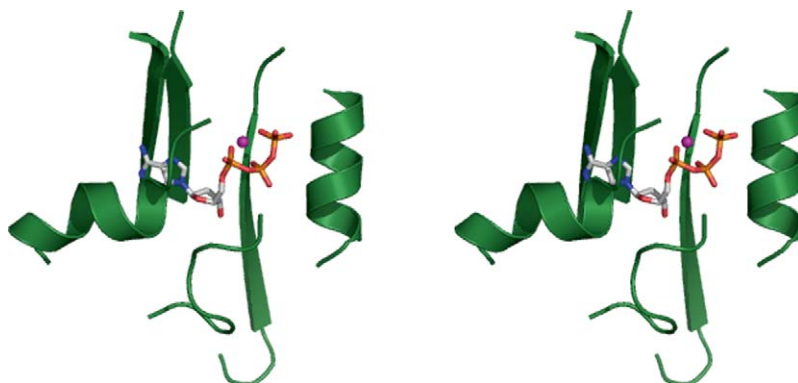
Figure 16 Stereoimage showing the novel ATP binding motif in the auxiliary binding site of the *Thermotoga maritima* lgPurL/ ATP complex (2HS0). A bound Mg^{2+} ion is rendered as a magenta sphere, and protein secondary structural elements are shown as cartoon representations (green). ATP coloring: C – gray, O – red, N – blue, and P – orange. Image rendered in PYMOL.
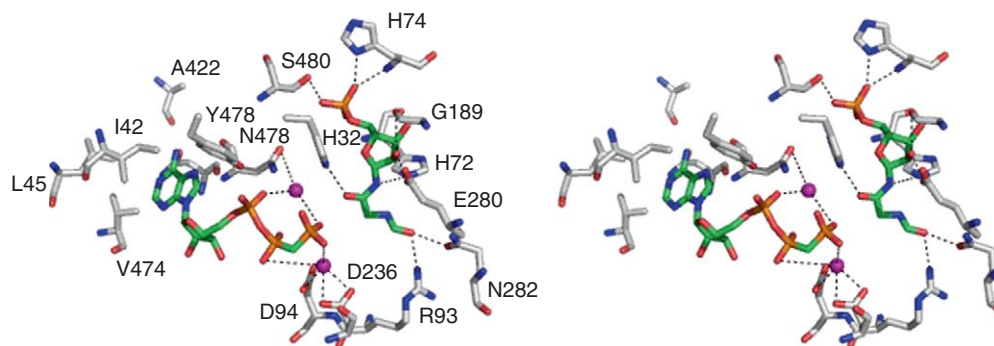
Figure 17 Stereoimage showing the active site of the *Thermotoga maritima* lgPurL/AMPPCP/FGAR ternary complex (2HS4). Bound Mg^{2+} ions are rendered as magenta spheres, and hydrogen bonds and metal–ligand bonds are indicated by dashed lines. Coloring: C – green (adenosine 5′ (β,γ-methylene) triphosphate (AMPPCP) and FGAR ligands) and gray (protein), O – red, N – blue, and P – orange. Image rendered in PYMOL.

FGAR was obtained (**Figure 17**). Substrate binding reorganized the conformation of an active site loop, and suggested catalytic roles for two histidine residues. This hypothesis was subsequently confirmed by site-directed mutagenesis experiments.

8.06.3.3.3   Coordinating the catalytic activities in FGAR-AT

To date, there have been no experiments similar to those performed on CPS that directly demonstrate the functional importance of the ammonia tunnels hypothesized to be present in either lgPurL or the *B. subtilis* ternary smPurL/PurS/PurQ complex. Of course, this reflects the relatively recent success in obtaining structural information for FGAR-AT, but the importance of glutamine as a nitrogen source is clearly evident from the very low ammonia-dependent synthetase activity observed for *E. coli* FGAR-AT, which is only 2% of the glutamine-dependent reaction.[159] Studies on the γ-glutamyl enzyme, isolated by incubation of chicken liver FGAR-AT with [^{14}C]-glutamine,[176] have provided evidence for 'enzyme-bound' ammonia. In addition, this form of the enzyme exhibits almost no glutaminase activity in the absence of other substrates.[177] The existence of interdomain or intersubunit interactions that coordinate the glutaminase and synthetase activities therefore seems likely, although these remain to be identified.

8.06.3.4   Cytidine 5′-Triphosphate Synthetase

8.06.3.4.1   Catalytic mechanism

Cytidine triphosphate synthetase (CTPS) mediates the conversion of uridine triphosphate (UTP) into cytidine triphosphate (CTP) (**Scheme 17**), which is the rate-limiting step of *de novo* CTP biosynthesis.[46,178] As a consequence, CTPS activity regulates the intracellular rates of RNA, DNA, and phospholipid synthesis.[179,180] This enzyme has therefore been selected as a target for the development of drugs against leukemia[181] and parasitic infections,[182,183] and its kinetic properties have been the subject of numerous studies.[184–187] GTP is an
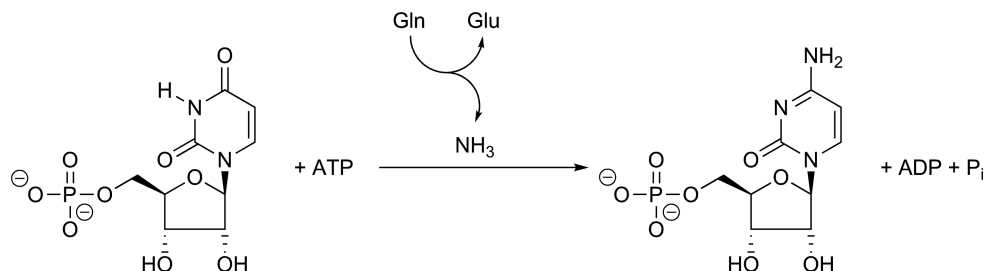

Scheme 17
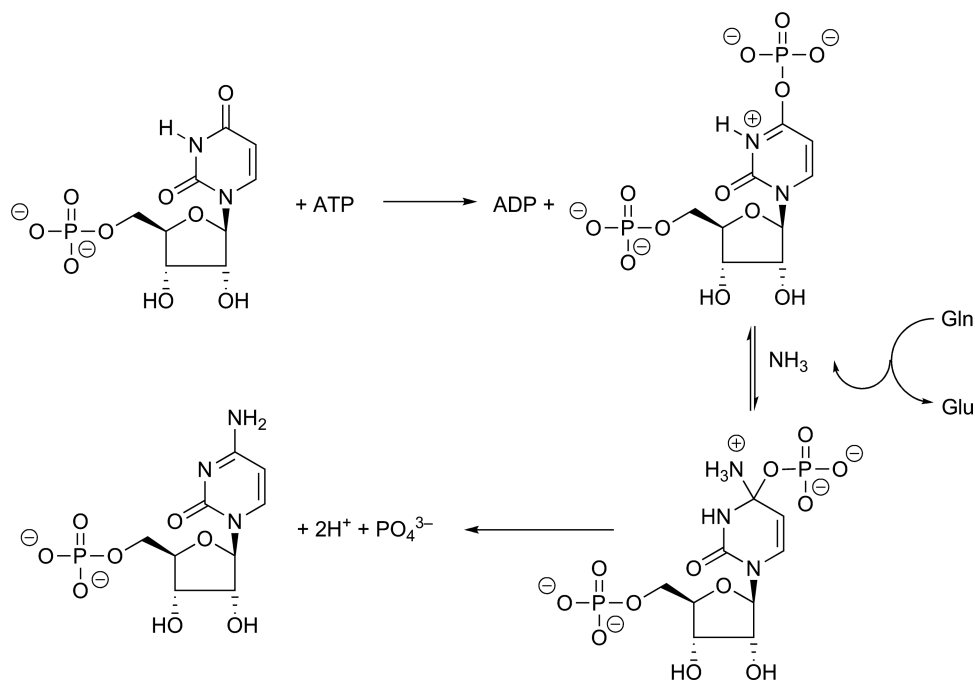
Scheme 18

allosteric activator of CTPS,[188] and GTP binding enhances the utilization of ammonia released in the glutaminase domain of the CTPS in *Lactococcus lactis*.[189] Enzyme activity is downregulated by CTP, which competes with UTP for binding to the synthetase site,[190] and there is evidence that phosphorylation at a number of serine residues is an important mechanism for the regulation of both the human and yeast CTP synthetases.[191,192] Substrate-induced changes in CTPS oligomerization also appear to regulate the activity of the *E. coli* enzyme.[193] The catalytic mechanism of the enzyme is straightforward, with activation of UTP by phosphorylation using ATP to give an iminophosphate intermediate that can undergo an addition–elimination mechanism with ammonia to yield CTP (**Scheme 18**). Evidence for this mechanism has been provided by isotope partitioning experiments, and the isolation of the phosphorylated intermediate.[184]

8.06.3.4.2   *Enzyme structure*

High-resolution, X-ray crystal structures for the CTP synthetases from *E. coli* and *T. thermophilus* have been reported,[194–196] and these show both the glutaminase and synthetase sites of the enzyme to be spatially separated by a distance of ∼25 Å (**Figure 18**). The structure of the synthetase domain of the human CTPS has also been determined.[197] As anticipated on the basis of sequence alignment studies, the glutaminase domains of the two bacterial enzymes have a three-dimensional fold that is very similar to that seen for the CPS small subunit, providing further evidence for the idea that all the glutaminase domains, or subunits, in Class I amidotransferases have evolved from a common ancestor. In addition, the synthetase domain is homologous to dethiobiotin synthase,[198] an enzyme that forms an iminophosphate intermediate during catalytic turnover.[199] The structure of *T. thermophilus* CTPS complexed with three sulfate anions suggested the location of the UTP- and ATP-binding sites to be located at the interface of the glutaminase and synthetase domains.[196] Unfortunately, both catalytic sites were solvent accessible, implying that the formation of an ammonia tunnel in this enzyme requires conformational changes that are driven by binding of the three substrates and/or allosteric CTPS regulators. Better evidence for the existence of a tunnel in CTPS was obtained from the crystal structure of the free *E. coli* enzyme,[194] which revealed a solvent-filled 'vestibule' between the interface of the N-terminal synthetase (residues 1–266) and C-terminal glutaminase (residues 287–544) domains and the catalytic Cys–His–Glu triad (**Figure 19**). As anticipated from mutagenesis experiments probing the molecular mechanisms by which GTP enhances the glutaminase activity of CTPS,[200–202] this cavity is located close to the GTP-binding site. The cavity was lined

(a)

(b)
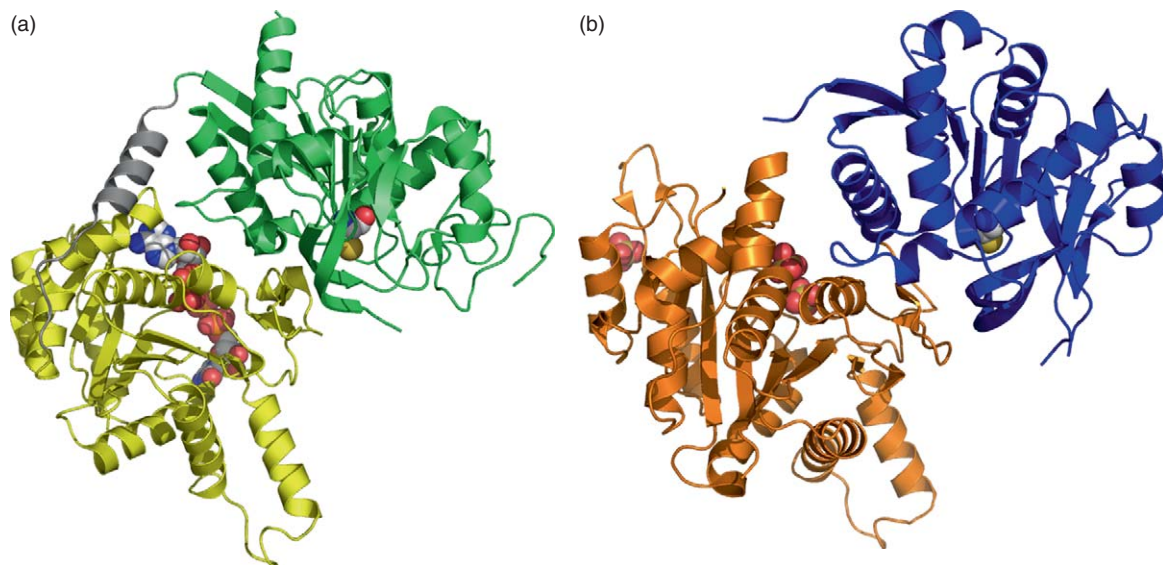
Figure 18 (a) *Escherichia coli* CTP synthetase bound to ADP and CTP (2AD5). The glutaminase and synthase domains are colored green and yellow, respectively, and the linker peptide is shown in gray. The catalytic cysteine in the glutaminase domain (Cys379) and the ADP and CTP ligands are rendered as CPK models. (b) *Thermus thermophilus* CTP synthetase bound to three sulfate ions (1VCN). The glutaminase and synthase domains are colored blue and red, respectively, and the catalytic cysteine in the glutaminase domain (Cys391) and sulfate ions are rendered as CPK models: C – gray, O – red, N –blue, P – orange, and S – yellow. Image rendered in PYMOL.
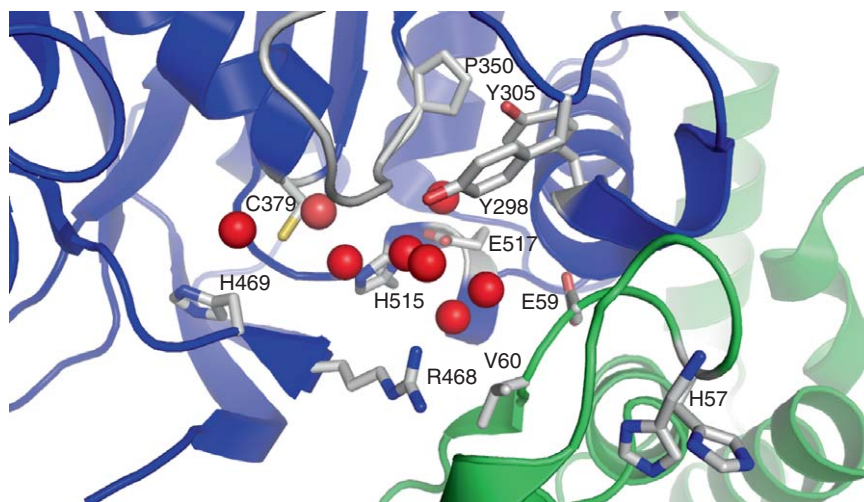
Figure 19 Close-up view of the solvent-filled vestibule in *Escherichia coli* CTP synthetase, which is proposed to form part of the ammonia tunnel in the enzyme (1S1M). Red spheres indicate water molecules, and the glutaminase and synthetase domains are colored blue and green, respectively. Residues in the flexible loop (L11) are colored gray, while those lining the tunnel are represented as 'sticks'. Coloring: C – gray, O – red, N – blue, and S – yellow. Image rendered in PYMOL.

by polar residues (Tyr298, Tyr305, and Arg468) and a flexible loop (L11), and could provide a means by which exogenous ammonia, which can be used by CTPS as an alternate nitrogen source,[203] enters the synthetase site of the enzyme. Mutations in the L11 loop have also been shown to affect the GTP-dependent activation of CTPS glutaminase activity.[204] In a further interesting observation, the His57 side chain adopted two conformations in the structure, one of which blocks access to the synthetase site from the cavity. It was therefore hypothesized that UTP and/or ATP binding to the N-terminal domain might be an essential element in causing conformational changes

that open the tunnel by moving the His57 side chain. An obvious tunnel connecting the two active sites was not observed, however, in subsequent studies of the ternary complex involving *E. coli* CTPS, ADP, and CTP,[195] although this structure provided significant insights into the role of mutations conferring resistance to a number of clinically important pyrimidine antimetabolites.[205,206] Tunnel stabilization may therefore depend on glutamine binding and/or thioester formation in the presence of the iminophosphate intermediate.

8.06.3.4.3   *Investigating intramolecular ammonia translocation*

The primary chemical evidence to support intramolecular tunneling of ammonia released in the glutaminase active site has been provided by studies examining the pH dependence of the glutamine- and ammonia-dependent synthetase activities of *E. coli* CTPS.[185] At pH 10.4, the turnover number of the enzyme when ammonia is the nitrogen source exceeds that observed for glutamine. Lowering the solution pH, however, abolishes ammonia-dependent activity because the ammonium ion is not a substrate for the enzyme. At pH 7.2, equilibration of ammonia released in the glutaminase site with that in free solution was calculated to give an overall rate that was 0.45% of that observed for the glutamine-dependent synthetase activity of *E. coli* CTPS. It therefore seems likely that ammonia released from glutamine remains bound to the enzyme and is translocated along a tunnel connecting the two active sites. Further evidence for this hypothesis has been provided by more recent experiments that examined the effects of modifying Leu109 to alanine in *E. coli* CTPS on the reactivity of alternate substrates for the enzyme.[207] Thus, CTPS was shown capable of catalyzing the hydrolysis of L-γ-glutamyl hydroxamate and L-γ-glutamyl hydrazide to yield hydroxylamine and hydrazine, respectively, which can be regarded as bulkier analogs of ammonia. Moreover, both hydroxylamine and hydrazine could be used in place of ammonia to give the corresponding CTP analogs (**Scheme 19**). Kinetic measurements showed that ammonia, hydroxylamine, and hydrazine exhibited very similar values of k_{cat}/K_M in the presence of saturating ATP and UTP for both the wild-type CTPS and the L109A CTPS variant under steady-state conditions (**Table 1**). Both L-γ-glutamyl hydroxamate and L-γ-glutamyl hydrazide could also be used as substrates by the wild-type enzyme, albeit with reduced k_{cat}/K_M values relative to that for glutamine. The replacement of Leu109 by alanine, however, resulted in an inability of the L109A CTPS variant to use L-γ-glutamyl hydrazide as a substrate, and reduced k_{cat}/K_M for L-γ-
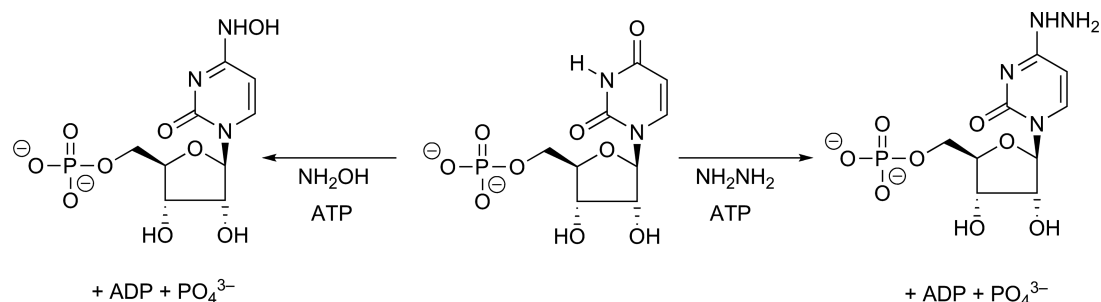
Scheme 19

Table 1   Steady-state kinetic parameters for CTP formation catalyzed by wild-type CTP synthetase and the L109A CTP synthetase mutant using various nitrogen sources

Substrate	CTP synthetase			L109A CTP synthetase mutant		
	k_{cat} (s^{-1})	K_M (mmol l^{-1})	k_{cat}/K_M (mmol^{-1}/s^{-1})	k_{cat} (s^{-1})	K_M (mmol l^{-1})	k_{cat}/K_M (mmol^{-1}/s^{-1})
NH$_3$	9.5 ± 0.5	2.1 ± 0.1	4.4 ± 0.1	10.1 ± 0.3	2.2 ± 0.1	4.63 ± 0.04
NH$_2$OH	14 ± 2	83 ± 7	0.17 ± 0.02	14 ± 2	80 ± 10	0.187 ± 0.003
H$_2$NNH$_2$	–[a]	–[a]	0.15 ± 0.02	–[a]	–[a]	0.13 ± 0.01
Glutamine	6.1 ± 0.8	0.35 ± 0.06	18 ± 2	1.9 ± 0.3	0.5 ± 0.1	3.8 ± 0.8

[a] Not determined. Saturation could not be achieved and the k_{cat}/K_M value was determined at values of [H$_2$NNH$_2$] << K_M.
Data adapted from F. A. Lunn; S. L. Bearne, *Eur. J. Biochem.* **2004**, *271*, 4204–4212.

glutamyl hydroxamate by an order of magnitude. Modeling studies using the *E. coli* crystal structure suggested that removal of the leucine side chain causes a 'kink' in the putative ammonia tunnel, thereby preventing the translocation of hydroxylamine and hydrazine released in the glutaminase site.

8.06.3.4.4   Coordinating the catalytic activities of CTPS

In addition to being implicated in the structural changes that lead to ammonia tunnel formation, Leu109 also appears to be important in coordinating the glutaminase and synthetase activities of *E. coli* CTPS.[201] For example, when this residue was changed to alanine, the L109A CTPS mutant exhibited similar steady-state ammonia-dependent synthetase and glutaminase activities to those seen for the wild-type enzyme. The replacement of this conserved leucine residue, however, impaired the glutamine-dependent formation of CTP even though the affinity of the mutant enzyme for GTP was reduced by an insignificant amount. Leu109 is located within a contiguous segment of 16 amino acids, all of which are highly conserved in known CTP synthetases (**Figure 20**). Scanning alanine mutagenesis experiments demonstrated that the R105A, D107A, and G110A CTPS mutants all exhibited decoupling of glutamine-dependent CTP formation while retaining the ammonia-dependent synthetase activity of the wild-type enzyme. These findings support the hypothesis that this loop segment is involved in stabilizing the ammonia tunnel and perhaps in coordinating the catalytic activities of the spatially separated active sites, perhaps by interacting with residues in the C-terminal glutaminase domain. The subsequent X-ray crystal structure of free *E. coli* CTPS,[194] which was reported after completion of the steady-state kinetic experiments,[201] provides some support for this idea because the Leu109 side lies close to the 'vestibule' that is thought to form the tunnel after substrate/intermediate-induced conformational changes. The absence of structural information for the enzyme in its active, tetrameric form has precluded the identification of additional residues that might mediate other domain–domain interactions, which coordinate the two catalytic activities of CTPS.

8.06.3.5   Pyridoxal Phosphate Synthase

8.06.3.5.1   Catalytic mechanism

Despite the importance of pyridoxal phosphate (PLP) in cellular metabolism, the pathways by which this cofactor is synthesized have been characterized relatively recently.[43,208] Two mutually exclusive pathways have been described, with eubacteria using deoxyxylulose-5-phosphate (DXP) as a key intermediate (**Scheme 20**).[209] More recent work has identified a second, DXP-independent pathway (**Scheme 21**),[210,211] which appears to be used by a wide range of organisms and is distinguished by the involvement of PLP synthase.[212,213] In *B. subtilis*,[212] fungi,[214] *Plasmodium falciparum*,[215] and plants,[211] this enzyme is a heterodimer composed of two subunits encoded
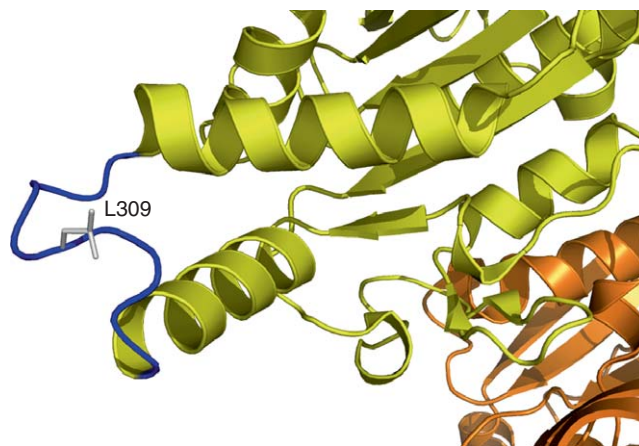
Figure 20   Highly conserved loop region of CTPS (1S1M). The loop containing 16 highly conserved residues (102–118, *Escherichia coli* numbering) is shown in blue, with bonds in the Leu309 being rendered as cylinders. Residues in the synthase and glutaminase domains are colored yellow and orange, respectively. Image rendered in PYMOL.
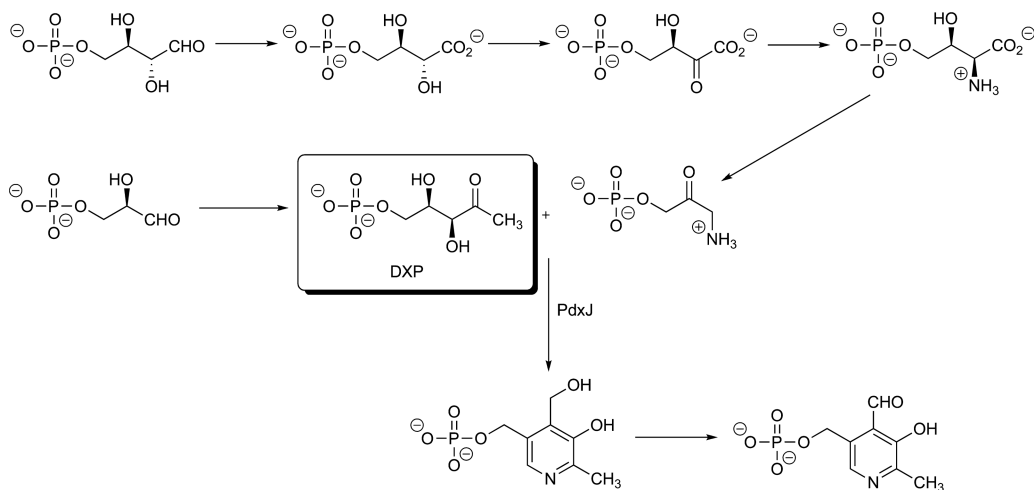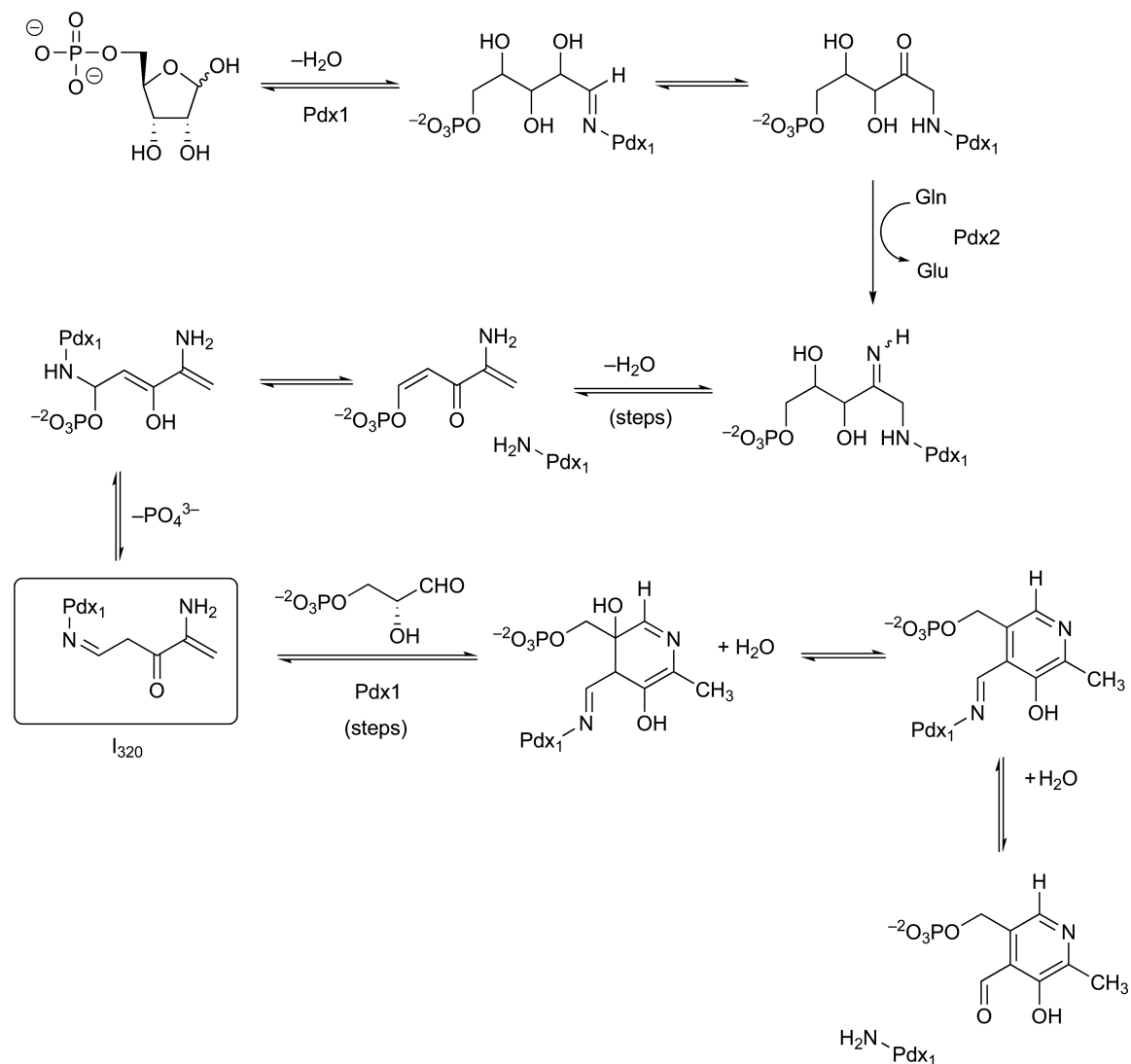
Scheme 20
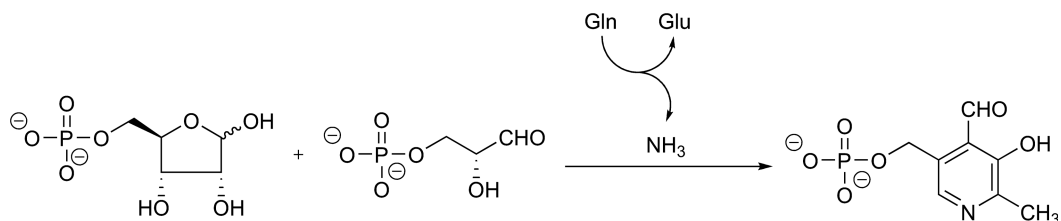
Scheme 21

by the *Pdx1* and *Pdx2* genes. In agreement with the results of ^{15}N-labeling experiments, which showed that the nitrogen atom of the pyridine ring in PLP was derived from glutamine rather than glutamate,[216] Pdx2 is a glutaminase that is inhibited by acivicin,[212] and is homologous to the glutaminase subunit/domains of other Class I amidotransferases. As is the case for the HisH subunit of imidazole glycerol-3-phosphate synthase (IGPS),[134] Pdx2 appears to be inactive in the absence of the Pdx1 protein.[215,217–220] PLP synthase is a remarkable catalyst, mediating the synthesis of PLP from ribulose-5-phosphate, glyceraldehyde-3-phosphate, and glutamine (**Scheme 21**). The enzyme can also use ribose-5-phosphate, dihydroxyacetonephosphate, and ammonia as alternate substrates, albeit less efficiently.[213,221] Given that the function of Pdx2 is merely to catalyze the release of ammonia from glutamine, residues in the Pdx1 active site must perform a plethora of chemical steps, including the isomerization of triose and pentose intermediates, C–C bond formation, and aromatic ring synthesis. Although the detection of an enzyme-bound chromophoric reaction intermediate has placed constraints on possible catalytic mechanisms used by PLP synthase,[221–223] the details of product formation remain ill-defined. It is known, however, that ribose-5-phosphate is isomerized to ribulose-5-phosphate, which forms an imine by reaction with the side chain of Lys81. Subsequent elimination of phosphate then yields the experimentally observed chromophore.[222,223] Reaction with ammonia, tautomerization, and aldol condensation with glyceraldehyde-3-phosphate then yields an intermediate that probably undergoes cyclization to yield the pyridine moiety to regenerate the Lys81 side chain (**Scheme 22**).

8.06.3.5.2 Enzyme structure

Several X-ray crystal structures have been reported for both the isolated Pdx1 and Pdx2 subunits,[215,224,225] and the PLP synthase heterodimer (**Figure 21(a)**).[226,227] Despite the unique primary structure of Pdx1, this protein possesses a $(\beta/\alpha)_8$ barrel fold similar to that of the HisF protein in IGPS.[144] In its active form, PLP synthase likely exists as a dodecamer composed of Pdx1/Pdx2 heterodimers,[226,227] with the Pdx1 subunits binding one Pdx2 monomer and interacting with other Pdx2 proteins to form a ring (**Figure 21(b)**). There are no direct

Scheme 22

interactions between the glutaminase subunits in the molecular assembly, which is unique among glutamine-dependent amidotransferases that have been structurally characterized to date. Activation of the Pdx2 subunit on binding to Pdx1 seems to be associated with a conformational rearrangement that reorients residues defining the oxyanion hole of the active site so that the initial tetrahedral adduct can be stabilized. Ammonia tunnels having a length of $\sim 26\,\text{Å}$ are seen in each of the individual *B. subtilis* heterodimers, and, as in the case of IGPS, ammonia must be translocated between hydrophobic, methionine side chains (Met13, Met43, Met79, and Met145) located in the interior of the $(\beta/\alpha)_8$-barrel of the Pdx1 subunit (**Figure 22**). The entrance to the synthase active site, however, is defined by charged residues (Asp25, Asp103, and Arg148).[226] Two absolutely conserved residues (Glu47 and Arg135) in the Pdx2 subunit are thought to function as a gate through which ammonia released in the glutaminase site can enter the putative tunnel. A similar tunnel exists in the holoenzyme from *T. maritima*.[227]

8.06.3.5.3   Coordinating the catalytic activities of PLP synthase

To date, relatively few kinetic studies have been reported for site-directed mutants of PLP synthase, and the ability of glutamine to suppress the incorporation of exogenous ammonia has not been examined. In contrast to
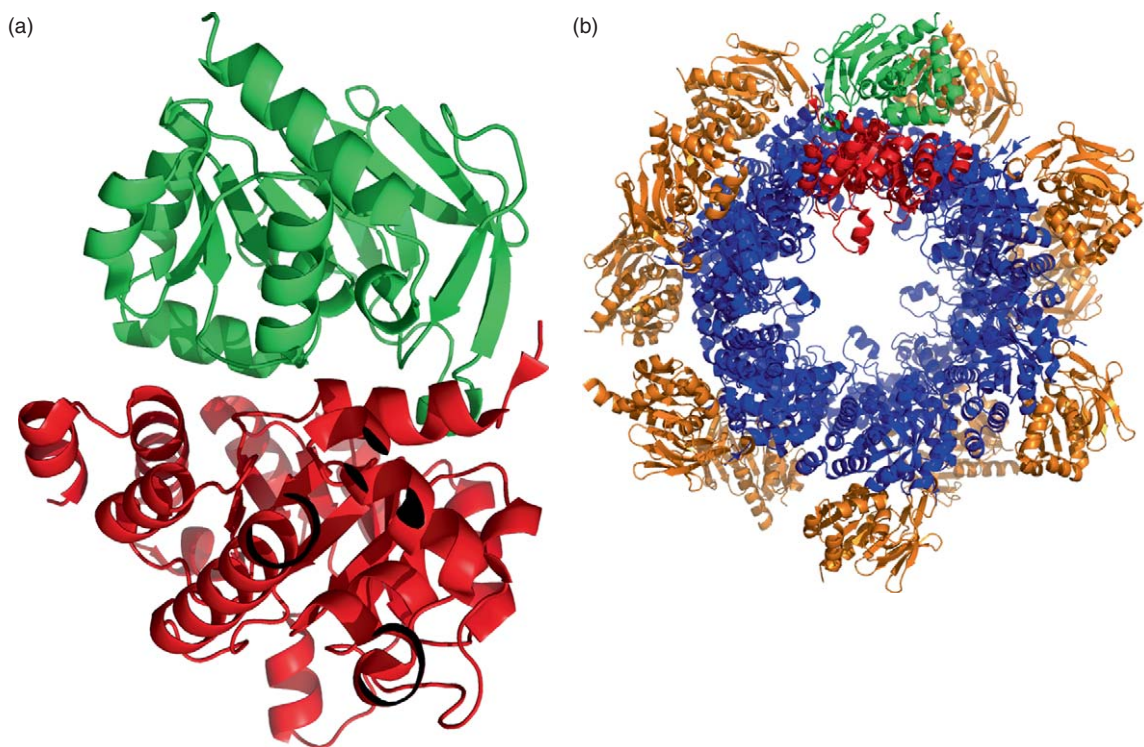
Figure 21   (a) Cartoon representation of the wild-type *Bacillus subtilis* PLP synthase heterodimer (2NV2). The glutaminase (Pdx2) and synthase (Pdx1) subunits are colored green and red, respectively. (b) The quaternary structure of *B. subtilis* PLP synthase showing the organization of the 12 Pdx1 (blue) and Pdx2 (orange) subunits. One of the 12 heterodimers is highlighted using the subunit coloring shown in **Figure 20**. Image rendered in PYMOL.

other Class I amidotransferases,[67] the coordination of glutaminase and synthase activities in this enzyme appears to be dependent on the presence of glutamine rather than the nitrogen acceptor. This almost certainly is a consequence of the importance of glutamine for the association of Pdx1 and Pdx2 to form a complex given that the binding interaction between the subunits from *P. falciparum* and *B. subtilis* is enhanced 30- and 25-fold, respectively, in the presence of this substrate.[228,229] Experimental work on the PLP synthase from *Arabidopsis thaliana*[220] has provided some support for ammonia translocation along the putative tunnel seen in the bacterial forms of the enzyme. The glutaminase and synthase activities are coordinated much more effectively in the plant enzyme compared to the bacterial PLP synthases. Site-specific mutation of a leucine residue to alanine in the plant Pdx1 subunit resulted in a fourfold reduction of coupling efficiency in the resulting PLP synthase mutant.[220] Sequence alignments suggest that this leucine residue occupies the same spatial location as a methionine residue in the putative ammonia tunnel in the bacterial Pdx1 subunit (**Figure 23**).

8.06.4   Class II Amidotransferases

### 8.06.4.1   Glutamine 5′-Phosphoribosylpyrophosphate Amidotransferase

8.06.4.1.1   *Catalytic mechanism*

The synthesis of PRA is the first, committed step of *de novo* purine biosynthesis, and is catalyzed by the enzyme GPATase.[230] This catalytic activity is therefore highly regulated, both at the gene level and by the binding of end products in the purine biosynthetic pathway.[45,231,232] It has been proposed that GPATase forms a transient complex to the second enzyme in the pathway, glycinamide ribonucleotide (GAR) synthetase,[4,233] in order to transfer PRA, which is unstable in aqueous solution and has a half-life of under 5 s at physiological pH and room temperature.[234] Evidence for such an interaction has been provided by experiments using the GPATase and GAR synthetase from
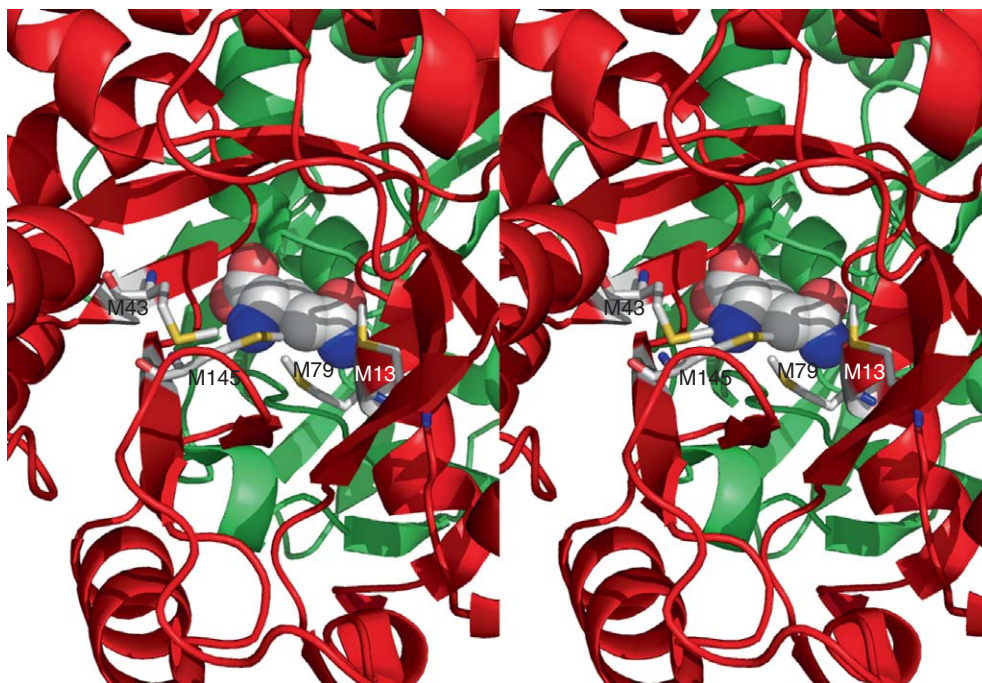
Figure 22　Stereoview showing the methionine side chains (Met13, Met43, Met79, and Met145) that line the ammonia tunnel in the interior of the Pdx1 $(\beta/\alpha)_8$ barrel (red). The glutamine ligand bound in the Pdx2 subunit (green) is rendered as a CPK model. Coloring: C – gray, O – red, N – blue, and S – yellow. Image rendered in PYMOL.
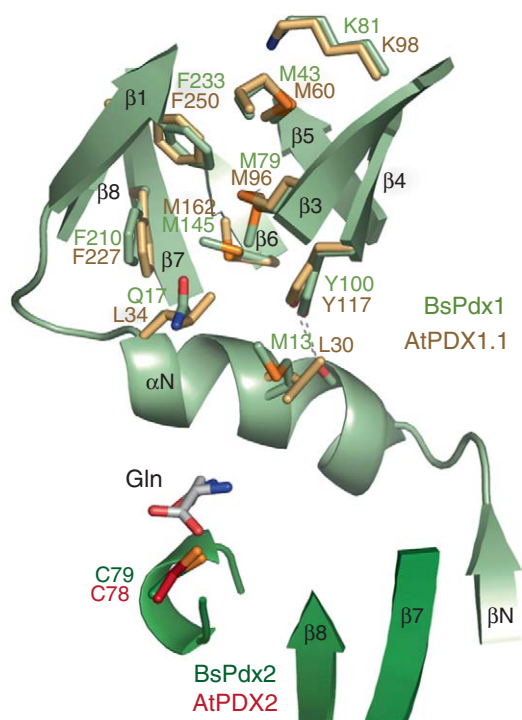
Figure 23　Cartoon representation of PLP synthase containing the putative ammonia tunnel showing residue differences between the enzymes from *Bacillus subtilis* (green) and *Arabidopsis thaliana* (gold). Side chains defining the tunnel are shown as 'sticks'. Reproduced from M. Tambasco-Studart; I. Tews; N. Amrhein; T. B. Fitzpatrick, *Plant Physiol.* **2007**, *144*, 915–925, with permission from the American Society of Plant Biologists.
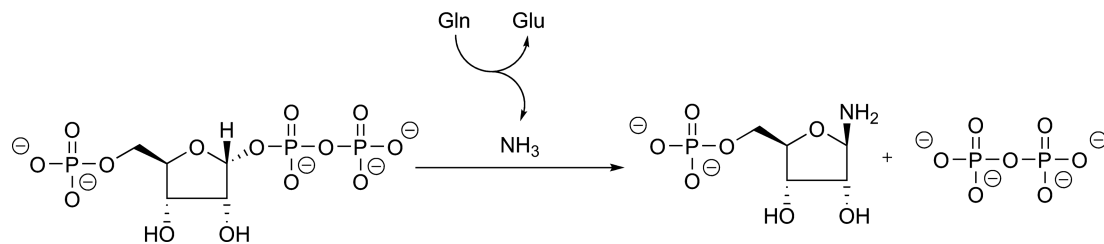
Scheme 23

the hyperthermophile *Aquifex aeolicus*.[235] The best characterized forms of GPATase are those present in *E. coli* and *B. subtilis*, in which both can accept ammonia as an alternate nitrogen source in place of glutamine. For (as yet) unresolved reasons, the *Bacillus* enzyme contains a [4Fe–4S] cluster, rendering the protein sensitive to aerobic conditions.[236,237] The *in vivo* role of this cluster does not appear to be related to catalysis, although it has been suggested that this moiety might function as an oxygen sensor.[238] In the absence of detailed kinetic and mechanistic studies of GPATase, the reaction catalyzed by the enzyme likely proceeds via direct displacement of the pyrophosphate (PP$_i$) leaving group by an ammonia molecule that is generated in the glutaminase site of the enzyme (**Scheme 23**).[239] To date, no residues have been identified in the synthase site that appear to play a role in catalyzing this reaction, a finding that has been rationalized by the excellent leaving group behavior of PP$_i$ when coordinated with metals such as Mg^{2+}, and the nucleophilicity of free ammonia. The role of the enzyme therefore seems to involve decreasing the entropic component of the activation energy by organizing the substrates for reaction. Although detailed kinetic measurements that confirm such a hypothesis do not appear to have been reported, indirect evidence for this proposal is provided by the lack of invariant active site residues in Type I phosphoribosyltransferases (PRTases).[240,241]

8.06.4.1.2   *Enzyme structure*

High-resolution X-ray crystal structures have been reported for the GPATases present in *E. coli*,[79,242–244] and *B. subtilis*,[233,245] which show that the two enzymes are composed of two domains joined by a flexible peptide segment (**Figure 24**). As anticipated on the basis of sequence analysis,[67] the glutaminase active site is located in the N-terminal domain. PRPP binding and activation for nucleophilic attack takes place in the C-terminal domain, which has a similar fold to that seen in members of the Type I PRTase superfamily,[240,241] including orotate phosphoribosyltransferase,[246] uracil PRTase,[247] and hypoxanthine–guanine phosphoribosyltransferase.[248] This domain also contains the sites through which purine nucleotides exert their allosteric regulation of enzyme activity. In the crystal, *B. subtilis* GPATase is a tetramer (dimer of homodimers),[232,245] although it appear that the *E. coli* enzyme can exist both as the tetramer and the homodimer.[242,243] The availability of several structures for the *E. coli* GPATase complexed to a variety of substrate analogs have shown that a key loop region (residues 326–349) possesses a remarkable conformational flexibility (**Figure 25**).[79,242] In the both the apo- and DON-modified forms of *E. coli* GPATase, this loop segment is conformationally disordered with the result that both the glutaminase and synthase sites are solvent accessible. When a nonhydrolyzable PRPP analog (cPRPP),[249] which is a competitive inhibitor with respect to PRPP,[250] is also present in the synthase site of the DON-modified enzyme; however, the loop residues adopt an α-helical secondary structure that not only prevents solvent access into the protein interior but also creates an intramolecular tunnel, of ~20 Å, linking the two active sites.[242] Thus, unlike the ammonia tunnels that are observed in several Class I amidotransferases,[11–13] this structural element is not maintained throughout the entire catalytic cycle of GPATase, and its formation is controlled by PRPP binding. The tunnel itself is constructed predominantly of hydrophobic residues, many of which are invariant (Phe259, Tyr258, Phe334, Ile335, Ile369, Val370, and Leu415) or conservatively substituted (Leu253, Phe254, Val257, Ala396, Ile399, and Phe461). There are also a few polar side chains that can interact with ammonia as it is translocated between the active sites (Tyr258, Thr333, Ser368, and Thr374). It was therefore proposed that water and ammonium cation are excluded from entering the channel.[242] In addition to creating the ammonia tunnel, cPRPP (and presumably PRPP) binding exerts a more global conformational effect that permits glutamine to enter the glutaminase active site. Tryptophan fluorescence studies on a series of GPATase triple mutants have provided evidence for the occurrence of these conformational changes in
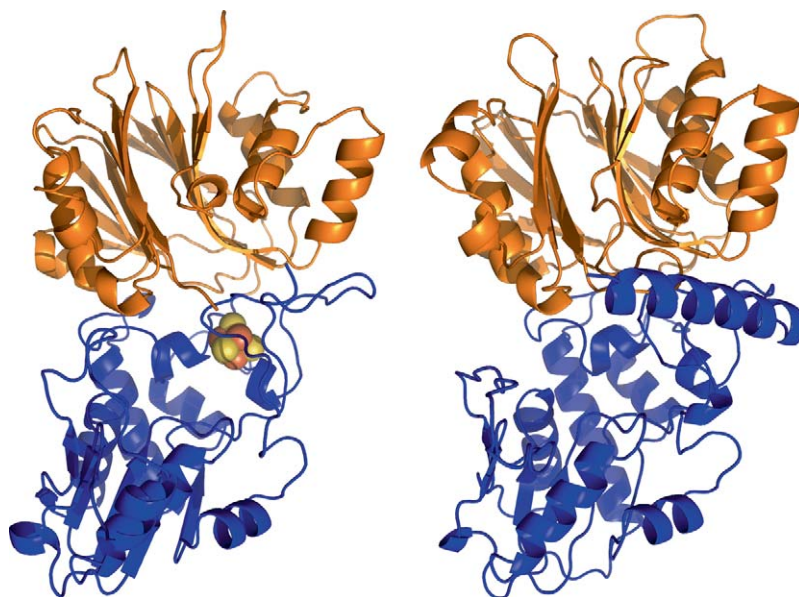
Figure 24 (Left) Cartoon representation of the *Bacillus subtilis* GPATase monomer (1AO0) showing the domain structure and the location of the [4Fe–4S] cluster (spheres). Coloring: S – yellow and Fe – red. (Right) Cartoon representation of the *Escherichia coli* GPATase monomer (1ECF). The glutaminase and synthase domains in both images are colored orange and blue, respectively. Image rendered in PYMOL.
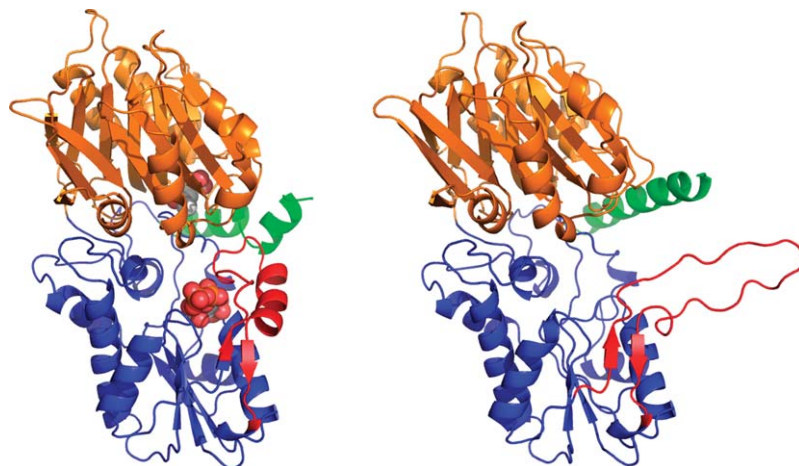
Figure 25 (Left) Cartoon representation of the *Escherichia coli* GPATase/DON/cPRPP complex (1ECC) showing the enzyme monomer in its 'active' conformation. The covalently modified Cys1 and cPRPP are rendered as CPK models. Coloring: C – gray, O – red, N – blue, and S – yellow. (Right) Cartoon representation of unbound *E. coli* GPATase (1ECF) adopting the 'inactive' conformation of the enzyme. In both images, the glutaminase and synthase domains are colored orange and blue, respectively. Peptide segments that undergo conformational changes on ligand binding are colored red (residue 325–354) and green (residues 471–492). Image rendered in PYMOL.

solution,[251] and suggest the existence of three functional states comprising (1) the ligand-free enzyme in which the synthase site is open and the glutaminase site is closed, (2) the PRPP complex in which the synthase site is closed and the glutaminase site open, and (3) the ternary glutamine/PRPP/GPATase complex that contains the tunnel and a restructured loop segment in the glutaminase site.

8.06.4.1.3   Investigating intramolecular ammonia translocation

Evidence to support the hypothesis that the catalytic mechanism of the enzyme includes ammonia translocation through the tunnel linking the two active sites has come almost completely from studies of the stoichiometry seen for the glutamine-dependent synthase reaction under steady-state conditions.[252] Such an approach to estimating the efficiency of ammonia transfer is possible for GPATase because the enzyme has negligible glutaminase activity in the absence of PRPP.[79,253] As a result, glutamate and PRA are formed with an almost exact 1:1 stoichiometry in the GPATase-catalyzed reaction. The effect on ammonia transfer efficiency (as measured by the glutamate:PRA ratio) of mutating a series of invariant, hydrophobic residues (Phe254, Tyr258, Phe259, and Phe334) located in the tunnel was therefore undertaken. This series of GPATase variants exhibited substantially reduced levels of ammonia transfer, ranging from 0.009 to 0.91 of that observed for the wild-type enzyme (**Table 2**), suggesting that the conformational change leading to tunnel formation had been sufficiently disrupted to permit leakage of ammonia into bulk solution. One especially interesting GPATase variant was that for which Leu415, a residue that lies at a relatively large distance from the putative intramolecular tunnel (**Figure 26**), was replaced by alanine (L415A). The L415A GPATase mutant exhibited rate increases for both the glutaminase and ammonia-dependent PRA synthase activities of 1.8- and 1.3-fold, respectively, showing that this modification had no significant effect on either of the active site structures. The efficiency of ammonia transfer from glutamine to form PRA in the L415A variant, however, was only about 15% that of the wild-type GPATase. It is likely that the replacement of Leu451 by alanine results in a defective tunnel structure in the active conformation of the enzyme thereby permitting ammonia to leak from the enzyme.[252] The removal of the large hydrophobic side chain may also affect dynamical properties of the enzyme that are important for catalysis.[254–256]

8.06.4.1.4   Coordinating the catalytic activities of GPATase

As for many other glutamine-dependent amidotransferases, the glutaminase and synthetase activities of GPATase are tightly coupled. For example, the K_M for glutamine decreases by two orders of magnitude when PRPP is present in solution.[253] As outlined above, X-ray crystallography[250–252] and tryptophan fluorescence measurements[253,257] have provided considerable evidence that residues in the glutaminase site are incorrectly positioned for catalysis until conformational ordering of a loop comprising residues 326–350, which is triggered by PRPP binding to the synthase active site. Because this loop ordering also forms one wall of the intramolecular tunnel, interdomain signaling is highly correlated with ammonia translocation. The interaction between residues Tyr74 and Ile335, located in the N- and C-terminal domains, respectively, in the active form of the enzyme appears to play a critical role in organizing residue side chains in the glutaminase site for catalysis (**Figure 27**).[242] Thus, when this interaction is present, the side chain of Arg73 is repositioned so as to make a key salt bridge with the carboxylate of bound glutamine. Replacing Ile335 by valine gave a GPATase variant that bound glutamine 30–60-fold less

Table 2   Effect of mutating *Escherichia coli* GPATase tunnel residues on the efficiency of ammonia translocation and glutamine binding

Enzyme	Glutaminase activity (U mg^{-1})	Translocation efficiency[a]	Glutamine K_D (mmol l^{-1})
WT	60 ± 7	1.0	0.25
Y74F	54 ± 6	0.45	7.25
F254V	48 ± 10	0.68	9
F254Y	35 ± 4	0.37	1
Y258F	27 ± 1	0.08	1.5
F259V	58 ± 4	0.63	0.25
T333A	30 ± 5	0.61	1.25
F334A	82 ± 7	0.009	10
I335V	54 ± 8	0.91	30.5
P337A	65 ± 9	0.77	0.63
K339A	51 ± 2	0.50	2.75

[a] These values are calculated from the stoichiometry of glutamate to 5′-phosphoribosylamine, and therefore represent the fraction of ammonia released by glutamine hydrolysis that is used in the synthase reaction. All values for the mutant enzymes are normalized to that of WT GPATase (0.90).
WT, wild type.
Data adapted from A. K. Bera; J. L. Smith; H. Zalkin; *J. Biol. Chem.* **2000**, *275*, 7975–7979.
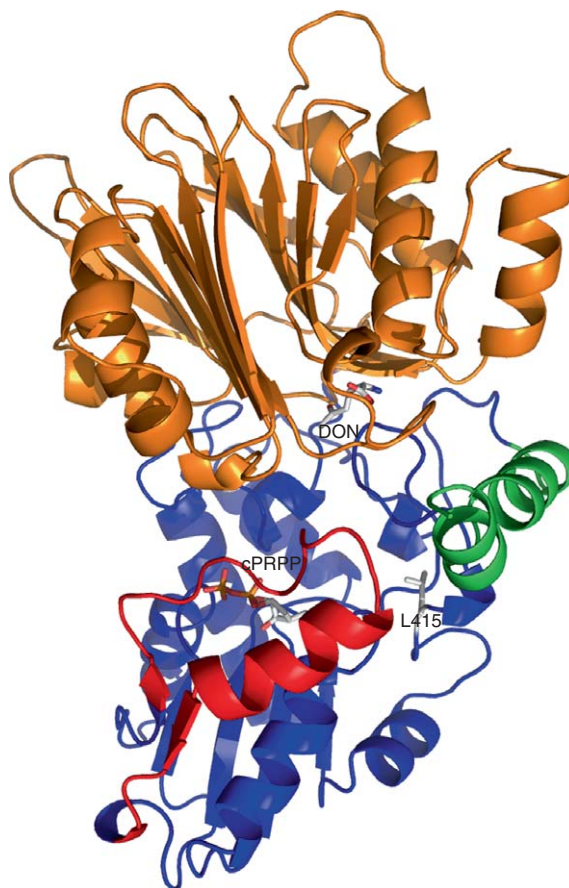
Figure 26   Cartoon representation of the *Escherichia coli* GPATase/DON/cPRPP complex (1ECC) showing the location of the Leu415 side chain relative to the two active sites. The glutaminase and synthase domains are colored orange and blue, respectively, and contiguous peptide segments that undergo conformational changes on ligand binding are colored red (residues 325–354) and green (residues 471–492). DON, cPRPP, and the Leu415 side chain are shown as sticks. Coloring: C–gray, O – red, N – blue, and S – yellow. Image rendered in PYMOL.

tightly than the wild-type enzyme, but which retained a similar affinity for PRPP.[257] The importance of the hydrophobic side chain at this position was further emphasized by the kinetic behavior of the I335A GPATase mutant for which glutamine- and ammonia-dependent synthase activity were less than 3% of those of the wild-type enzyme. Similarly, I335A GPATase mutant exhibited a nearly 50-fold decrease in glutaminase activity. When Tyr74 was replaced by phenylalanine, the dissociation constant for glutamine in the presence of PRPP was increased 29-fold although glutamine-dependent synthase activity was relatively unchanged when compared to wild-type GPATase (**Table 2**). This observation suggests that (1) the tunnel structure is unaffected in this mutant and (2) only interdomain signaling is significantly perturbed by removal of the hydrogen bonding group. When the aromatic side chain was removed, however, as in the Y74A GPATase mutant, glutamine-dependent PRA production was almost eliminated even though PRPP binding was unaffected. Although it is possible that this mutation caused structural defects in the ammonia tunnel, the glutaminase activity of the Y74A GPATase mutant was extremely low due to a severely impaired ability to bind glutamine.

## 8.06.4.2   Glutamine Fructose-6-Phosphate Amidotransferase

8.06.4.2.1   *Catalytic mechanism*

Glucosamine-containing polymers, which are formed from precursors such as uridine 5′-diphospho-*N*-acetyl-D-glucosamine (UDP-GlcNAc),[258] play key roles in cellular metabolism and intercellular signaling.[259,260] For
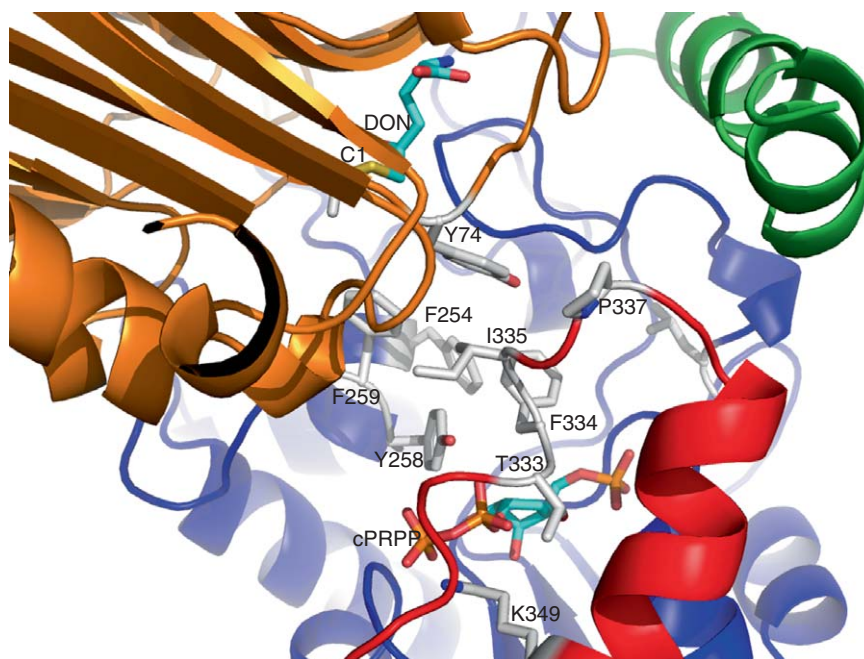
Figure 27 View of the intramolecular tunnel in the *Escherichia coli* GPATase/DON/cPRPP complex (1ECC) showing the proximity of the Tyr74 and Ile335 side chains. Protein backbone coloring is identical to that in **Figure 25**. DON, cPRPP, and selected residue side chains are shown as sticks. Coloring: C – gray, O – red, N – blue, and S – yellow. Image rendered in PYMOL.

example, GlcNAc is found in a variety of compounds, including bacterial peptidoglycan,[261–263] chitin, and mucopolysaccharides.[264] The first committed step in the *de novo* biosynthesis of glucosamine-containing compounds is mediated by GFAT,[50] which catalyzes the addition of glutamine-derived ammonia to fructose-6-phosphate to give glucosamine-6-phosphate after an enzyme-catalyzed isomerization reaction (**Scheme 24**).[78,265] The enzyme exhibits very low glutaminase activity in the absence of fructose-6-phosphate, and in sharp contrast to other Class II amidotransferases,[67] GFAT cannot use free ammonia as an alternate nitrogen source.[50,78] Recent studies have also implicated GFAT in the regulation of human hexose metabolism,[266–268] suggesting that GFAT inhibitors, in addition to being potential antibiotics and/or antifungal agents,[269–272] may have clinical utility in the treatment of noninsulin-dependent (type 2) diabetes.[264,273,274] The best characterized forms of GFAT are those present in *E. coli* (on which almost all the detailed mechanistic studies have been performed),[50,84,274–278] *Candida albicans*,[279–281] and *Homo sapiens*.[282–285] The bacterial enzyme is a single polypeptide, which is active as a homodimer.[50] Partial digestion of *E. coli* GFAT with trypsin gives two catalytically active fragments,[286] which were used (1) to provide the first insights into enzyme structure[287,288] and (2) to assign the location of the synthase active site in the C-terminal domain.[274,286] A series of
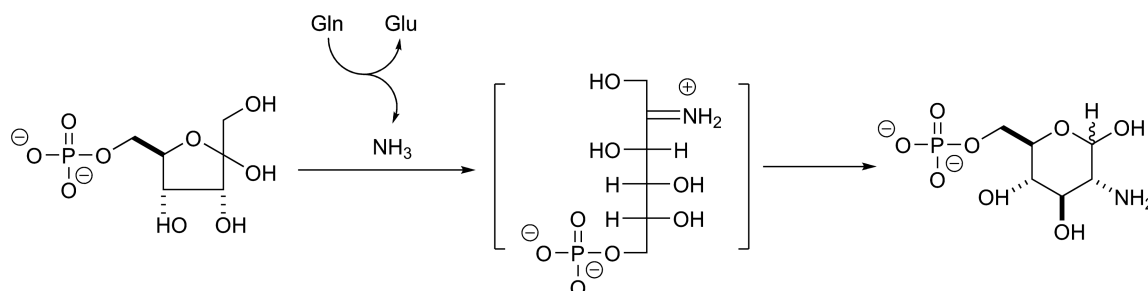

Scheme 24
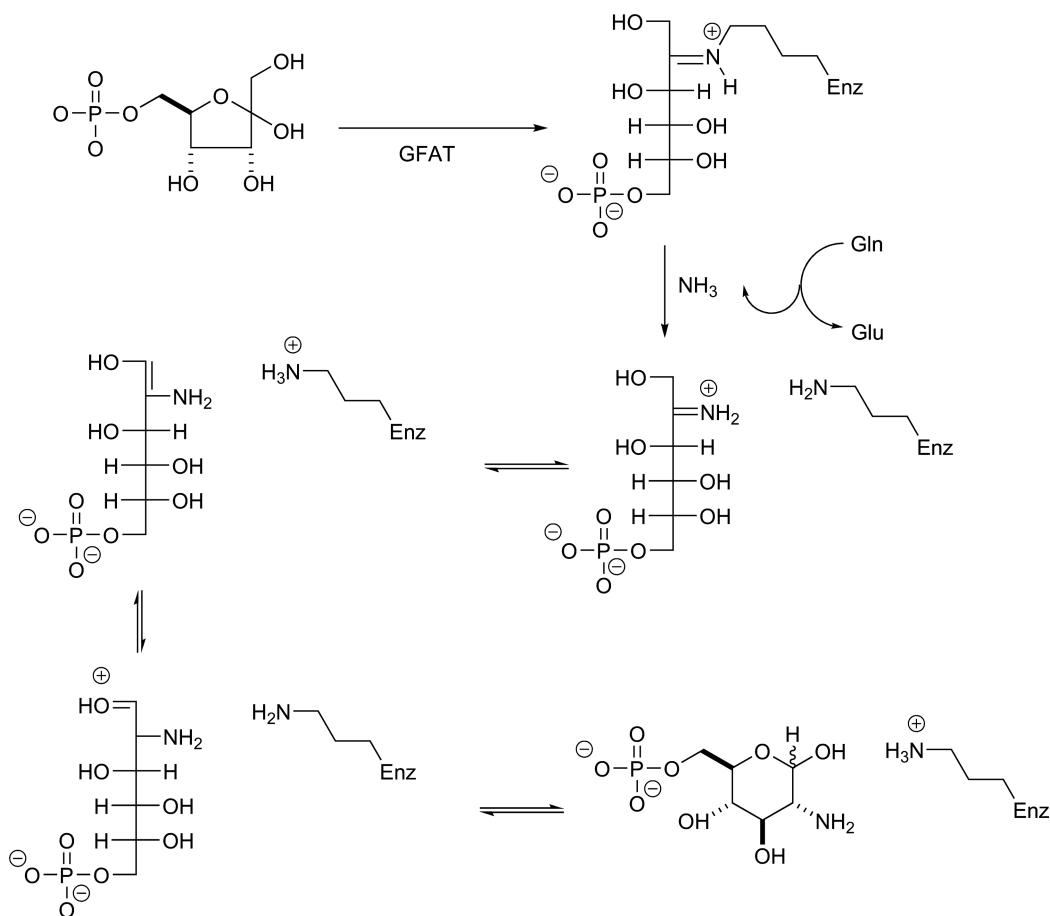
Scheme 25

kinetic studies on both the wild-type GFAT and a series of site-specific mutants have suggested that fructose-6-phosphate likely binds to the synthase site in its furanose form, and undergoes enzyme-catalyzed ring opening to generate a ketone moiety that subsequently reacts with the side chain of Lys603 (**Scheme 25**).[289] Transimination of this Schiff base intermediate with ammonia, released in the glutaminase site of the enzyme, then yields an imine that can be isomerized to glucosamine-6-phosphate by a series of proton transfer steps. In *B. subtilis*, GFAT is regulated at the transcriptional level by a riboswitch that binds to glucosamine-6-phosphate.[290,291] Thus, like other riboswitches that control gene expression,[292] high concentrations of glucosamine-6-phosphate facilitate the formation of the active complex, which then downregulates gene transcription and the cellular levels of the enzyme in the bacterium. The eukaryotic GFAT is larger than its prokaryotic counterpart, and, although it likely uses an identical catalytic mechanism, is inhibited by UDP-GlcNAc,[283] the final product of the reaction pathway in which GFAT catalyzes the first step. Recent work has also shown that phosphorylation/dephosphorylation is used as a mechanism of regulating activity in eukaryotic forms of the enzyme.[280,293–296]

8.06.4.2.2   Enzyme structure

Building on early structural studies of domain fragments obtained by limited proteolysis of the *E. coli* GFAT,[287,288] a series of high-resolution X-ray crystal structures have been obtained for this enzyme and various complexes corresponding to putative intermediates in the catalytic cycle (**Figure 28**).[80,297–299] The first demonstration that there was an ammonia tunnel linking the glutaminase and synthase sites came from a relatively low-resolution (3 Å) structure of the intact enzyme from *E. coli* bound to fructose-6-phosphate
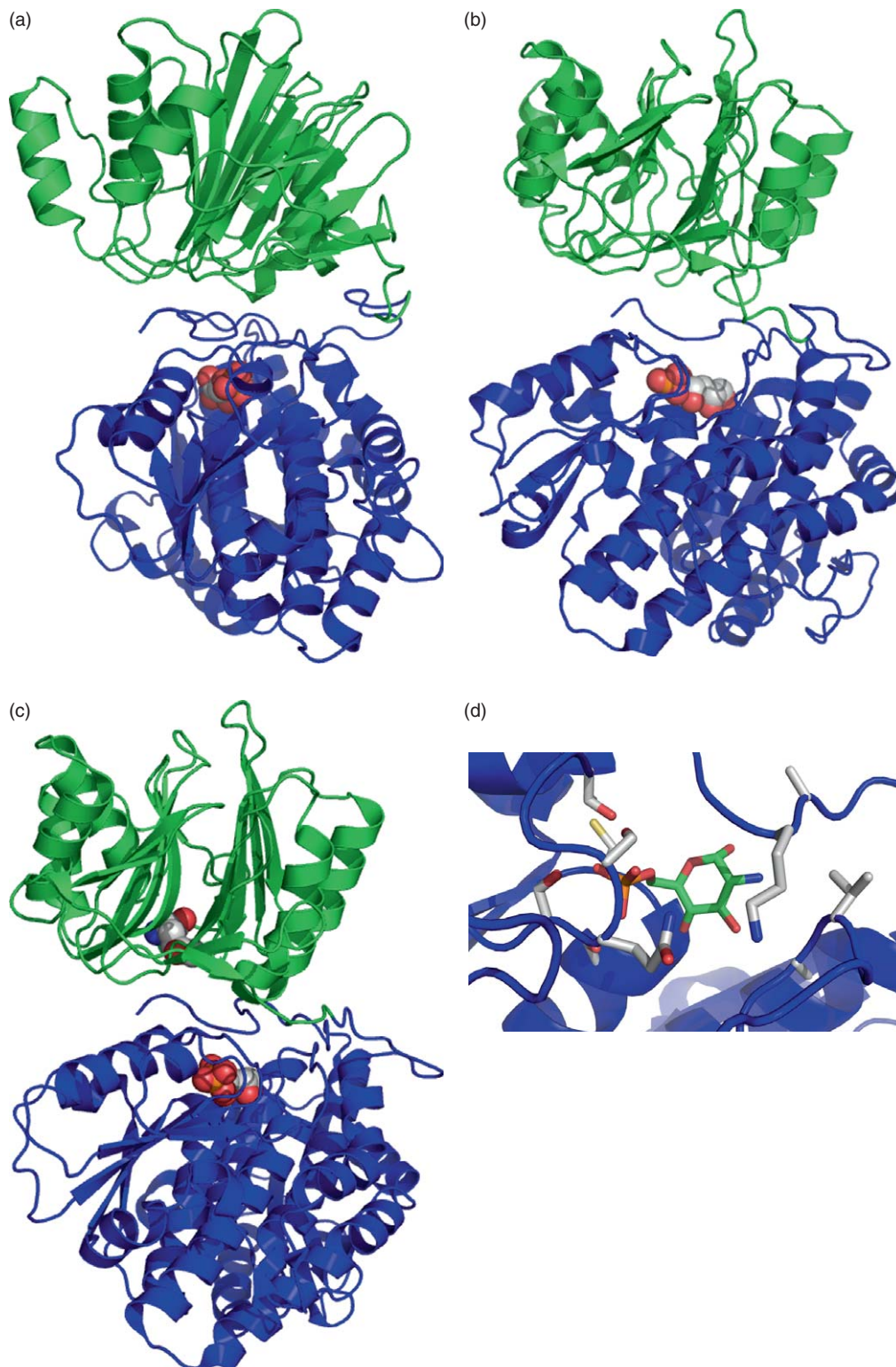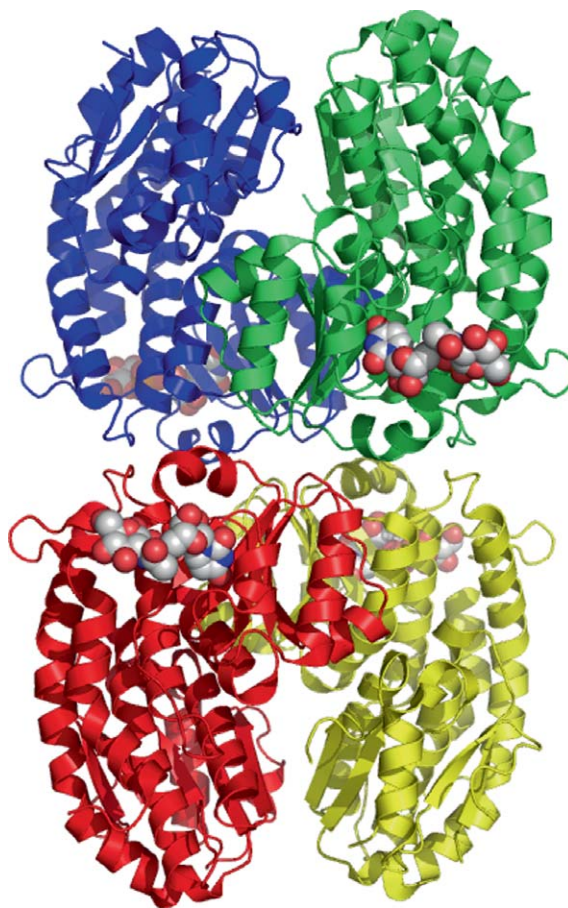
Figure 28 Cartoon representation of *Escherichia coli* GFAT complexed to (a) fructose-6-phosphate (2BPL) and (b) glucose-6-phosphate (1JXA). (c) DON-modified *E. coli* GFAT complexed to glucose-6-phosphate (2BPJ). In all three images, the glutaminase and synthase domains are colored green and blue, respectively, and bound ligands are shown as CPK models. (d) Close-up view of glucosamine-6-phosphate bound in the synthase active site of *E. coli* GFAT (2VF5). The ligand and nearby active site residues are rendered as sticks. Coloring: C – green (Glc-6-P) and gray (protein), O – red, N – blue, P – orange, and S – yellow. Image rendered in PYMOL.

(**Figure 28(a)**).[297] The complex crystallizes as a dimer in which all protein/protein contacts are formed between the C-terminal, isomerase domains and the two domains in each monomer are linked by a flexible segment (residues 240–247). The bacterial enzyme has been shown to exhibit an ordered Bi–Bi kinetic mechanism in which fructose-6-phosphate binds first followed by glutamine.[50] Hence, this structure corresponds to the first Michaelis complex in the catalytic mechanism, and the glutaminase site is exposed to solvent through a conformational rearrangement in a loop that is observed to 'close down' on the active site when glutamine or glutamine analogs are bound within this domain.[287] Given its location relative to the bound substrate, His504 in the isomerase/synthase domain was assigned to be the general acid/base catalyst needed for the ring opening/closure and proton transfer steps. Conserved residues from both the glutaminase and synthase domains were found to be involved in defining the 18 Å tunnel, and, as observed in GPATase, these were predominantly hydrophobic. Thus, the central portion of the channel comprises the C-terminal tail (Leu601 to Val607), with Leu601, Ala602, Val603, and Val607 forming a hydrophobic entrance to the isomerase/synthase site (**Figure 29**). Conformational flexibility in the C-terminal tail also appears to control the access of fructose-6-phosphate, implying that tunnel formation is initiated by substrate binding as observed for GPATase. Arg26 and Trp74 in the glutaminase domain of the bacterial enzyme define the N-terminal end of the tunnel, with loop segments from the C-terminal domain of the other GFAT monomer also participating in defining the tunnel (**Figure 29**). In an interesting finding, access to the tunnel was blocked by the side chain of Trp74 and it was therefore suggested that glutamine binding gave rise to a form of the enzyme in which intramolecular ammonia transfer could take place. This suggestion was confirmed by the subsequent determination of a structure for the *E. coli* enzyme bound to glucose-6-phosphate in which the side chain of Cys1 was covalently modified by treatment with DON (**Figure 28(c)**). This revealed that changes in the relative orientation of the two domains caused the tunnel to become inaccessible to bulk solvent. In addition, the side chains of Trp74 and Asn98 underwent conformational changes to open the ammonia tunnel and form an oxyanion-binding site, respectively, thereby rendering the glutaminase site catalytically active.[80] These conformational changes have been compared to, and contrasted with, those that give rise to tunnels in other amidotransferases.[10] Finally, recent work has also yielded a structure for the *E. coli* enzyme, both as the free enzyme and complexed to glucosamine-6-phosphate (**Figure 28(d)**). The latter structure corresponds to the
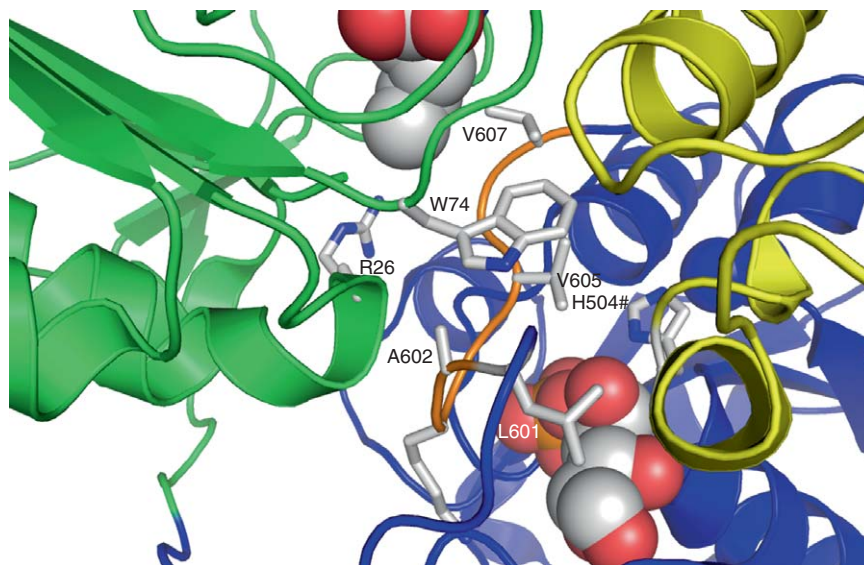
Figure 29   The intramolecular tunnel in the *Escherichia coli* GFAT dimer (2BPJ). The glutaminase and synthase domains of monomer A are colored green and blue, respectively, and the C-terminal domain of monomer B is yellow. The C-terminal tail of monomer A is orange. Residues Arg26, Trp74 (defining the N-terminal tunnel end in A), Leu601, Ala602, Val603, Val607 (defining the entrance to the synthase site in A), and His504 (from monomer B) are shown as sticks. Bound DON and glucose-6-phosphate are rendered as CPK models. Coloring: C – gray, O – red, N – blue, and S – yellow. Image rendered in PYMOL.

Figure 30 Cartoon representation of the isomerase/synthase domain of *Candida albicans* GFAT complexed to UDP-GlcNAc (2PUT). Each monomer in the homotetramer is colored differently. UDP-GlcNAc is shown as a CPK model. Coloring: C – gray, O – red, and N – blue. Image rendered in PYMOL.

last product complex formed during catalytic turnover.[299] Once again, these reveal the subtle conformational changes in the protein structure that permit product release and tunnel formation.[298]

Given the medical importance of the human and fungal forms of GFAT, and the interesting differences in their regulation, considerable efforts have been made to obtain high-resolution crystal structures for both of these enzymes. The recent success in solving the structure of the isomerase/synthase domain of the GFAT from *C. albicans* complexed to UDP-GlcNAc is therefore an important development (**Figure 30**).[300,301] The domain crystallizes as a tetramer (dimer of homodimers) and exhibits a number of differences in the relative orientation of subunits when compared to the cognate structure for the isomerase domain of the *E. coli* enzyme.[287]

8.06.4.2.3 Investigating intramolecular ammonia translocation

GFAT is unusual among amidotransferases in being unable to use ammonia as an alternate nitrogen source under *in vitro* conditions,[50] precluding any direct assessment of the ability of glutamine to suppress ammonia-dependent synthase activity. The process of ammonia translocation has therefore been studied by steered MD simulations on a model of the active enzyme built from the crystal structure of the DON-modified form of *E. coli* GFAT complexed with fructose-6-phosphate.[302] Hence, the DON residue was replaced by glutamate and a molecule of ammonia, and the pyranose form of glucose-6-phosphate was modeled into the site occupied by the linear form of this molecule present in the original crystal structure.[80] The steered MD simulations followed similar computational protocols to those used in studies of IGP synthase, and confirmed the importance of the Trp74 side chain dynamics in controlling the passage of ammonia through the tunnel.

Table 3   Effect of mutating critical tunnel residues on the steady-state glutaminase and synthase activities of *Escherichia coli* GFAT[a]

Enzyme	Glutaminase			Synthase		
	k_{cat} $(min^{-1})^b$	K_M $(\mu mol\,l^{-1})^b$	k_{cat}/K_M $(mmol^{-1}\,l\,s^{-1})^b$	k_{cat} $(min^{-1})^c$	K_M $(\mu mol\,l^{-1})^c$	k_{cat}/K_M $(mmol^{-1}\,l\,s^{-1})^c$
WT	1030 ± 5	270 ± 20	63 600	865 ± 1	360 ± 20	40 000
W74A	181 ± 5	380 ± 50	7900	7.00 ± 0.01	930 ± 40	125
W74L	93 ± 3	40 ± 6	38 700	5 ± 1	1100 ± 200	75
W74F	430 ± 5	48 ± 3	148 600	55 ± 1	1450 ± 60	630
A602L	23.0 ± 0.5	100 ± 20	3833	13.0 ± 0.2	1150 ± 50	190
V605L	17.0 ± 0.2	72 ± 6	3935	9.0 ± 0.1	670 ± 40	228

[a] The glutaminase activity is measured in the presence of fructose-6-phosphate, and so the ratios of k_{cat} for the glutaminase and synthase activities are a measure of the efficiency of ammonia translocation between the two active sites of the enzyme.
[b] The kinetic parameters are for glutamine.
[c] The kinetic parameters are for fructose-6-phosphate.
WT, wild type.
Data adapted from N. Floquet; S. Mouilleron; R. Daher; B. Maigret; B. Badet; M.-A. Badet-Denisot, *FEBS Lett.* **2007**, *581*, 2981–2987.

In addition, these simulations suggested that this side chain might play a role in excluding bulk water from passing through the tunnel in place of ammonia. The side chains of Ala602 and Val605, which are located in the conserved C-terminal tail of the enzyme,[287] also appeared to control the entry of ammonia into the C-terminal isomerase/synthase site, with the alanine residue moving in response to the presence of ammonia within the tunnel. These conclusions were probed by experimental studies of ammonia translocation efficiency in a series of site-specific *E. coli* GFAT mutants (**Table 3**).[302] Although none of these mutants were able to use exogenous ammonia as an alternate substrate for glutamine, the efficiency of ammonia transfer, at least as measured by the ratio of the turnover numbers for glutaminase-dependent glucosamine-6-phosphate formation and glutaminase activity in the absence of fructose-6-phosphate, was significantly decreased when Trp74 was replaced by either alanine or leucine. In the case of the W74A variant, MD simulations suggest that this kinetic behavior results from perforation of the tunnel linking the glutaminase and synthase sites in this enzyme. Modifications to the side chains of either Ala602 or Val605 had a smaller impact on ammonia transfer efficiency.

8.06.4.3   Glutamate Synthase

8.06.4.3.1   *Catalytic mechanism*

Glutamate synthase (GltS) is a complex iron–sulfur flavoprotein that plays a key role in the ammonia assimilation pathways found in bacteria and plants.[303] GltS catalyzes an interesting conversion in which nitrogen from the glutamine side chain is transferred to α-ketoglutarate (αkG) to form two molecules of glutamate after the acquisition of two electrons from a suitable reductant (**Scheme 26**).[304–306] Detailed information on the details of the kinetic and chemical mechanism of this complex enzyme has primarily been obtained through studies of the enzyme from the nitrogen-fixing bacterium *Azospirillum brasilense*.[307,308] This prokaryotic GltS is a heterodimer in which both the glutaminase and synthase sites are located in the α-subunit of the enzyme together with an FMN cofactor and one [3Fe–4S] cluster, which mediate electron
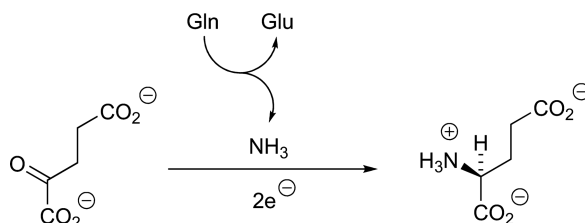
Scheme 26
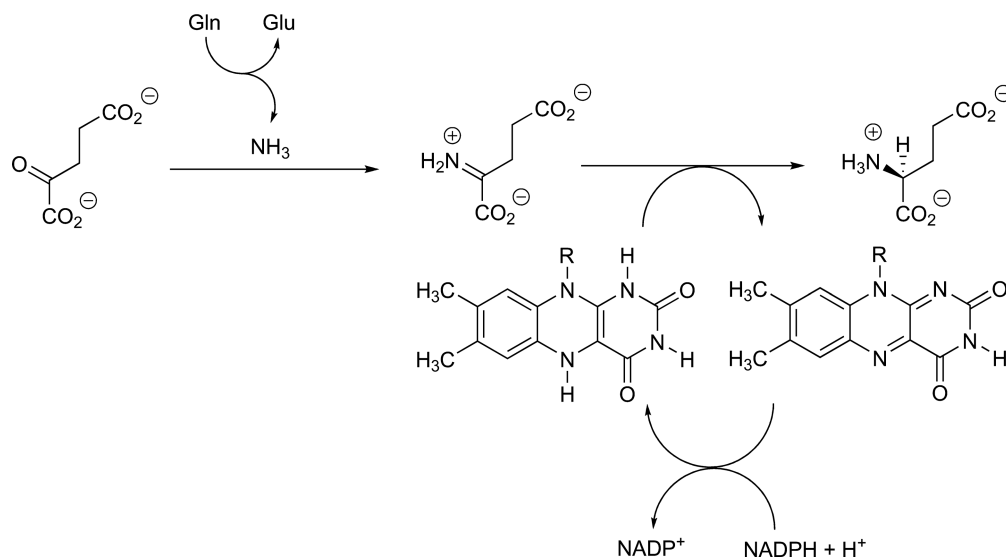
Scheme 27

transfer.[309] For bacteria other than *B. subtilis*, which uses NADH as the source of reducing power, the β-subunit of the prokaryotic enzyme contains two FAD cofactors, two [4Fe–4S] clusters that also function as structural elements mediating heterodimer formation,[310] and the binding site for NADPH.[311–313] The function of the β-subunit is to provide the electrons needed to reduce the imine intermediate that is formed by reaction of αkG with ammonia in the synthase site of the α-subunit (**Scheme 27**).[304] The plant variant of GltS,[303,314] which is evolutionarily related to the form of the enzyme present in cyanobacteria,[315] is a single polypeptide chain that contains an FMN cofactor and a [3Fe–4S] cluster.[316] For this enzyme, however, the source of reducing power is ferredoxin, which transfers electrons into the FMN cofactor by forming a transient complex with the α-subunit. Despite sharing considerable similarity in primary structure, the ferredoxin-dependent GltS (Fd-GltS) and the α-subunit of the NADPH-dependent bacterial enzyme exhibit remarkable differences in the synchronization of the reactions taking place at their glutaminase and synthase sites.[304] For example, the ability of Fd-GltS to catalyze glutamine hydrolysis is negligible in the absence of bound αkG, whereas the bacterial enzyme has significant glutaminase activity.[313] This difference in the strength of 'reaction coupling' is a consequence of modified molecular interactions between the two independent active sites of these GltS variants.

8.06.4.3.2   Enzyme structure

Experimental studies aimed at understanding how the complex series of chemical and electron transfer steps have been placed on a firm basis by the availability of X-ray crystal structures for the α-subunit of the NADPH-dependent GltS from *A. brasilense*,[317] and the Fd-dependent GltS from *Syncheocystis*,[318,319] as complexes with a variety of substrates and substrate analogs. This work has shown that the two proteins have very similar three-dimensional folds and domain organizations (**Figure 31**). Each protein is a single polypeptide composed of four domains, with the Ntn amidohydrolase domain being located at the N-terminus. A second domain with a β/α topology then connects this functional unit to the third domain, which is a $(\beta/\alpha)_8$ barrel resembling the HisF protein in IGP synthase, an evolutionarily unrelated Class I amidotransferase, and other FMN-binding proteins, including glycolate oxidase[320,321] and flavocytochrome b_2.[322] This domain contains both the αkG-binding site and the FMN and [3Fe–4S] redox centers. The fourth, C-terminal domain, composed of residues 1224–1523 has an unusual ellipsoidal shape, which is similar to that observed in *Erwinia carotovora* polygalacturonase,[323] and extends for a distance of 43 Å. The functional role of this domain may be to organize the relative orientation of the glutaminase and synthase domains, and several residues participate in forming the ammonia tunnel linking the two active sites, which are separated by ~31 and 24 Å in the two enzymes (**Figure 32**). The tunnel is better defined in the Fd-dependent GltS, however, because the side chains of Thr507, Asn508, Ser976, and Glu978 in the *Azospirillum* α-subunit divide the putative tunnel into two cavities (**Figure 33(Left)**). The first of these cavities is
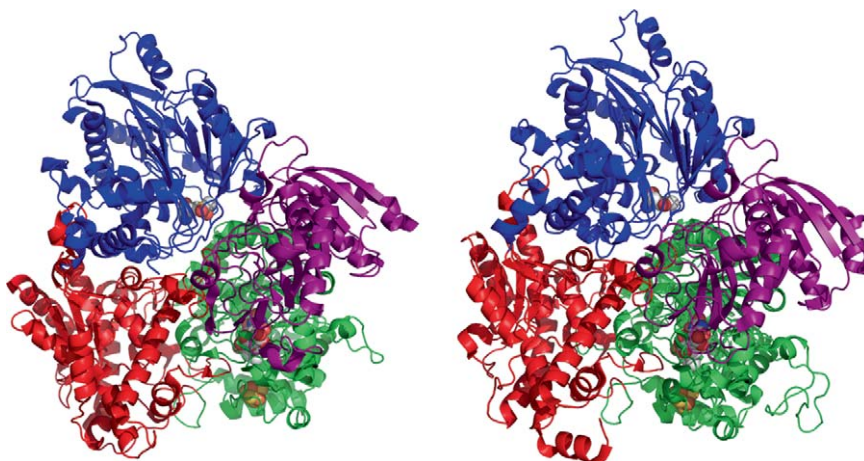
Figure 31   (Left) Cartoon representation of the α-subunit of the NADPH-dependent *Azospirillum brasilense* GltS (1EA0) complexed to α-ketoglutarate (αkG) and methioninesulfone. (Right) The DON-modified form of Fd-dependent *Synechocystis* GltS (1OFE) complexed to αkG. Both proteins are composed of glutaminase (blue), central (red), FMN-binding (green) and C-terminal, β-helical (purple) domains, and contain a [3Fe–4S] cluster and FMN cofactor (rendered as CPK models). Coloring: C – gray, O – red, N – blue, S – yellow, and Fe – red. Image rendered in PYMOL.
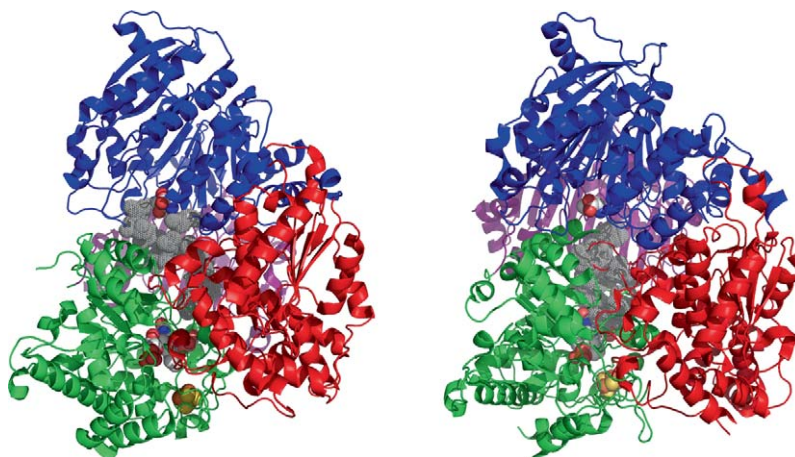

Figure 32   Cartoon representations of the α-subunit of the NADPH-dependent *Azospirillum brasilense* GltS (1EA0) complexed to α-ketoglutarate (αkG) and methioninesulfone (left) and DON-modified Fd-dependent *Synechocystis* GltS (1OFE) complexed to αkG (right) showing residues (gray dotted spheres) that define the tunnels linking the active sites. Domains and ligands are colored as described in **Figure 30**. Image rendered in PYMOL.

lined by the side chains of Tyr211, Ser212, Thr213, and Glu233 in the N-terminal domain, residues from the central linker (Val506, Thr507, and Asn508), and Ile977, which is located on a conformationally flexible loop (loop 4) in the FMN-binding domain. This part of the tunnel is therefore hydrophilic, in contrast to the hydrophobic nature of the tunnels observed in other glutamine-dependent amidotransferases.[11,12] The walls of the second cavity, which is proximal to the synthase site in the third domain, has a different chemical character, being lined only by carbonyl oxygens from the backbone amide moieties, and aliphatic side chains, including those of six proline residues that are conserved throughout all known GltSs (**Figure 33(Right)**). The tunnel observed in the *Syncheocystis* enzyme is structurally and chemically similar to that of the NADPH-dependent *Azospirillum* α-subunit, but contains a network of hydrogen-bonded water molecules, which become visible at the higher resolution (2.0– 2.45 Å) of the Fd-dependent GltS crystal structures (**Figure 34**).[306,318] These waters completely fill the cavity, however, raising questions about how ammonia can pass through the tunnel and access the synthase site.
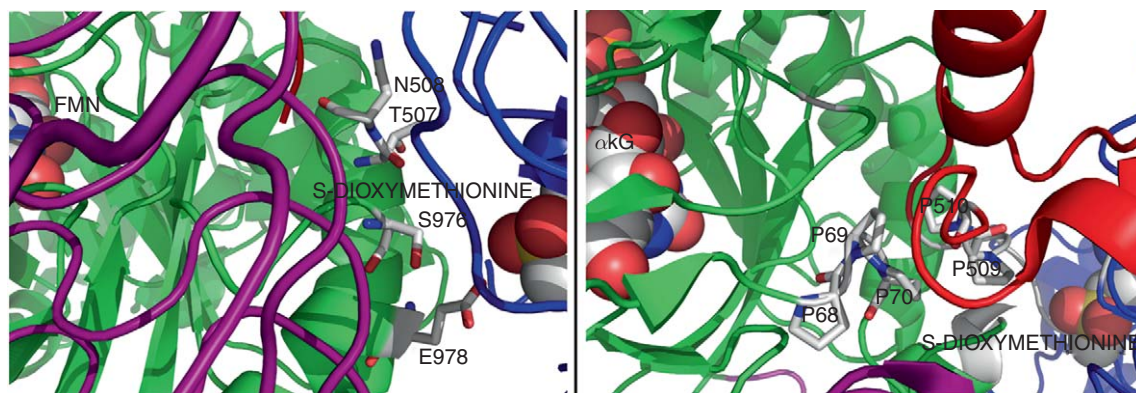
Figure 33 Close-up views of the α-subunit of NADPH-dependent *Azospirillum brasilense* GltS (1EA0) complexed to α-ketoglutarate and methioninesulfone. (Left) Residues Thr507, Asn508, Ser976, and Asp978, which divide the putative ammonia tunnel into two cavities. (Right) The six conserved proline residues that line the walls of the tunnel. Residue side chains are shown as sticks. Domains and ligands are colored as described in **Figure 30**. Image rendered in PYMOL.
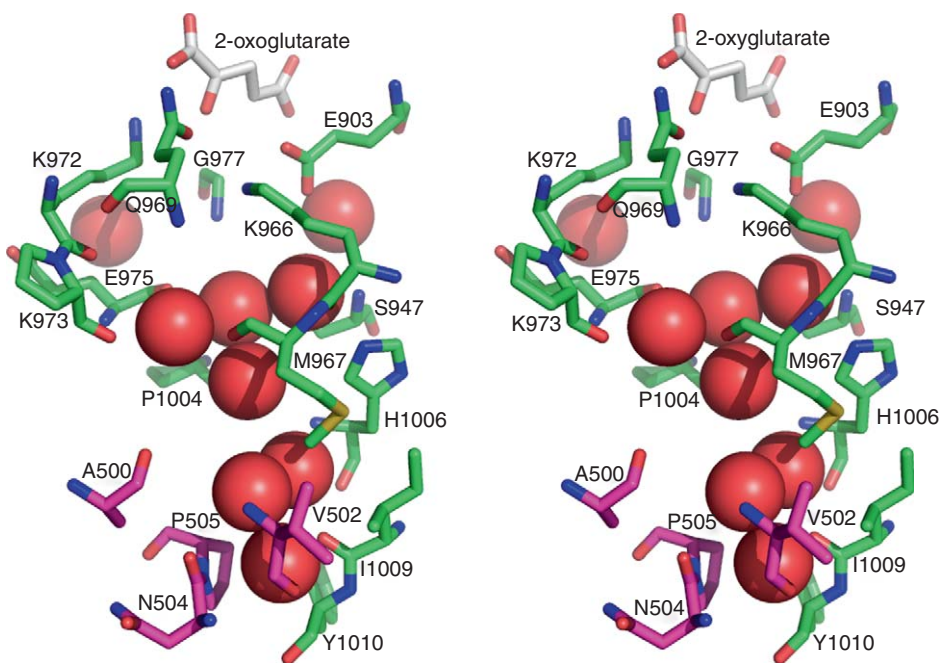
Figure 34 Stereoview of the residues defining the ammonia tunnel in Fd-dependent *Synechocystis* GltS (1OFE) and crystallographic waters (red spheres). Residue side chains are shown as sticks. Coloring: C – magenta (central domain residues) and green (FMN-binding domain residues), O – red, N – blue, P – orange, and S – yellow. Image rendered in PYMOL.

Unlike other Class II enzymes, such as GPATase or asparagine synthetase, in which reactive, phosphorylated intermediates must be prevented from undergoing hydrolysis, the presence of the water is unlikely to impact imine formation during catalytic turnover. Finally, the ammonia tunnel is present in all GltS structures obtained to date, including that of the *Syncheocystis* enzyme containing the reduced form of the FMN cofactor.[319] This again contrasts with the finding that substrate-dependent conformational changes are needed to form a well-defined, solvent-inaccessible ammonia tunnel in GPATase, GFAT, and many of the Class I amidotransferases.

It has not yet been possible to crystallize either the intact *Azospirillum* GltS αβ-heterodimer or the *Syncheocystis* GltS/ferredoxin complex. Studies using computational modeling,[319,324] small-angle X-ray scattering (SAXS) measurements,[324,325] cryo-electron microscopy (cryo-EM),[326] and electrospray ionization (ESI)

mass spectrometry[327] have provided some insight into the quaternary structure of GltS in solution and intersubunit electron transfer. For example, SAXS methods showed that the stoichiometry of the *Syncheocystis* GltS/ferredoxin complex was 1:1, suggesting that two ferredoxins bind consecutively during reduction of the FMN cofactor.[324] Computer-based docking was then used to obtain a model of the 1:1 complex from the crystal structures of DON-modified *Syncheocystis* GltS complexed with αkG and the *Syncheocystis* ferredoxin.[319] Building on the results of SAXS measurements of the intact *Azospirillum* GltS αβ-heterodimer,[324] cryo-EM methods have yielded a structure for the hexameric $(αβ)_6$ complex of the NADPH-dependent enzyme.[326] In combination with homology modeling, and previous biochemical observations,[310] this structure provides the first insights into how the C-terminal domain interacts with residues in the β-subunit to facilitate electron transfer between the two subunits.

8.06.4.3.3   *Coordinating the catalytic activities of GltS*

Although there has been little work aimed at delineating the functional role of residues in the GltS ammonia tunnel, structural predictions concerning the molecular mechanisms that coordinate catalysis in the two active sites and the role of redox state in modulating enzyme activity have been investigated. In the case of the NAPDH-dependent GltS αβ-heterodimer, no glutaminase activity is observed in the absence of αkG, and ^{15}N-labeling experiments have shown that ammonia transfer from glutamine takes place 100% efficiently, at least within the limits of the assay.[313] In the absence of the β-subunit, however, this tight coupling of catalytic activities is lost, suggesting that the redox state of the FMN and [3Fe–4S] cluster may also be important in the conformational changes required for controlling ammonia transfer. In order to obtain insights into such effects, MD simulations were performed for both the apo-form of the *Azospirillum* GltS α-subunit and its ternary complex with glutamine and αkG.[328] Loops that were not observed in the original crystal structures were built into the subunit models, which both contained the reduced form of the FMN cofactor. Trajectories (4 ns) computed for the fully solvated proteins were then analyzed for conformational motions, using an interesting algorithm,[329,330] that might impact tunnel structure and/or the organization of the catalytic residues in each active site. This analysis showed that the presence of ligands in both active sites did significantly impact the motional properties of the two systems,[330] and the relative orientations of the domains in the structure.[331] More importantly, substrate binding caused (1) a conformational change in a flexible loop (residues 263–271), which modulates access to the glutaminase active site and (2) changes in the diameter of the ammonia tunnel that allowed the entry of up to 16 water molecules during the course of the simulation of the ternary complex.

   The tight coupling of glutaminase and synthase activities is maintained for the Fd-dependent GltS in the absence of ferredoxin as an electron source.[316] Conformational changes seen in the crystal structures of different complexes of the *Syncheocystis* enzyme have therefore been analyzed to evaluate how catalytic activity in separate domains might be coordinated.[318] These experiments suggested that a strictly conserved loop (loop 4) composed of residues 968–1013 in the FMN-binding domain might play an essential role in mediating communication between bound ferredoxin, the αkG-binding site, and the N-terminal, glutaminase domain (**Figure 35**). Thus, residues (Lys972 and Arg992) in this loop form hydrogen bonds to bound αkG, and the side chain of Tyr987 interacts with the carboxylate of Asp907, which is part of the loop to which ferredoxin binds in the active heterodimer.[310] Perhaps most importantly, in the GltS/MetS/FMN complex, the Glu1013 side chain is hydrogen bonded to the N-terminal amino group of Cys1, which is thought to mediate general acid/ base catalysis in the glutaminase reaction,[67,310] and the side chain of Arg31. These interactions position Cys1 in an inactive orientation, leading to the proposal that the presence of αkG and cofactor reduction results in a conformational rearrangement of loop 4 that disrupts this hydrogen bond so that the Cys1 thiolate can react with the side chain amide of the glutamine substrate. A series of Fd-GltS mutants were therefore kinetically characterized in which Glu1013 was replaced by aspartate (E1013D), asparagine (E1013N), and alanine.[332] The effects of these mutations on glutaminase and glutamine-dependent synthase activity were significant, these activities being reduced by 2–4 orders of magnitude, even though the K_M of glutamine was almost unaffected. In addition, the tight coupling of glutaminase and synthase activities in the wild-type enzyme was only maintained in the E1013D mutant, with the rate of glutamine hydrolysis being twice that of glutamate synthesis in the E1013N and E1013A mutants. This observation therefore confirms the importance of the negatively charged side chain of Glu1013 for the coordination of glutaminase and synthase activities.
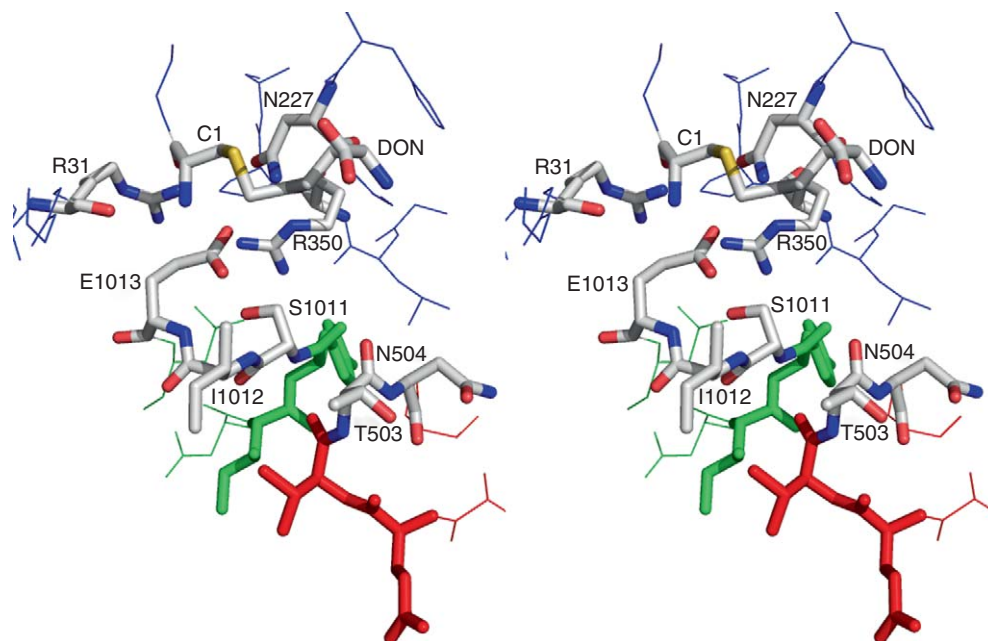
Figure 35 Stereoview of residues in 'loop 4' (Ser1011 and Ile1012) and the central domain (Thr503 and Asn504) of Fd-dependent *Synechocystis* GltS (1OFE), colored green and red, respectively, which obstruct the entrance to the putative ammonia tunnel. Nearby residues, DON and α-ketoglutarate are shown as sticks. Coloring: C – gray, O – red, N – blue, and S – yellow. Image rendered in PYMOL.

8.06.4.4   Asparagine Synthetase

8.06.4.4.1   Catalytic mechanism

Glutamine-dependent ASNS catalyzes the synthesis of asparagine from aspartic acid, ATP, and glutamine (**Scheme 28**).[47] Under *in vitro* conditions, ammonia, hydroxylamine, and hydrazine can also be used as alternate nitrogen sources.[333] ASNS is of clinical interest because it has been implicated in the cellular mechanisms that cause resistance to L-asparaginase[334] in clinical protocols for treating acute lymphoblastic leukemia,[335–337] and an ASNS inhibitor with nanomolar affinity for the enzyme has been shown to suppress the *in vitro* proliferation of a drug-resistant leukemia cell line.[338] Asparagine biosynthesis is also an important element of nitrogen metabolism in plants,[339] and differential expression of ASNS has been correlated with molecular processes underlying the germination and senescence of sunflowers.[340] The enzyme has been cloned, or isolated, from a number of organisms including yeast,[341] bacteria,[342,343] and mammals.[344–346] Early work on ASNS isolated from mammalian sources was complicated by the low abundance and instability of these proteins,[344,345] and the ASNS from bovine pancreas remains the only native mammalian enzyme that has been studied in detail.[346–349] Large amounts of active, recombinant human ASNS can be obtained using a baculovirus-based expression system,[350] however, which should facilitate the detailed characterization of the eukaryotic form of the enzyme. Methods have also been developed to obtain recombinant forms of the glutamine-dependent ASNS present in *E. coli* (AS-B),[61] and *Vibrio*
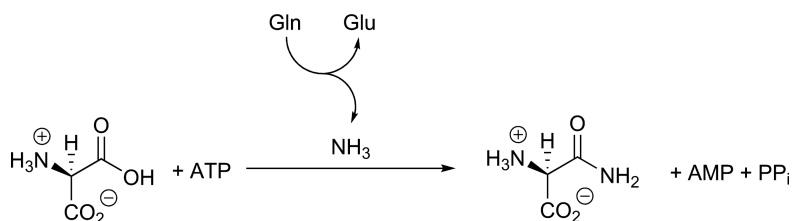

Scheme 28
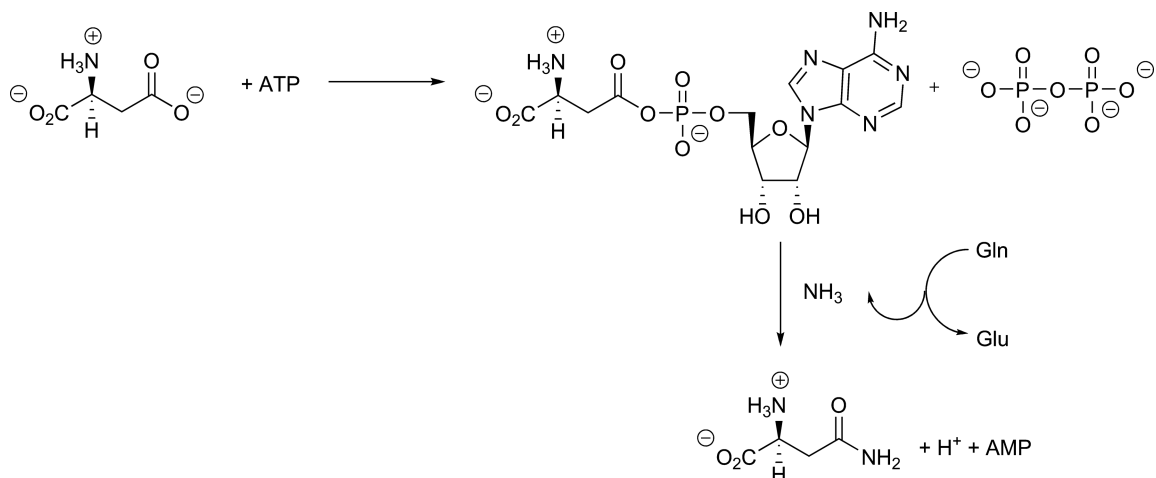
Scheme 29

cholerae.[351] Efforts to determine the kinetic mechanism used by the enzyme are complicated by the very high glutaminase activity of the enzyme in the absence of the other substrates, and several different proposals for the order of substrate binding and product release have been advanced.[349,351–355] Only one kinetic scheme appears to be consistent with the observed dependence of the glutamate:asparagine ratio on the initial glutamine concentration, and this model predicts that the catalytic activities of the glutaminase and synthase sites are only weakly coupled.[352] This kinetic behavior contrasts sharply with that of almost all other amidotransferases for which the rate of glutamine hydrolysis is significantly enhanced by the presence of the appropriate nitrogen acceptor. Recent *in vitro* studies on AS-B[356] and the *Vibrio* enzyme[351] have shown, however, that asparagine can bind to the N-terminal active site and is a competitive inhibitor of glutaminase activity (K_I: 50–60 μmol l^{-1}). The presence of this amino acid in the cell therefore represents a mechanism for regulating the glutamine-dependent activities of the enzyme, and preventing futile glutamine hydrolysis. Despite the kinetic complexity of ASNS, isotope-labeling experiments using ^{18}O have clearly demonstrated that the side chain carboxylate of aspartic acid is activated by adenylation to yield a β-aspartyl-AMP intermediate,[346,353] which can then undergo reaction with the ammonia molecule liberated in the glutaminase active site (**Scheme 29**).

8.06.4.4.2   *Enzyme structure*

An X-ray crystal structure of the ternary complex involving AMP, glutamine, and the C1A mutant of *E. coli* AS-B, which lacks glutaminase activity due to replacement of Cys1 by alanine, has been determined at a resolution of 2.0 Å (**Figure 36**).[65] As expected on the basis of sequence alignments,[8,47,67] ASNS is composed of two distinct domains. The location of the synthetase site was assigned from that of the bound AMP, and the C-terminal domain was observed to be homologous to the synthetase domain of GMPS,[25] rather than the ammonia-dependent asparagine synthetase (AS-A) present in prokaryotes.[357–359] Although ASNS and GMPS share very little sequence similarity, both enzymes possess an SGGXDS sequence 'signature', which appears to be present in all enzymes that couple the conversion of ATP to AMP and PP$_i$ to their overall chemical transformation,[360] including argininosuccinate synthetase,[361,362] ATP sulfurylase,[363,364] β-lactam synthetase (BLS),[365–367] and ThiI (4-thiouridine synthetase).[368,369] Disorder in the C-terminal domain of the C1A AS-B/Gln/AMP complex, however, prevented the observation of electron density for two loop regions (Ala250 to Leu267 and Cys422 to Ala426) and the final 40 residues of the enzyme, suggesting that aspartate binding may be important in organizing the synthetase domain for catalysis. Insight into the residues that likely catalyze β-aspartyl-AMP and/or asparagine formation in the synthetase site has been provided by crystallographic studies on BLS,[365,370] a key enzyme in clavulanic acid biosynthesis,[371] which has either evolved from (or shares a common ancestor with) ASNS (**Figure 37**).[367,372] Thus, these two enzymes catalyze reactions with similar chemistry in that both substrates are activated as their AMP derivatives and the carbonyl group is attacked by a nitrogen nucleophile
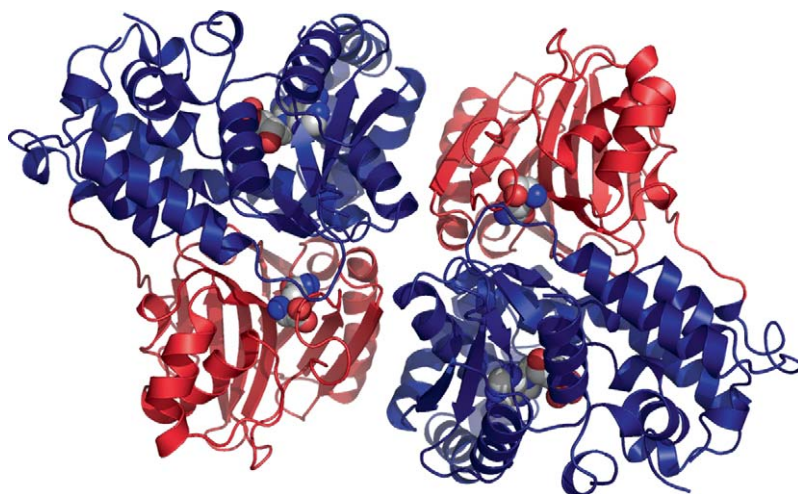
Figure 36 Cartoon representation of the *Escherichia coli* AS-B C1A mutant complexed to glutamine and AMP (1CT9). The glutaminase and synthase domains of each monomer are colored red and blue, respectively, showing the 'head-to-tail' packing in the dimer. Bound glutamine and AMP are shown as CPK models. Coloring: C – gray, O – red, N – blue, and P – orange. Image rendered in PYMOL.

(a)   (b)
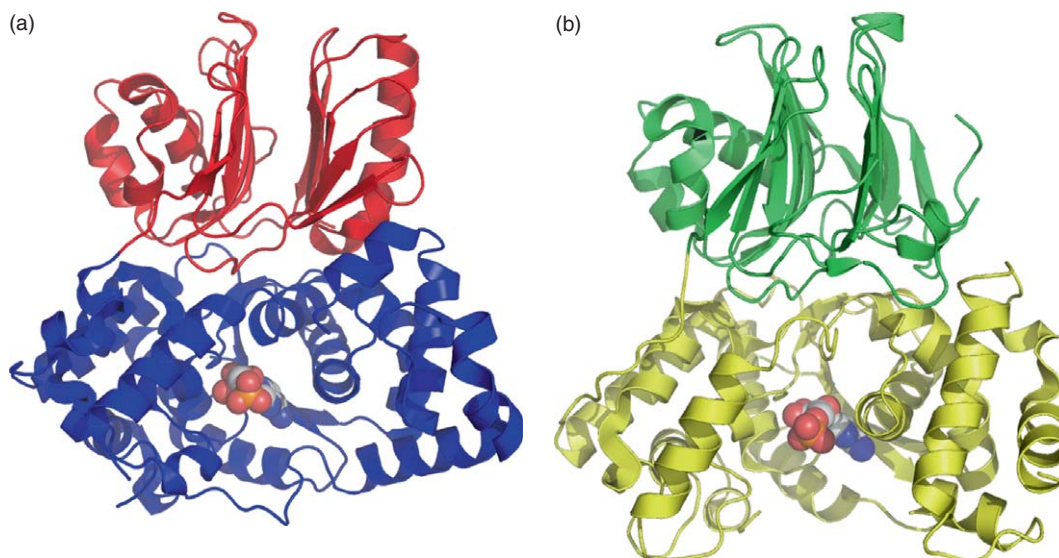
Figure 37 Structural comparison of (a) the *Escherichia coli* AS-B C1A/AMP complex (1CT9) and (b) *Streptomyces clavuligerus* BLS bound to ATP (1MB9). Domain coloring for AS-B is as described in **Figure 36**. The N- and C-terminal domains of BLS are colored yellow and green, respectively. AMP and ATP are rendered as CPK models. Coloring: C – gray, O – red, N – blue, and P – orange. Image rendered in PYMOL.

(**Scheme 30**). It therefore seems reasonable to assume that the catalytic machinery for substrate adenylation is conserved in both enzymes.

The C1A AS-B/Gln/AMP structure also shows that ammonia released in the glutaminase site travels a distance of ∼20 Å through a solvent-inaccessible tunnel to reach the C-terminal synthetase site (**Figure 38**).[65] Most of the residues defining the interior surface of this tunnel are hydrophobic (Met120, Ile142, Ile143, Leu232, Met329, Ala399, and Val401), although conserved polar residues are found close to the glutaminase (Arg30) and synthetase (Ser346 and Glu348). In contrast to GPATase, for which a significant conformational change in a disordered loop is
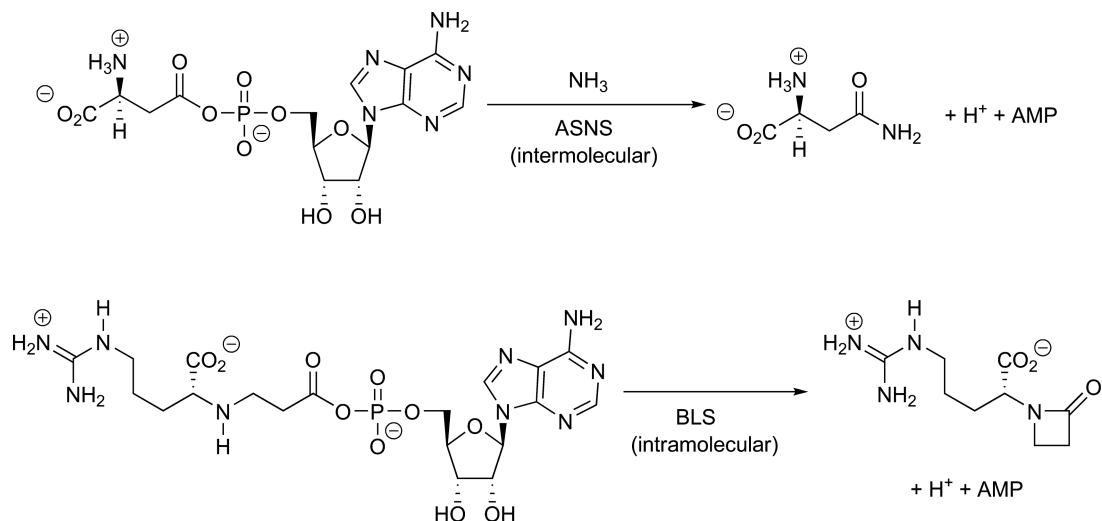
Scheme 30
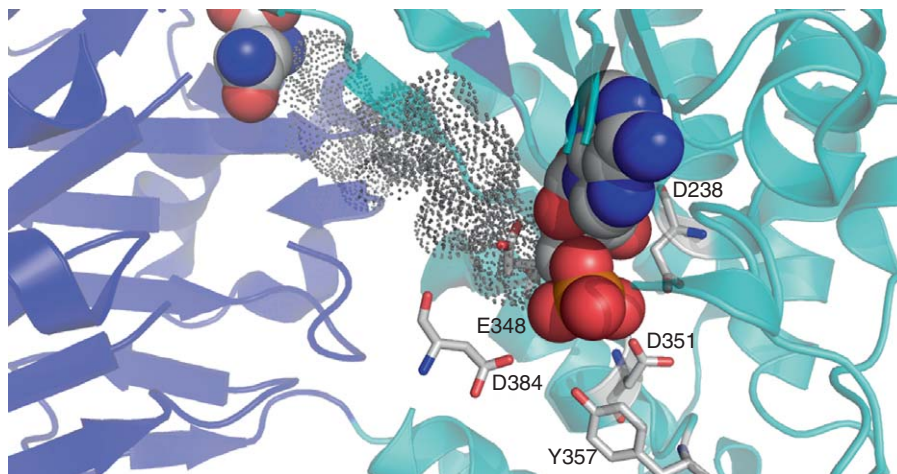
Figure 38 Close-up of the intramolecular ammonia tunnel (gray dotted spheres) present in the AS-B C1A/glutamine/AMP complex (1CT9). Charged residues defining the C-terminal region of the tunnel are shown as sticks, and the glutaminase and synthetase domains are colored blue and cyan, respectively. Glutamine and AMP are rendered as CPK models. Coloring: C – gray, O – red, N – blue, and P – orange. Image rendered in PYMOL.

needed to create the ammonia tunnel connecting the active sites, most of these hydrophobic residues in AS-B are located in the protein interior on secondary structural elements, making it entirely possible that the tunnel structure in ASNS is maintained throughout the entire catalytic cycle. Several water molecules are observed at the interface of the two domains, close to the glutaminase site, suggesting that exogenous ammonia might be able to access the tunnel when the enzyme adopts conformations other than that seen in the crystal.

8.06.4.4.3 Investigating intramolecular ammonia translocation

The apparent lack of significant coupling of the glutaminase and synthetase activities in ASNS raises questions about the efficiency with which ammonia is transferred through the tunnel that links the two active sites.[352] A competition experiment using *E. coli* AS-B, similar to that performed for CPS,[107] was therefore undertaken to assess the ability of exogenous ^{15}N-labeled ammonia to suppress ^{15}N incorporation from unlabeled glutamine.[356] This study used a novel isotope-edited ^{1}H NMR-based (gradient heteronuclear multiple

quantum coherence (gHMQC)) assay[373] having a significantly enhanced level of sensitivity relative to the ^{15}N NMR measurements used for studying ammonia transfer in CPS. The results of these gHMQC measurements showed that glutamine could not completely suppress ^{15}N incorporation into product asparagine, even at concentrations of $40 \, \text{mmol} \, l^{-1}$. In addition, ^{15}N-labeled ammonia from bulk solution was able to attack the thioester intermediate formed as an intermediate in the N-terminal glutaminase active site,[374,375] even though aspartate and ATP were present at saturating levels. This unexpected activity must therefore occur when these substrates are both bound to the C-terminal synthetase site, suggesting that exogenous ammonia can access the ammonia tunnel when glutamine, ATP, and aspartate are all present on the enzyme. The pathway by which exogenous ammonia can access the glutaminase site during turnover remains to be established, although it is possible that the tunnel becomes accessible to solvent by a conformational change in the protein that leads to uncoupling of the synthetase and glutaminase activities with increasing concentrations of glutamine.[352] This unexpected observation sharply contrasts with the tight kinetic coupling of the glutaminase and synthetase activities seen in almost all other amidotransferases,[108,158] and with the suppression of ammonia incorporation in the CPS-catalyzed reaction at saturating glutamine concentrations.[107]

The situation is complicated, however, by the observation that, in the absence of aspartate, the glutaminase activity of AS-B is stimulated approximately twofold by the presence of ATP, AMP-PNP (a nonhydrolyzable ATP analog[376]), or AMP and PP$_i$.[83] The fact that this effect is associated with binding to the C-terminal synthetase site by these compounds is further supported by the observation that no such stimulation is seen when only AMP or PP$_i$ are present in the assay.[83] Kinetic studies showed that this effect was abolished in AS-B variants in which Arg30 was replaced by alanine or lysine,[83] suggesting the presence of 'vestigial' interdomain interactions leading to conformational changes that 'report' on active site occupancy, similar to those seen in GPATase,[242] GFAT,[10,80] and GltS.[318,319] Given that the enzyme crystallizes as a 'head-to-tail' homodimer, it remains unclear whether such conformational changes are transmitted within an AS-B monomer or via intermolecular contacts between the glutaminase domain of one monomer with the synthetase domain of the other (**Figure 36**). The importance of Arg30, a residue that is conserved in all known glutamine-dependent asparagine synthetases, most likely arises from hydrogen bonding interactions[65] that control the side chain orientation of Asn74, which is part of the oxyanion hole needed for stabilization of the tetrahedral intermediate formed during the glutaminase reaction.[82] The 'optimization' of this vestigial level of active site communication may have been precluded by the ability of asparagine to compete with glutamine for the N-terminal-binding site, ameliorating any detrimental cellular effects associated with the unusually high intrinsic glutaminase activity of the enzyme.

8.06.5 Class III Amidotransferases

8.06.5.1 Aminoacyl-tRNA Amidotransferase

8.06.5.1.1 *Catalytic mechanism*

Aminoacyl tRNA synthetases ensure the accuracy of ribosomal protein biosynthesis by ensuring that amino acids are correctly esterified to their corresponding tRNA molecules.[377,378] The molecular strategies by which these enzymes achieve their exquisite levels of specificity have therefore been the subject of extensive structural,[379–381] kinetic,[382,383] and computational investigations.[384,385] The discovery that the synthesis of Gln-tRNA$^{\text{Gln}}$ in archaea and many bacteria did not proceed in a single step, as is the case in eukaryotes, was therefore unexpected.[386–388] Subsequent experiments showed that archaea generate Gln-tRNA$^{\text{Gln}}$ using a two-step pathway in which tRNA$^{\text{Gln}}$ is mischarged with glutamate to give Glu-tRNA$^{\text{Gln}}$, which is then amidated to form Gln-tRNA$^{\text{Gln}}$. The latter reaction is catalyzed by a Class III glutamine-dependent amidotransferase using ATP as an energy source (**Scheme 4**). This two-step pathway is also used by many eubacteria but is absent in eukaryotes, and so this amidotransferase represents a potential target for the development of novel antibiotics.[62,389–391]

Two variants of glutamine-dependent Glu-tRNA$^{\text{Gln}}$ amidotransferase have been identified and characterized. One form (GatDE) is present only in archaea and is a heterodimer composed of two subunits.[387,392] The GatD subunit catalyzes the hydrolysis of glutamine to yield glutamate and ammonia, and seems evolutionarily related to L-asparaginase.[393,394] GatE activates Glu-tRNA$^{\text{Gln}}$ for reaction with ammonia by catalyzing phosphoryl transfer from ATP to yield an acylphosphate intermediate.[392] Ammonia, released from glutamine in the GatD active site, can then attack the acylphosphate to yield Gln-tRNA$^{\text{Gln}}$ and P$_i$ (**Scheme 31**). The second
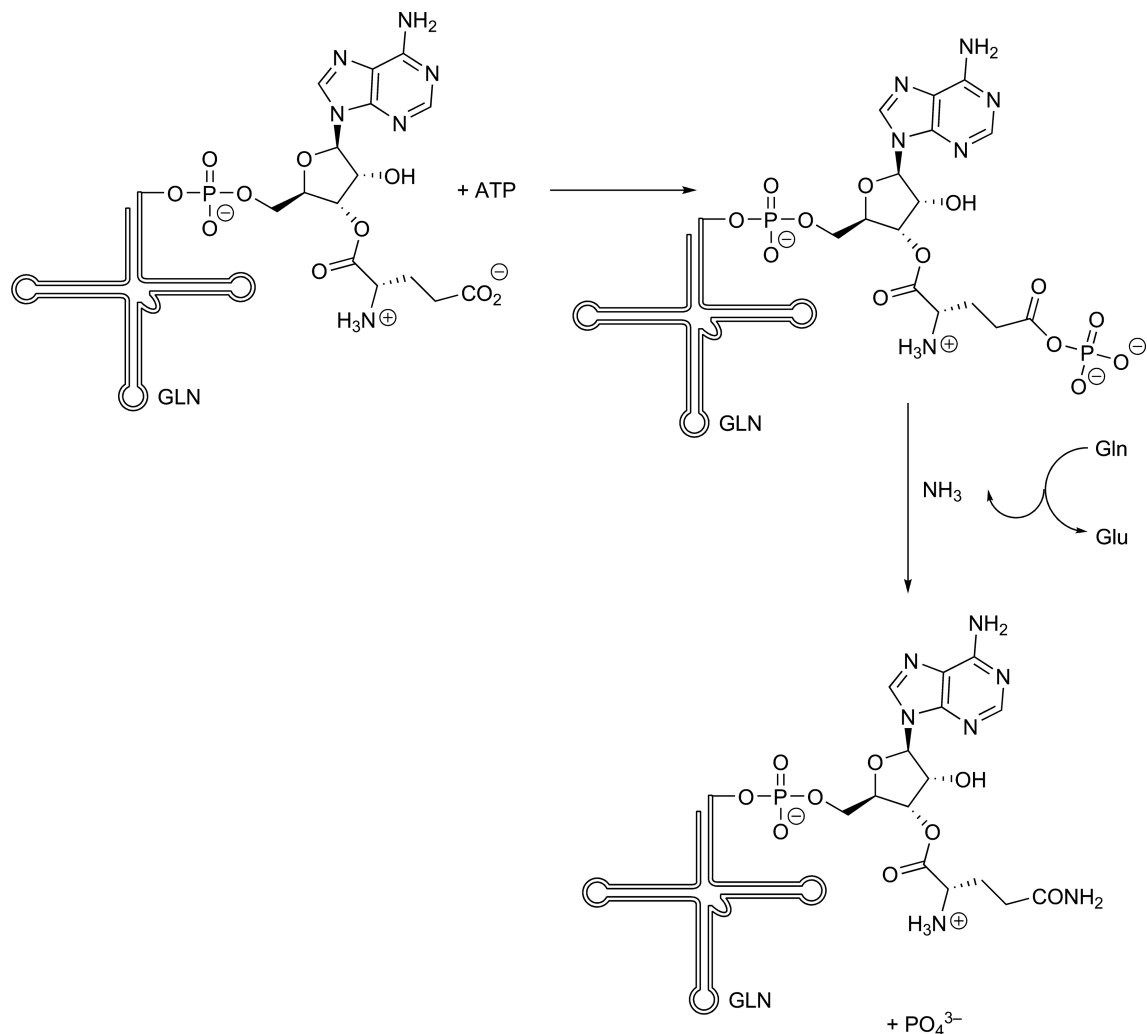
Scheme 31

variant of the amidotransferase, which is found in many eubacteria and some archaea,[93,395–400] is a heterotrimer (GatCAB) that not only catalyzes the conversion of Glu-tRNA^{Gln} to Gln-tRNA^{Gln} but also the formation of Asn-tRNA^{Asn} from the 'mischarged' precursor Asp-tRNA^{Asn}. The genes encoding all three subunits, GatA, GatB, and GatC, are located on the same operon. The ability of this heterotrimer to produce Asn-tRNA^{Asn} is especially important in bacteria that lack either ammonia- or glutamine-dependent asparagine synthetases,[55,401,402] and the components (aspartyl-tRNA synthetase, GatCAB) needed to synthesize Asn-tRNA^{Asn} in *Thermus themophilus* seem to be localized in a 'transamidosome' ribonucleoprotein complex that remains stable throughout the entire process.[403] As expected on the basis of sequence similarity to amidases,[94] GatA mediates the hydrolysis of glutamine to glutamate and ammonia. Although GatA and GatD are not structurally similar and use different residues to catalyze amide bond cleavage, GatB is highly similar to its archaeal counterpart GatE in that ATP is used to activate the unreactive carboxylate by formation of an acylphosphate intermediate and ADP (Scheme 31).[387] Phosphoryl transfer likely requires the presence of two Mg^{2+} ions, as observed for GS,[404] and GTP cannot function as an alternate donor.[405] On the other hand, whereas GatE is specific for Glu-tRNA^{Gln}, GatB can use both Glu-tRNA^{Gln} or Asp-tRNA^{Asn} as substrates. Given that both the kinase and synthetase activities required for glutamine-dependent amidation are associated with the GatB subunit, the function of GatC remained ill-defined until the determination of a crystal structure for the GatCAB enzyme from *Staphylococcus aureus*.[406]

8.06.5.1.2 Enzyme structure

Crystal structures have been solved for *S. aureus* GatCAB,[406] and the GatDE heterodimers present in *Methanothermobacter thermoautotrophicus* and *Pyrococcus abyssi*,[92,407] albeit at relatively low resolution (2.5–3.15 Å). These studies have provided information concerning not only the structure of the ammonia tunnel linking the glutaminase and synthetase sites in the two enzymes, but also the molecular interactions responsible for discrimination of the aminoacyl-tRNA substrates. Moreover, the function of GatC seems to be in stabilizing the GatA/GatB heterodimer, as this protein is 'wrapped' about the interface formed by these subunits and forms numerous intermolecular interactions (**Figure 39(a)**). This finding therefore explains the fact that coexpression of the three subunits is essential for GatCAB activity.[93] Soaking crystals of GatCAB with glutamine gave cocrystals that could be used to solve the structure of the GatCAB/glutamine complex, which showed density for a bound ligand close to the Ser-*cis*-Ser–Lys catalytic triad present in the GatA subunit of the complex. Closer examination of the density suggested that the side chain of Ser178 had become covalently bonded to the ligand, and so this species was assigned as an oxyanion intermediate, suggesting that GatA uses a catalytic mechanism that is identical to that of other amidase enzymes.[99–101] The apparent ability of the enzyme to react with glutamine in the absence of a Glu-tRNA^{Gln} molecule bound to the GatB domain is also consistent with the relatively high intrinsic glutaminase activity that is observed for GatCAB in the absence of other substrates.[408] The other catalytic subunit, GatB, is composed of two domains connected by a long (60 Å) peptide linker. Although it was not possible to cocrystallize Glu-tRNA^{Gln} or tRNA^{Gln} with the enzyme, complicating efforts to delineate the basis for how GatCAB discriminates against Glu-tRNA^{Glu} or tRNA^{Glu}, the location of the kinase/synthetase site was determined from the structure of the enzyme complexed to ADP.ALF$_4^-$. As in all other structurally characterized amidotransferases, a tunnel of 30 Å in length was identified, using the CAVER software package,[409] which linked the glutaminase and synthetase sites. The side chains of several conserved polar, charged amino acids (Arg200, Asp211, and Arg323 from GatA, and Lys79, Lys88, Glu125, Glu272, and Asp274 from GatB) line the tunnel between the active sites, which is also filled with an organized network of water molecules (**Figure 40**). This finding is somewhat surprising given
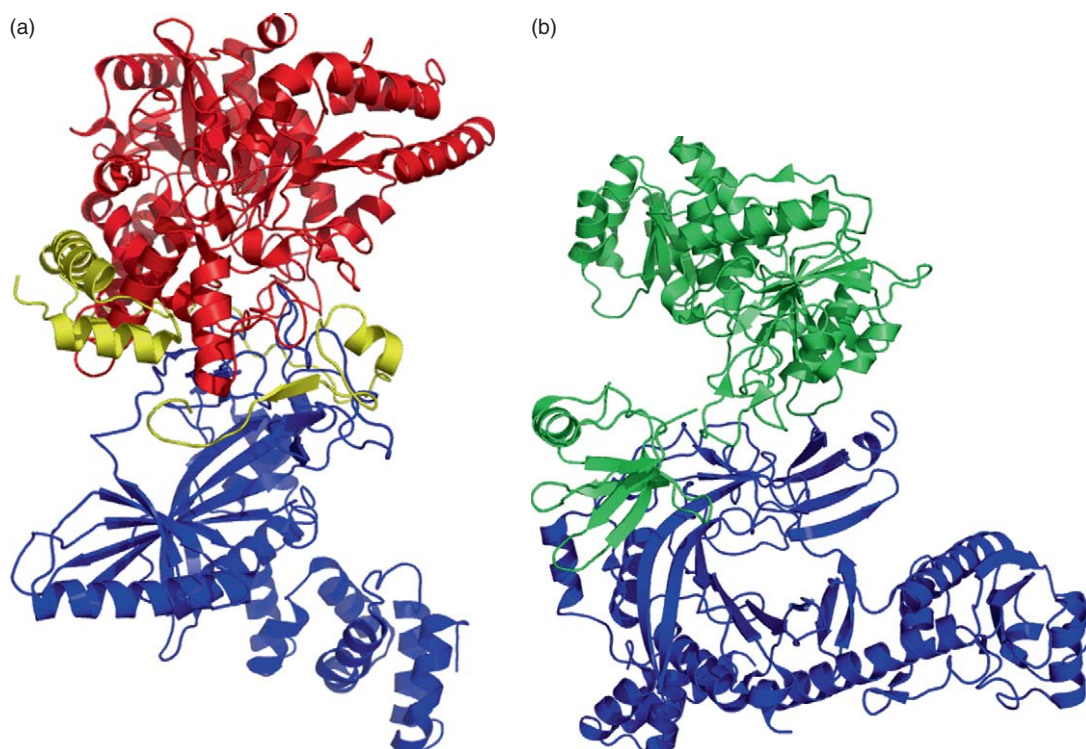
Figure 39 (a) Cartoon representation of *Staphylococcus aureus* GatCAB showing the interaction of the GatC (yellow) subunit with the GatA (red)/GatB (blue) heterodimer (2G5H). (b) Cartoon representation of *Methanothermobacter thermoautotrophicus* GatDE (2D6F). The GatD and GatE subunits are colored green and blue, respectively.
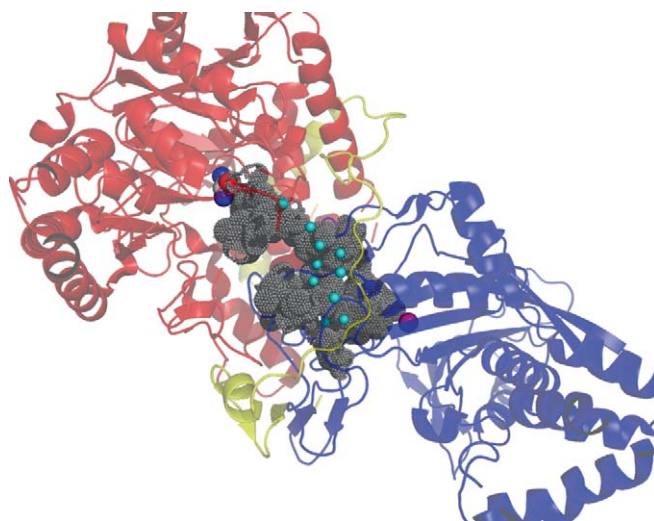
Figure 40   The intramolecular tunnel (gray dotted spheres) seen in *Staphylococcus aureus* GatCAB heterotrimer. Subunit coloring is as described in **Figure 39**. Water molecules are shown as cyan spheres, and the location of Mg^{2+} is indicated by the magenta sphere. Bound glutamine is shown as a CPK model. Coloring: C – gray, O – red, N – blue. Image rendered in PYMOL.

that predominantly hydrophobic residues line the tunnels observed in other amidotransferases,[11–13] and it has been suggested that ammonia does not pass along the tunnel in a neutral form as hypothesized for IGPS[153] or GFAT.[302] Instead, a 'proton-relay' was envisaged in which ammonia is moved through the tunnel as a result of consecutive protonation/deprotonation steps involving its interaction with alternating positively and negatively charged side chains.[406] This unique translocation mechanism remains to be experimentally validated, although the Lys79 side chain (from the GatB subunit), which is located at the entrance to the synthetase site, might function as a base to generate ammonia from an ammonium cation.

The structure of the GatDE form of the amidotransferase shows many features similar to those observed in GatCAB, including widely separated active sites linked by a tunnel composed of highly charged side chains even though the glutaminase domains GatD and GatA have different folds and are evolutionarily unrelated. Unlike GatCAB, GatDE only amidates Glu-tRNA^{Gln} and exhibits no significant glutaminase activity in the absence of this substrate.[392,407] Insights into the molecular basis for this ability to recognize its substrate were possible for this enzyme, however, because GatDE could be cocrystallized with tRNA^{Gln} (**Figure 39(b)**). The structure of this complex showed that (1) the transfer RNA bound only to the GatE subunit and (2) no protein residues formed intermolecular interactions with the tRNA^{Gln} anticodon. Having identified the substrate-binding site, the significant structural similarity of the GatB and GatE domains permitted identification of the ATP-binding site in GatE by comparison with the GatB/ADP.AlF$_4^-$ complex.[406] This yielded a model in which the glutaminase and kinase/synthetase active sites were separated by a distance of 40 Å. As for GatCAB, however, detailed analysis suggested that these sites in GatDE are linked by a tunnel lined predominantly with the side chains of polar residues that are strictly conserved in all archaeal sequences.

The unexpected finding that GatE residues do not interact with the tRNA^{Gln} anticodon, even though GatDE can discriminate tRNA^{Gln} from tRNA^{Glu} and tRNA^{Asn} with high levels of specificity, has been rationalized on the basis of model building studies. Thus, computational docking of the nondiscriminating *T. thermophilus* glutamyl-tRNA^{Gln} aminoacyl synthetase (GluRS)[410] with the GatDE/tRNA^{Gln} complex was accomplished by superimposing the tRNA molecules in each structure. This gave a protein complex in which GluRS could be inserted into the concave surface of two domains in GatE without causing major steric clashes between tRNA^{Gln}, which forms complementary interactions with the C-terminal helical and tail domains of GatE, and GluRS. Efforts to build a similar model using the structure of aspartyl-tRNA^{Asn} aminoacyl synthetase (AspRS)[411] failed, however, because of significant steric interactions between AspRS and the 'AspRS-like insertion' domain of GatE, which is not present in the GatB subunit. Thus, it is possible that substrate discrimination results from the formation of a GatDE/GluRS complex that permits direct

transfer of Glu-tRNA^{Gln} to the amidotransferase, thereby preventing transport of the misacylated tRNA to the ribosome by EF-Tu.[412,413] This is an interesting hypothesis for which there is some support from studies of the evolution of GluRS, GatDE, and GatCAB.[414–416]

8.06.5.1.3   Investigating intramolecular ammonia translocation

A limited set of mutagenesis studies have been reported aimed at demonstrating the functional importance of ammonia translocation through the 40 Å tunnel present in the GatDE/tRNA^{Gln} complex (**Figure 41**).[92] Thus, the replacement of either Leu336 in the GatD subunit or Leu178 in the GatE subunit by phenylalanine gave GatDE variants with glutaminase and kinase activities similar to those of the wild-type enzyme but effectively abolished the formation of the Gln-tRNA^{Gln} product. Although this result is consistent with the idea that the larger side chains block movement of ammonia through the tunnel, no structural studies that confirm this assertion have yet to be reported. When Tyr373 in the GatD subunit, which lies at the N-terminal end of the putative tunnel, was replaced by phenylalanine, however, the glutaminase activity was significantly reduced. Although it has been suggested that this residue plays a role in ensuring ammonia access into the tunnel, this hypothesis remains to be confirmed.

8.06.6   Other Aspects

8.06.6.1   Intramolecular Ammonia Tunnel Structure

Based on the evidence discussed above, it is likely that all glutamine-dependent amidotransferases contain an intramolecular tunnel linking multiple catalytic sites in their active conformation. This is a remarkable finding in light of the large variation in the three-dimensional folds of the synthase/synthetase domains in these enzymes, and illustrates the extent to which the structural 'plasticity' of proteins can be exploited by natural selection.[417] In consequence, there are few 'rules' governing the length and shape of the tunnels that have been structurally characterized, and, perhaps unexpectedly given the largely hydrophobic nature of known ammonia
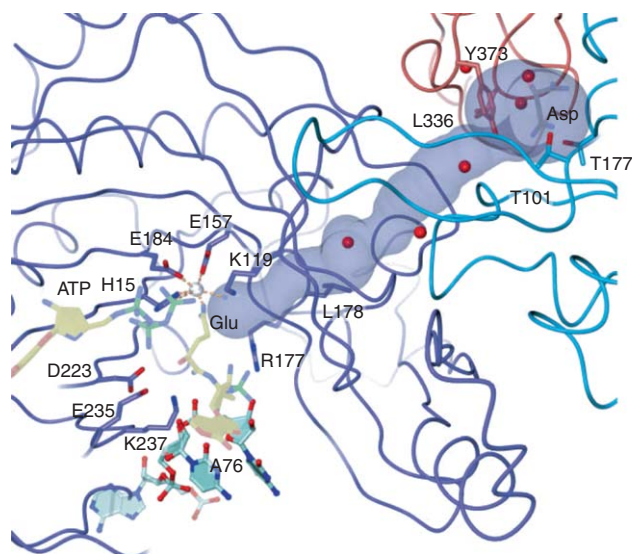
Figure 41   The ammonia tunnel in *Methanothermobacter thermoautotrophicus* GatDE showing the positions of the Leu178 and Leu336 side chains. GatE is represented by a violet-colored wire model, and similar representations of two GatD molecules are colored cyan and orange. In the GatD glutaminase site, the location of bound L-aspartate is modeled from the crystal structure of *Pyrococcus abyssi* GatDE.[407] The tunnel is shown by gray shading and water, or possibly ammonia, molecules are rendered as red spheres. From H. Oshikane; K. Sheppard; S. Fukai; Y. Nakamura; R. Ishitani; T. Numata; R. L. Sherrer; L. Feng; E. Schmitt; M. Panvert; S. Blanquet; Y. Mechulam; D. Söll; O. Nureki, *Science* **2006**, *312*, 1950–1954. Reprinted with permission from AAAS.

tunnels, Glu-tRNA^{Gln} amidotransferase possesses a tunnel that is lined with the side chains of polar amino acids. In contrast to membrane-bound ammonia transporters,[416–420] ammonia translocation through known amidotransferases does not seem to be 'driven' by differences in electrostatic potential at the two ends. This is probably because the relatively small lengths of these intramolecular tunnels cause the time required for ammonia to travel between active sites to be an insignificant fraction of that required for the enzyme to complete a single turnover.[13] In addition, the two tunnels that have been studied computationally appear to discriminate against transporting ammonium ions between the active sites,[151,302] reducing the magnitude of electrostatic interactions between the substrate and the protein. The seeming ability of tunnel residues to discriminate between ammonia and ammonium ion is also interesting because the nitrogen atom is translocated to the synthase/synthetase active site in a chemically reactive form. The involvement of enzyme-bound water in the transport process remains unclear, even though computational evidence has been cited as supporting the notion that the tunnel in IGP synthase can discriminate between ammonia and bulk water.[151] Indeed, the extent to which amidotransferase tunnels contain water prior to the release of ammonia from the glutaminase active site remains uncertain, and there is considerable variation in the number and location of ordered water molecules observed in the crystal structures of these enzymes. For enzymes such as GPATase, in which significant conformational changes must take place in a disordered loop of the protein (**Figure 25**), however, it is hard to imagine that solvent water can easily be excluded from the tunnel during catalytic turnover. Equally, even if a tunnel appears to be inaccessible to solvent by crystallography and is predominantly lined by hydrophobic side chains, exogenous ammonia may still be able to access the structure, as revealed by the glutamine/^{15}N-NH$_4$Cl competition experiments for AS-B.[356] Of course, the importance of excluding water from the region in which ammonia is utilized is dependent on the nature of the substrate. Thus, the presence of water is unlikely to interfere with the addition of ammonia to ketones or aldehydes in the active site, and, given the higher nucleophilicity of ammonia relative to that of water,[421] this might also prove to be the case for enzymes in which substrates are activated by phosphorylation.

8.06.6.2   Protein Dynamics and Ammonia Translocation

The extent to which protein dynamics play a role in facilitating ammonia translocation is another issue that remains to be resolved. Energy minimization of amidotransferase X-ray crystal structures using well-established force fields, such as CHARMM[422] or AMBER,[423] often yields optimized structures in which there is insufficient space for ammonia pass through the protein (Wang and Richards, unpublished results). Whether this is merely a consequence of the absence of water molecules in the crystal structure that are normally present during turnover, or is a reflection of the true resting forms of these enzymes remains to be established. 'Steered' MD simulations suggest that dynamical motions in the protein can lead to widening of tunnel diameter,[151,302] presumably enhancing the ability of ammonia to pass through the tunnel. Additional support for the coupling of tunnel size to ammonia motions within the protein is provided by the side chain displacements that take place as *E. coli* GFAT moves through its catalytic cycle.[10] On the other hand, this does not have to be true for enzymes containing a 'permanent' tunnel, such as CPS. Recent MD simulations of the small subunit of CPS have, however, suggested that side chain dynamics may be important to 'guide' ammonia through the protein,[424] raising the question of whether this is a general strategy used in all amidotransferases.

8.06.6.3   Molecular Evolution of Amidotransferases

The modular construction of glutamine-dependent amidotransferases is an elegant solution to the problem of developing a plethora of active sites capable of transferring nitrogen from a single molecular carrier to a wide range of electrophiles. On the other hand, such a strategy requires (1) the optimization of molecular interactions at the interface of the two domains (subunits) so as to permit intersite communication and (2) the development of a tunnel between the two active sites through which ammonia can be translocated. Both issues likely constrain the types of folds that can be 'combined' to yield functional amidotransferases, thereby precluding the use of only one type of glutaminase domain (subunit) for all amidotransferase-catalyzed reactions. Indeed, the molecular difficulties of assembling such multidomain, or multisubunit, enzymes are well illustrated by recent efforts to engineer 'hybrid' enzymes using protein engineering and library selection methods.[425–429] In part, this

reflects a lack of understanding of the evolutionary pathways that have lead to the amidotransferases that have been identified. In one model for constructing these 'complex' enzymes, a surface mutation in a glutaminase permits its weak, or transient, association with a catalytic module from another enzyme superfamily. In this situation, one might imagine ammonia transfer to be inefficient, perhaps taking place by release and rebinding. Optimization residues at the protein–protein interface by selection would then take place, resulting in the formation of long-lived heterodimers. Insights into how natural selection has 'optimized' domain–domain, or protein–protein, contacts in specific amidotransferases, however, are complicated by the lack of variation in the molecular interactions at the interface of the glutaminase and synthase/synthetase components across modern species, suggesting that these enzymes were assembled and optimized relatively early in evolutionary history. The optimization of tunnel structure and coordination of catalysis in the two active sites could then proceed subsequent to the creation of a stable heterodimer. To date, there have been few detailed studies of interfacial interactions in many of the enzymes outlined above, although modern strategies for probing protein–protein association[430,431] offer the possibility of evaluating the 'plasticity' in the interactions of known glutaminase domains with other catalytic modules.

8.06.7   Conclusion

That all glutamine-dependent amidotransferases contain tunnels, which sequester ammonia from bulk solution and translocate this reactive intermediate between multiple active sites, is a remarkable finding given the diversity of structure and chemistry exhibited by these enzymes. Although there has been remarkable progress in determining structure–function relationships for residues defining these intramolecular tunnels over the past decade, the paucity of knowledge concerning the role of conformational dynamics and protein 'plasticity" is still evident from the unexpected structural consequences of efforts to engineer 'holes' and 'blockages' within these structures. Moreover, even though it has been possible to identify certain critical molecular interactions that couple glutaminase activity to the presence of suitable substrates and/or intermediates in the synthase/synthetase active site for a small subset of amidotransferases, much remains to be learnt about the energetics of such interactions and their importance for the coordinated dynamical motions that are needed for catalysis.[432] Efforts to explore these problems may also be pertinent to obtaining a general understanding of the molecular mechanisms involved in channeling reaction products in biosynthetically important multienzyme complexes.[433–436]

References

1. S. W. Ragsdale, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 165–195.
2. E. W. Miles; S. Rhee; D. R. Davies, *J. Biol. Chem.* **1999**, *274*, 12193–12196.
3. D. Leys; J. Basran; N. S. Scrutton, *EMBO J.* **2003**, *22*, 4038–4048.
4. W. Wang; T. J. Kappock; J. Stubbe; S. E. Ealick, *Biochemistry* **1998**, *37*, 15647–15662.
5. D. C. Hess; S. Lu; J. D. Rabinowitz; D. Botstein, *PLoS Biol.* **2006**, *4*, 2012–2023.
6. J. Ricard; B. Gontero; L. Avilan; S. Lebreton, *Cell. Mol. Life Sci.* **1998**, *54*, 1231–1248.
7. P. Srere, *Trends Biochem. Sci.* **1994**, *19*, 519–520.
8. H. Zalkin, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1993**, *66*, 203–309.
9. J. M. Buchanan, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1973**, *39*, 91–183.
10. S. Mouilleron; B. Golinelli-Pimpaneau, *Curr. Opin. Struct. Biol.* **2007**, *17*, 653–664.
11. A. Weeks; L. Lund; F. M. Raushel, *Curr. Opin. Chem. Biol.* **2006**, *10*, 465–472.
12. F. M. Raushel; J. B. Thoden; H. M. Holden, *Acc. Chem. Res.* **2003**, *36*, 539–548.
13. X. Huang; H. M. Holden; F. M. Raushel, *Annu. Rev. Biochem.* **2001**, *70*, 149–180.
14. M. Wojcik; H. F. Seidle; P. Bieganowski; C. Brenner, *J. Biol. Chem.* **2006**, *281*, 33395–33402.

15. C. K. Yu; L. S. Dietrich, *J. Biol. Chem.* **1972**, *247*, 4794–4802.
16. P. Bieganowski; H. C. Pace; C. Brenner, *J. Biol. Chem.* **2003**, *278*, 33049–33055.
17. V. Fresquet; L. Williams; F. M. Raushel, *Biochemistry* **2007**, *46*, 13983–13993.
18. L. Williams; V. Fresquet; P. J. Santander; F. M. Raushel, *J. Am. Chem. Soc.* **2007**, *129*, 294–295.
19. F. Blanche; M. Couder; L. Debussche; D. Thibaut; B. Cameron; J. Crouzet, *J. Bacteriol.* **1991**, *173*, 6046–6051.
20. Q.-Z. Ye; J. Liu; C. T. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9391–9395.
21. J. F. Parsons; P. Y. Jensen; A. S. Pachikara; A. J. Howard; E. Eisenstein; J. E. Ladner, *Biochemistry* **2002**, *41*, 2198–2208.
22. S. Keller; H. S. Schadt; I. Ortel; R. D. Süssmuth, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 8284–8286.
23. M. Hirst; E. Haliday; J. Nakamura; L. Lou, *J. Biol. Chem.* **1994**, *269*, 23830–23837.
24. J. Y. Bhat; B. G. Shastri; H. Balaram, *Biochem. J.* **2008**, *409*, 263–273.
25. J. J. G. Tesmer; T. J. Klem; M. L. Deras; V. J. Davisson; J. L. Smith, *Nat. Struct. Biol.* **1996**, *3*, 74–86.
26. A. A. Morollo; M. J. Eck, *Nat. Struct. Biol.* **2001**, *8*, 243–247.
27. G. Spraggon; C. Kim; X. Nguyen-Huu; M.-C. Yee; C. Yanofsky; S. E. Mills, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6021–6026.
28. T. Knöchel; A. Ivens; G. Hester; A. Gonzalez; R. Bauerle; M. Wilmanns; K. Kirschner; J. N. Jasonius, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9479–9484.
29. A. A. Morollo; R. Bauerle, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9983–9987.
30. C. T. Walsh; J. Liu; F. Rusnak; M. Sakaitani, *Chem. Rev.* **1990**, *90*, 1105–1129.
31. C. Gaille; C. Reimmann; D. Haas, *J. Biol. Chem.* **2003**, *278*, 16893–16898.
32. J. Liu; N. Quinn; G. A. Berchtold; C. T. Walsh, *Biochemistry* **1990**, *29*, 1417–1425.
33. A. R. Knaggs, *Nat. Prod. Rep.* **2003**, *20*, 119–136.
34. L. Reitzer, *Annu. Rev. Microbiol.* **2003**, *57*, 155–176.
35. M. G. Klotz; L. Y. Stein, *FEMS Microbiol. Lett.* **2008**, *278*, 146–156.
36. D. C. Rees; J. B. Howard, *Curr. Opin. Chem. Biol.* **2000**, *4*, 559–566.
37. J. B. Howard; D. C. Rees, *Chem. Rev.* **1996**, *96*, 2965–2982.
38. S. G. Rhee; P. B. Chock; E. R. Stadtman, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1989**, *62*, 37–92.
39. W. W. Krajewski; R. Collins; L. Holmberg-Schiavone; T. A. Jones; T. Karlberg; S. L. Mowbray, *J. Mol. Biol.* **2008**, *375*, 217–228.
40. D. Eisenberg; H. S. Gill; G. M. U. Pfluegl; S. H. Rotstein, *Biochim. Biophys. Acta* **2000**, *1477*, 122–145.
41. A. Meister, *Meth. Enzymol.* **1985**, *113*, 185–199.
42. J. Hiratake, *Chem. Record* **2005**, *5*, 209–228.
43. J. M. Manning; S. Moore; W. B. Rowe; A. Meister, *Biochemistry* **1969**, *8*, 2681–2685.
44. S.-H. Liaw; C. Pan; D. Eisenberg, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4996–5000.
45. T. J. Kappock; S. E. Ealick; J. Stubbe, *Curr. Opin. Chem. Biol.* **2000**, *4*, 567–572.
46. M. E. Jones, *Annu. Rev. Biochem.* **1980**, *49*, 253–279.
47. N. G. J. Richards; S. M. Schuster, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1998**, *72*, 145–198.
48. M. A. Vanoni; B. Curti, *Cell. Mol. Life Sci.* **1999**, *55*, 617–638.
49. P. Alifano; R. Fani; P. Liòo; A. Lazcano; M. Bazzicalupo; M. S. Carlomagno; C. B. Bruni, *Microbiol. Rev.* **1996**, *60*, 44–69.
50. B. Badet; P. Vermoote; P.-Y. Haumont; F. Lederer; F. Le Goffic, *Biochemistry* **1987**, *26*, 1940–1948.
51. T. B. Fitzpatrick; N. Amrhein; B. Kappes; F. Macheroux; I. Tews; T. Raschle, *Biochem. J.* **2007**, *407*, 1–13.
52. A. I. Scott; C. A. Roessner, *Biochem. Soc. Trans.* **2002**, *30*, 613–620.
53. B. Roux; C. T. Walsh, *Biochemistry* **1992**, *31*, 6904–6910.
54. M. Ibba; H. D. Becker; C. Stathopoulos; D. L. Tumbula; D. Söll, *Trends Biochem. Sci.* **2000**, *25*, 311–316.
55. A. W. Curnow; M. Ibba; D. Söll, *Nature* **1996**, *382*, 589–590.
56. A. Lazcano; S. L. Miller, *J. Mol. Evol.* **1999**, *49*, 424–431.
57. G. Caetano-Anollés; H. S. Kim; J. E. Mittenthal, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9358–9363.
58. J. L. Smith, *Biochem. Soc. Trans.* **1995**, *23*, 894–898.
59. A. C. Storer; R. Ménard, *Meth. Enzymol.* **1994**, *244*, 486–500.
60. W. C. Earnshaw; L. M. Martins; S. H. Kaufmann, *Annu. Rev. Biochem.* **1999**, *68*, 383–424.
61. S. K. Boehlein; N. G. J. Richards; S. M. Schuster, *J. Biol. Chem.* **1994**, *269*, 7450–7457.
62. C. P. Decicco; D. J. Nelson; Y. Luo; L. Shen; K. Y. Horiuchi; K. M. Amsler; L. A. Foster; S. M. Spitz; J. J. Merrill; C. F. Sizemore; K. C. Rogers; R. A. Copeland; M. R. Harpel, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2561–2564.
63. M. E. Glasner; J. A. Gerlt; P. C. Babbitt, *Curr. Opin. Chem. Biol.* **2006**, *10*, 492–497.
64. J. A. Gerlt; P. C. Babbitt, *Annu. Rev. Biochem.* **2001**, *70*, 209–246.
65. T. M. Larsen; S. K. Boehlein; S. M. Schuster; N. G. J. Richards; J. B. Thoden; H. M. Holden; I. Rayment, *Biochemistry* **1999**, *38*, 16146–16157.
66. P. Bork; E. V. Koonin, *Proteins* **1994**, *20*, 347–355.
67. H. Zalkin; J. L. Smith, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1998**, *72*, 87–144.
68. J. B. Thoden; X. Huang; F. M. Raushel; H. M. Holden, *Biochemistry* **1999**, *38*, 16158–16166.
69. J. Kraut, *Annu. Rev. Biochem.* **1977**, *46*, 331–356.
70. M. A. Rishavy; W. W. Cleland; C. J. Lusty, *Biochemistry* **2000**, *39*, 7309–7315.
71. C. J. Lusty, *FEBS Lett.* **1992**, *314*, 135–138.
72. M. G. Chaparian; D. R. Evans, *J. Biol. Chem.* **1991**, *266*, 3387–3395.
73. F. J. Schendel; E. Mueller; J. Stubbe; A. Shiau; J. M. Smith, *Biochemistry* **1989**, *28*, 2459–2471.
74. J. B. Thoden; S. G. Miran; J. C. Phillips; A. J. Howard; F. M. Raushel; H. M. Holden, *Biochemistry* **1998**, *37*, 8825–8831.
75. X. Huang; F. M. Raushel, *Biochemistry* **1999**, *38*, 15909–15914.
76. A. Hewagama; H. I. Guy; M. Chaparian; D. R. Evans, *Biochim. Biophys. Acta* **1998**, *1388*, 489–499.
77. B. Mei; H. Zalkin, *J. Biol. Chem.* **1989**, *264*, 16613–16619.
78. M.-A. Badet-Denisot; L. René; B. Badet, *Bull. Chim. Soc. Fr.* **1993**, *130*, 249–255.
79. J. H. Kim; J. M. Krahn; D. R. Tomchick; J. L. Smith; H. Zalkin, *J. Biol. Chem.* **1996**, *271*, 15549–15557.
80. S. Mouilleron; M.-A. Badet-Denisot; B. Golinelli-Pimpaneau, *J. Biol. Chem.* **2006**, *281*, 4404–4412.

81. R. E. Handschumacher; C. J. Bates; P. K. Chang; A. T. Andrews; G. A. Fischer, *Science* **1968**, *161*, 62–63.
82. S. K. Boehlein; J. G. Rosa-Rodriguez; S. M. Schuster; N. G. J. Richards, *J. Am. Chem. Soc.* **1997**, *119*, 5785–5791.
83. S. K. Boehlein; N. G. J. Richards; E. S. Walworth; S. M. Schuster, *J. Biol. Chem.* **1994**, *269*, 26789–26795.
84. F. Massière; M.-A. Badet-Denisot, *Cell. Mol. Life Sci.* **1998**, *54*, 205–222.
85. R. M. Czerwinski; T. K. Harris; M. A. Massiah; A. S. Mildvan; C. P. Whitman, *Biochemistry* **2001**, *40*, 1984–1995.
86. J. A. Brannigan; G. Dodson; H. J. Duggleby; P. C. E. Moody; J. L. Smith; D. R. Tomchick; A. G. Murzin, *Nature* **1995**, *378*, 416–419.
87. C. Oinonen; J. Rouvinen, *Protein Sci.* **2000**, *9*, 2329–2337.
88. Y. Kim; S. Kim; T. N. Earnest; W. G. Hol, *J. Biol. Chem.* **2002**, *277*, 2823–2829.
89. H. J. Duggleby; S. P. Tolley; C. P. Hill; E. J. Dodson; G. Moody; P. C. E. Moody, *Nature* **1995**, *373*, 264–268.
90. D. Voges; P. Zwicki; W. Baumeister, *Annu. Rev. Biochem.* **1999**, *68*, 1015–1068.
91. C. Oinonen; R. Tikkanen; J. Rouvinen; L. Peltonen, *Nat. Struct. Biol.* **1995**, *2*, 1102–1108.
92. H. Oshikane; K. Sheppard; S. Fukai; Y. Nakamura; R. Ishitani; T. Numata; R. L. Sherrer; L. Feng; E. Schmitt; M. Panvert; S. Blanquet; Y. Mechulam; D. Söll; O. Nureki, *Science* **2006**, *312*, 1950–1954.
93. A. W. Curnow; K. Hong; R. Yuan; S. Kim; O. Martins; W. Winkler; T. M. Henkin; D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11819–11826.
94. H. Chebrou; F. Bigey; A. Arnaud; P. Galzy, *Biochim. Biophys. Acta* **1996**, *1298*, 285–293.
95. D. Fournand; A. Arnaud, *J. Appl. Microbiol.* **2001**, *91*, 381–393.
96. M. Kobayashi; H. Komeda; T. Nagasawa; M. Nishiyama; S. Horinouchi; T. Beppu; H. Yamada; S. Shimizu, *Eur. J. Biochem.* **1993**, *217*, 327–336.
97. M. P. Patricelli; B. F. Cravatt, *J. Biol. Chem.* **2000**, *275*, 19177–19184.
98. S. Shin; T. H. Lee; N. C. Ha; H. M. Koo; S. Y. Kim; H. S. Lee; Y. S. Kim; B. H. Oh, *EMBO J.* **2002**, *21*, 2509–2516.
99. S. Shin; Y. S. Yun; H. M. Koo; Y. S. Kim; K. Y. Choi; B.-H. Oh, *J. Biol. Chem.* **2003**, *278*, 24937–24943.
100. M. K. McKinney; B. F. Cravatt, *Annu. Rev. Biochem.* **2005**, *74*, 411–432.
101. A. L. B. Valina; D. Mazumder-Shivakumar; T. C. Bruice, *Biochemistry* **2004**, *43*, 15657–15672.
102. A. Meister, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1989**, *62*, 315–374.
103. J. B. Thoden; H. M. Holden; G. Wesenberg; F. M. Raushel; I. Rayment, *Biochemistry* **1997**, *36*, 6305–6316.
104. H. M. Holden; J. B. Thoden; F. M. Raushel, *Cell. Mol. Life Sci.* **1999**, *56*, 507–522.
105. J. Kim; F. M. Raushel, *Biochemistry* **2004**, *43*, 5334–5340.
106. J. B. Thoden; X. Huang; F. M. Raushel; H. M. Holden, *J. Biol. Chem.* **2002**, *277*, 39722–39727.
107. L. S. Mullins; F. M. Raushel, *J. Am. Chem. Soc.* **1999**, *121*, 3803–3804.
108. B. W. Miles; F. M. Raushel, *Biochemistry* **2000**, *39*, 5051–5056.
109. W. L. Delano, DeLano Scientific Software LLC: Palo Alto, CA, 2002.
110. H. M. Berman; J. Westbrook; Z. Feng; G. Gilliland; T. N. Bhat; H. Wessig; I. N. Shindyalov; P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.
111. P. M. Anderson; A. Meister, *Biochemistry* **1966**, *5*, 3157–3163.
112. G. E. Gibson; L. S. Mullins; F. M. Raushel, *Bioorg. Chem.* **1998**, *26*, 255–268.
113. W. W. Cleland; A. C. Hengge, *Chem. Rev.* **2006**, *106*, 3252–3278.
114. H. Nyunoya; C. J. Lusty, *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 4629–4633.
115. G. L. Waldrop; I. Rayment; H. M. Holden, *Biochemistry* **1994**, *33*, 10249–10256.
116. M. A. Stapleton; F. Javid-Majd; M. F. Harmon; B. A. Hanks; J. L. Grahmann; L. S. Mullins; F. M. Raushel, *Biochemistry* **1996**, *35*, 14352–14361.
117. J. B. Thoden; F. M. Raushel; M. M. Benning; I. Rayment; H. M. Holden, *Acta Crystallogr.* **1999**, *D55*, 8–24.
118. T. T. Wang; S. H. Bishop; A. Himoe, *J. Biol. Chem.* **1972**, *247*, 4437–4440.
119. J. N. S. Evans, *Biomolecular NMR Spectroscopy*; Oxford University Press: New York, 1995.
120. X. Huang; F. M. Raushel, *Biochemistry* **2000**, *39*, 3240–3247.
121. X. Huang; F. M. Raushel, *J. Biol. Chem.* **2000**, *275*, 26233–26240.
122. F. M. Raushel; L. S. Mullins; G. E. Gibson, *Biochemistry* **1998**, *37*, 10272–10278.
123. J. Kim; S. Howell; X. Huang; F. M. Raushel, *Biochemistry* **2002**, *41*, 12575–12581.
124. B. W. Miles; J. A. Banzon; F. M. Raushel, *Biochemistry* **1998**, *37*, 16773–16779.
125. O. A. Pierrat; F. Javid-Majd; F. M. Raushel, *Arch. Biochem. Biophys.* **2002**, *400*, 26–33.
126. B. L. Braxton; L. S. Mullins; F. M. Raushel; G. D. Reinhart, *Biochemistry* **1999**, *38*, 1394–1401.
127. P. M. Anderson; S. V. Marvin, *Biochem. Biophys. Res. Commun.* **1968**, *32*, 928–934.
128. A. Pierard, *Science* **1966**, *154*, 1572–1573.
129. B. W. Miles; J. A. Banzon; F. M. Raushel, *Biochemistry* **1998**, *37*, 16773–16779.
130. S. M. Mareya; F. M. Raushel, *Biochemistry* **1994**, *33*, 2945–2950.
131. J. B. Thoden; X. Huang; J. Kim; F. M. Raushel; H. M. Holden, *Protein Sci.* **2004**, *13*, 2398–2405.
132. J. L. Johnson; J. K. West; A. D. L. Nelson; G. D. Reinhart, *Biochemistry* **2007**, *46*, 387–397.
133. P. Alifano; R. Fani; P. Liò; A. Lazcano; M. Bazzicalupo; M. S. Carlomagno; C. B. Bruni, *Microbiol. Rev.* **1996**, *60*, 44–69.
134. T. J. Klem; Y. Chen; V. J. Davisson, *J. Bacteriol.* **2001**, *183*, 989–996.
135. T. J. Klem; V. J. Davisson, *Biochemistry* **1993**, *32*, 5177–5186.
136. H. Zalkin; J. E. Dixon, *Prog. Nucleic Acid Res. Mol. Biol.* **1992**, *42*, 259–287.
137. M. Y. Galperin; E. V. Koonin, *Mol. Microbiol.* **1997**, *24*, 443–445.
138. R. Fani; P. Liò; A. Lazcano, *J. Mol. Evol.* **1995**, *41*, 760–774.
139. B. N. Chauduri; S. C. Lange; R. S. Myers; S. V. Chittur; V. J. Davisson; J. L. Smith, *Structure* **2001**, *9*, 987–997.
140. S. Beismann-Driemeyer; R. Sterner, *J. Biol. Chem.* **2001**, *276*, 20387–20396.
141. S. V. Chittur; Y. Chen; V. J. Davisson, *Protein Expr. Purif.* **2000**, *18*, 366–377.
142. B. N. Chauduri; S. C. Lange; R. S. Myers; V. J. Davisson; J. L. Smith, *Biochemistry* **2003**, *42*, 7003–7012.
143. A. Douangamath; M. Walker; S. Beismann-Driemeyer; M. C. Vega-Fernandez; R. Sterner; M. Wilmanns, *Structure* **2002**, *10*, 185–193.

144. D. Lang; R. Thoma; M. Henn-Sax; R. Sterner; M. Wilmanns, *Science* **2000**, *289*, 1546–1550.
145. R. Omi; H. Mizuguchi; M. Goto; I. Miyahara; H. Hayashi; H. Kagamiyama; K. Hirotsu, *J. Biochem.* **2002**, *132*, 759–765.
146. R. Sterner; B. Höcker, *Chem. Rev.* **2005**, *105*, 4038–4055.
147. E. L. Wise; I. Rayment, *Acc. Chem. Res.* **2004**, *37*, 149–158.
148. S. V. Chittur; T. J. Klem; C. M. Shafer; V. J. Davisson, *Biochemistry* **2001**, *40*, 876–887.
149. J. Y. Tso; S. G. Bower; H. Zalkin, *J. Biol. Chem.* **1980**, *255*, 6734–6738.
150. R. S. Myers; J. R. Jensen; I. L. Deras; J. L. Smith; V. J. Davisson, *Biochemistry* **2003**, *42*, 7013–7022.
151. R. E. Amaro; R. S. Myers; V. J. Davisson; Z. A. Luthey-Schulten, *Biophys. J.* **2005**, *89*, 475–487.
152. R. Amaro; Z. Luthey-Schulten, *Chem. Phys.* **2004**, *307*, 147–155.
153. R. Amaro; E. Tajkhorshid; Z. Luthey-Schulten, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7599–7604.
154. C. Jarzynski, *Phys. Rev. E* **1997**, *56*, 5018–5035.
155. M. Karplus; G. A. Petsko, *Nature* **1990**, *347*, 631–639.
156. S. Park; F. Khalili-Araghi; E. Tajkhorshid; K. Schulten, *J. Chem. Phys.* **2003**, *119*, 3559–3566.
157. R. E. Amaro; A. Sethi; R. S. Myers; V. J. Davisson; Z. A. Luthey-Schulten, *Biochemistry* **2007**, *46*, 2156–2173.
158. R. S. Myers; R. E. Amaro; Z. A. Luthey-Schulten; V. J. Davisson, *Biochemistry* **2005**, *44*, 11974–11985.
159. F. Schendel; E. Mueller; J. Stubbe; A. Shiau; J. M. Smith, *Biochemistry* **1989**, *28*, 2459–2471.
160. H. C. Li; J. M. Buchanan, *J. Biol. Chem.* **1971**, *246*, 4720–4726.
161. A. A. Hoskins; R. Anand; S. E. Ealick; J. Stubbe, *Biochemistry* **2004**, *43*, 10314–10327.
162. D. J. Ebbole; H. Zalkin, *J. Biol. Chem.* **1987**, *262*, 8274–8287.
163. H. H. Saxild; P. Nygaard, *Microbiology* **2000**, *146*, 807–814.
164. R. Anand; A. A. Hoskins; E. M. Bennett; M. D. Sintchak; J. Stubbe; S. E. Ealick, *Biochemistry* **2004**, *43*, 10343–10352.
165. F. Schendel; J. Stubbe, *Biochemistry* **1986**, *25*, 2256–2264.
166. J. M. Buchanan, *Meth. Enzymol.* **1982**, *87*, 76–84.
167. C. L. Li; T. J. Kappock; J. Stubbe; T. M. Weaver; S. E. Ealick, *Structure* **1999**, *7*, 1155–1166.
168. K. Matsuda; T. Nishioka; K. Kinoshita; T. Kawabata; N. Go, *Protein Sci.* **2003**, *12*, 2239–2251.
169. J. L. Schrimsher; F. Schendel; J. Stubbe, *Biochemistry* **1986**, *25*, 4366–4371.
170. R. Anand; A. A. Hoskins; J. Stubbe; S. E. Ealick, *Biochemistry* **2004**, *43*, 10328–10342.
171. E. S. Rangarajan; A. Asinas; A. Proteau; C. Munger; J. Baardsnes; P. Iannuzzi; A. Matte; M. Cygler, *J. Bacteriol.* **2008**, *190*, 1447–1458.
172. X.-M. Xu; B. A. Carlson; R. Irons; H. Mix; N. Zhong; V. N. Gladyshev; D. L. Hatfield, *Biochem. J.* **2007**, *404*, 115–120.
173. E. Settembre; T. P. Begley; S. E. Ealick, *Curr. Opin. Struct. Biol.* **2003**, *13*, 739–747.
174. R. Batra; D. Christendat; A. Edwards; C. Arrowsmith; L. Tong, *Proteins* **2002**, *49*, 285–288.
175. M. Morar; R. Anand; A. A. Hoskins; J. Stubbe; S. E. Ealick, *Biochemistry* **2006**, *45*, 14880–14895.
176. K. Mizobuchi; J. M. Buchanan, *J. Biol. Chem.* **1968**, *243*, 4853–4862.
177. K. Mizobuchi; J. M. Buchanan, *J. Biol. Chem.* **1968**, *243*, 4842–4852.
178. D. R. Evans; H. I. Guy, *J. Biol. Chem.* **2004**, *279*, 33035–33038.
179. D. B. Ostrander; D. J. O'Brien; J. A. Gorman; G. M. Carman, *J. Biol. Chem.* **1998**, *273*, 18992–19001.
180. G. M. Hatch; G. McClarty, *J. Biol. Chem.* **1996**, *271*, 25810–25816.
181. A. C. Verschuur; A. H. van Gennip; E. J. Muller; P. A. Voute; A. B. van Kuilenburg, *Adv. Exp. Med. Biol.* **1998**, *431*, 667–671.
182. A. Fijolek; A. Hofer; L. Thelander, *J. Biol. Chem.* **2007**, *282*, 11858–11865.
183. A. Hofer; D. Steverding; A. Chabes; R. Brun; L. Thelander, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6412–6416.
184. D. A. Lewis; J. J. Villafranca, *Biochemistry* **1989**, *28*, 8454–8459.
185. A. Levitzki; D. E. Koshland, Jr., *Biochemistry* **1971**, *10*, 3365–3371.
186. C. W. Long; A. B. Pardee, *J. Biol. Chem.* **1967**, *242*, 4715–4721.
187. K. H. Scheit; H. J. Linke, *Eur. J. Biochem.* **1982**, *126*, 57–60.
188. A. Levitzki; D. E. Koshland, Jr., *Biochemistry* **1972**, *11*, 241–246.
189. M. Willmoës; B. W. Sigurskjold, *Eur. J. Biochem.* **2002**, *269*, 4772–4779.
190. W. L. Yang; V. M. McDonough; O. Ozier-Kalogeropoulos; M. T. Adeline; M. T. Flocco; G. M. Carman, *Biochemistry* **1994**, *33*, 10785–10793.
191. Y.-F. Chang; S. S. Martin; E. P. Baldwin; G. M. Carman, *J. Biol. Chem.* **2007**, *282*, 17613–17622.
192. M.-G. Choi; T.-S. Park; G. M. Carman, *J. Biol. Chem.* **2003**, *278*, 23610–23616.
193. P. M. Anderson, *Biochemistry* **1983**, *22*, 3285–3292.
194. J. A. Endrizzi; H. Kim; P. M. Anderson; E. P. Baldwin, *Biochemistry* **2004**, *43*, 6447–6463.
195. J. A. Endrizzi; H. S. Kim; P. M. Anderson; E. P. Baldwin, *Biochemistry* **2005**, *44*, 13491–13499.
196. M. Goto; R. Omi; N. Nakagawa; I. Miyahara; K. Hirotsu, *Structure* **2004**, *12*, 1413–1423.
197. P. Kursula; S. Flodin; M. Ehn; M. Hammarström; H. Schüler; P. Nordlund; P. Stenmark, *Acta Crystallogr.* **2006**, *F62*, 613–617.
198. H. Käck; J. Sandmark; K. J. Gibson; G. Schneider; Y. Lindqvist, *Protein Sci.* **1998**, *7*, 2560–2566.
199. H. Käck; K. J. Gibson; Y. Lindqvist; G. Schneider, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5495–5500.
200. M. Willmoës, *J. Biol. Chem.* **2003**, *278*, 9407–9411.
201. A. Iyengar; S. L. Bearne, *Biochem. J.* **2003**, *369*, 497–507.
202. S. L. Bearne; O. Hekmat; J. E. Macdonnell, *Biochem. J.* **2001**, *356*, 223–232.
203. K. P. Chakraborty; R. B. Hurlbert, *Biochim. Biophys. Acta* **1961**, *47*, 607–609.
204. M. Willemoës; A. Mølgaard; E. Johansson; J. Martinussen, *FEBS J.* **2005**, *272*, 856–864.
205. J. Whelan; G. Phear; M. Yamauchi; M. Meuth, *Nat. Genet.* **1993**, *3*, 317–322.
206. J. L. Wylie; L. L. Wang; G. Tipples; G. McClarty, *J. Biol. Chem.* **1996**, *271*, 15393–15400.
207. F. A. Lunn; S. L. Bearne, *Eur. J. Biochem.* **2004**, *271*, 4204–4212.
208. T. P. Begley, *Nat. Prod. Rep.* **2006**, *23*, 15–25.
209. D. E. Cane; S. Du; J. K. Robinson; Y. Hsiung; I. D. Spenser, *J. Am. Chem. Soc.* **1999**, *121*, 7722–7723.
210. M. Ehrenshaft; P. Bilski; M. Y. Li; C. F. Chignell; M. E. Daub, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9374–9378.

211. M. Tambasco-Studart; O. Titiz; T. Raschle; G. Forster; N. Amrhein; T. B. Fitzpatrick, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13687–13692.
212. T. Raschle; N. Amrhein; T. B. Fitzpatrick, *J. Biol. Chem.* **2005**, *280*, 32291–32300.
213. K. E. Burns; Y. Xiang; C. L. Kinsland; F. W. McLafferty; T. P. Begley, *J. Am. Chem. Soc.* **2005**, *127*, 3682–3683.
214. M. Ehrenshaft; M. E. Daub, *J. Bacteriol.* **2001**, *183*, 3383–3390.
215. M. Gengenbacher; T. B. Fitzpatrick; T. Raschle; K. Flicker; I. Sinning; S. Müller; P. Macheroux; I. Tews; B. Kappes, *J. Biol. Chem.* **2006**, *281*, 3633–3641.
216. K. Tazuya; Y. Adachi; K. Masuda; K. Yamada; H. Kumaoka, *Biochim. Biophys. Acta* **1995**, *1244*, 113–116.
217. B. R. Belitsky, *J. Bacteriol.* **2004**, *186*, 1191–1196.
218. Y.-X. Dong; S. Sueda; J.-I. Nikawa; H. Kondo, *Eur. J. Biochem.* **2004**, *271*, 745–752.
219. C. Wrenger; M.-L. Eschbach; I. B. Müller; D. Warnecke; R. D. Walter, *J. Biol. Chem.* **2005**, *280*, 5242–5248.
220. M. Tambasco-Studart; I. Tews; N. Amrhein; T. B. Fitzpatrick, *Plant Physiol.* **2007**, *144*, 915–925.
221. J. W. Hanes; K. E. Burns; D. G. Hilmey; A. Chatterjee; P. C. Dorrestein; T. P. Begley, *J. Am. Chem. Soc.* **2008**, *130*, 3043–3052.
222. T. Raschle; D. Arigoni; R. Brunisholz; H. Rechsteiner; N. Amrhein; T. B. Fitzpatrick, *J. Biol. Chem.* **2007**, *282*, 6098–6105.
223. J. W. Hanes; I. Keresztes; T. P. Begley, *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 2102–2105.
224. J. A. Bauer; E. M. Bennett; T. P. Begley; S. E. Ealick, *J. Biol. Chem.* **2004**, *279*, 2704–2711.
225. J. Zhu; J. W. Burgner; E. Harms; B. R. Belitsky; J. L. Smith, *J. Biol. Chem.* **2005**, *280*, 27914–27923.
226. M. Strohmeier; T. Raschle; J. Mazurkiewicz; K. Rippe; I. Sinning; T. B. Fitzpatrick; I. Tews, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19284–19289.
227. F. Zein; Y. Zhang; Y.-N. Kang; K. Burns; T. P. Begley; S. E. Ealick, *Biochemistry* **2006**, *45*, 14609–14620.
228. M. Neuwirth; K. Flicker; M. Strohmeier; I. Tews; P. Macheroux, *Biochemistry* **2007**, *46*, 5131–5139.
229. K. Flicker; M. Neuwirth; M. Strohmeier; B. Kappes; I. Tews; P. Macheroux, *J. Mol. Biol.* **2007**, *374*, 732–748.
230. H. Zalkin; L. H. Hwang, *J. Biol. Chem.* **1971**, *246*, 6899–6907.
231. R. L. Switzer, *Biofactors* **1989**, *2*, 77–86.
232. S. Chen; D. R. Tomchick; D. Wolle; P. Hu; J. L. Smith; R. L. Switzer; H. Zalkin, *Biochemistry* **1997**, *36*, 10718–10726.
233. J. Rudolph; J. Stubbe, *Biochemistry* **1995**, *34*, 2241–2250.
234. E. J. Mueller; E. Meyer; J. Rudolph; V. J. Davisson; J. Stubbe, *Biochemistry* **1994**, *33*, 2269–2278.
235. A. K. Bera; S. Chen; J. L. Smith; H. Zalkin, *J. Bacteriol.* **2000**, *182*, 3734–3739.
236. J. A. Grandoni; R. L. Switzer; C. A. Makaroff; H. Zalkin, *J. Biol. Chem.* **1989**, *264*, 6058–6064.
237. D. A. Bernlohr; R. L. Switzer, *Biochemistry* **1981**, *20*, 5675–5681.
238. F. W. Outten, *Nat. Chem. Biol.* **2007**, *3*, 206–207.
239. V. L. Schramm; C. Grubmeyer, *Prog. Nucleic. Acid Res. Mol. Biol.* **2004**, *78*, 261–304.
240. S. C. Sinha; J. L. Smith, *Curr. Opin. Struct. Biol.* **2001**, *11*, 733–739.
241. W. D. L. Musick, *Crit. Rev. Biochem.* **1981**, *11*, 1–34.
242. J. M. Krahn; J. H. Kim; M. R. Burns; R. J. Parry; H. Zalkin; J. L. Smith, *Biochemistry* **1997**, *36*, 11061–11068.
243. C. R. Muchmore; J. M. Krahn; J. H. Kim; H. Zalkin; J. L. Smith, *Protein Sci.* **1998**, *7*, 39–51.
244. J. L. Smith, *Curr. Opin. Struct. Biol.* **1998**, *8*, 686–694.
245. J. L. Smith; E. J. Zaluzec; J. P. Wery; L. Niu; R. L. Switzer; H. Zalkin; Y. Satow, *Science* **1994**, *264*, 1427–1433.
246. G. Scapin; D. H. Ozturk; C. Grubmeyer; J. C. Sacchettini, *Biochemistry* **1995**, *34*, 10744–10754.
247. D. R. Tomchick; R. J. Turner; R. L. Switzer; J. L. Smith, *Structure* **1998**, *6*, 337–350.
248. J. C. Eads; G. Scapin; Y. Xu; C. Grubmeyer; J. C. Sacchettini, *Cell* **1994**, *78*, 325–334.
249. R. J. Parry; K. Haridas, *Tetrahedron Lett.* **1993**, *34*, 7013–7016.
250. J. H. Kim; D. Wolle; K. Haridas; R. J. Parry; J. L. Smith; H. Zalkin, *J. Biol. Chem.* **1995**, *270*, 17394–17399.
251. S. Chen; J. W. Burgner; J. M. Krahn; J. L. Smith; H. Zalkin, *Biochemistry* **1999**, *38*, 11659–11669.
252. A. K. Bera; J. L. Smith; H. Zalkin, *J. Biol. Chem.* **2000**, *275*, 7975–7979.
253. L. J. Messenger; H. Zalkin, *J. Biol. Chem.* **1979**, *254*, 3382–3392.
254. S. Hammes-Schiffer; S. J. Benkovic, *Annu. Rev. Biochem.* **2006**, *75*, 519–541.
255. K. A. Henzler-Wildman; M. Lei; V. Thai; S. J. Kerns; M. Karplus; D. Kern, *Nature* **2007**, *450*, 913–916.
256. D. D. Boehr; H. J. Dyson; P. E. Wright, *Chem. Rev.* **2006**, *106*, 3055–3079.
257. A. K. Bera; S. Chen; J. L. Smith; H. Zalkin, *J. Biol. Chem.* **1999**, *274*, 36498–36504.
258. D. Mengin-Lecreulx; J. van Heijenoort, *J. Bacteriol.* **1994**, *176*, 5788–5795.
259. G. Reuter; H.-J. Gabius, *Cell. Mol. Life Sci.* **1999**, *55*, 368–422.
260. R. A. Dwek, *Chem. Rev.* **1996**, *96*, 683–720.
261. H. Barreteau; A. Kovač; A. Boniface; M. Sova; S. Gobec; D. Blanot, *FEMS Microbiol. Rev.* **2008**, *32*, 168–207.
262. J. van Heijenoort, *Cell. Mol. Life Sci.* **1998**, *54*, 300–304.
263. T. D. H. Bugg; C. T. Walsh, *Nat. Prod. Rep.* **1992**, *9*, 199–215.
264. S. Milewski, *Biochim. Biophys. Acta* **2002**, *1597*, 173–192.
265. C. Leriche; M.-A. Badet-Denisot; B. Badet, *J. Am. Chem. Soc.* **1996**, *118*, 1797–1798.
266. L. F. Hebert; M. C. Daniels; J. Zhou; E. D. Crook; R. L. Turner; S. T. Simmons; J. L. Neidigh; J. S. Zhu; A. D. Baron; D. A. McClain, *J. Clin. Invest.* **1996**, *98*, 930–936.
267. R. R. Traxinger; S. Marshall, *J. Biol. Chem.* **1991**, *266*, 10148–10154.
268. S. Marshall; V. Bacote; R. R. Traxinger, *J. Biol. Chem.* **1991**, *266*, 4706–4712.
269. S. Milewski; A. Janiak; M. Wojciechowski, *Arch. Biochem. Biophys.* **2006**, *450*, 39–49.
270. N. Floquet; C. Richez; P. Durand; B. Maigret; B. Badet; M.-A. Badet-Denisot, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1966–1970.
271. F. Massière; M.-A. Badet-Denisot; L. René; B. Badet, *J. Am. Chem. Soc.* **1997**, *119*, 5748–5749.
272. S. L. Bearne; C. Blouin, *J. Biol. Chem.* **2000**, *275*, 135–140.
273. M. Nakata; R. O'Rourke; S. Wilson; K. Chilson; C. P. Selitrennikoff, *J. Antibiot.* **2001**, *54*, 737–743.
274. A. Teplyakov; C. Leriche; G. Obmolova; B. Badet; M.-A. Badet-Denisot, *Nat. Prod. Rep.* **2002**, *19*, 60–69.
275. A. Teplyakov; G. Obmolova; M.-A. Badet-Denisot; B. Badet, *Protein Sci.* **1999**, *8*, 596–602.

276. B. Golinelli-Pimpaneau; F. Le Goffic; B. Badet, *J. Am. Chem. Soc.* **1989**, *111*, 3029–3034.
277. B. Badet; P. Vermoote; F. Le Goffic, *Biochemistry* **1988**, *27*, 2282–2287.
278. M.-A. Badet-Denisot; C. Leriche; F. Massière; B. Badet, *Bioorg. Med. Chem. Lett.* **1995**, *5*, 815–820.
279. J. Olchowy; K. Kur; P. Sachadyn; S. Milewski, *Protein Expr. Purif.* **2006**, *46*, 309–315.
280. S. Milewski; D. Kuszczak; R. Jedrzejczak; R. J. Smith; A. J. P. Brown; G. W. Gooday, *J. Biol. Chem.* **1999**, *274*, 4000–4008.
281. S. Milewski; M. Hoffmann; R. Andruszkiewicz; E. Borowski, *Bioorg. Chem.* **1997**, *25*, 283–296.
282. K. O. Broschat; C. Gorka; J. D. Page; C. L. Martin-Berger; M. S. Davies; H. C. Huang; E. A. Gulve; W. J. Salsgiver; T. P. Kasten, *J. Biol. Chem.* **2002**, *277*, 14764–14770.
283. G. L. McKnight; S. L. Mudri; S. L. Mathewes; R. R. Traxinger; S. Marshall; P. O. Sheppard; P. J. O'Hara, *J. Biol. Chem.* **1992**, *267*, 25208–25212.
284. C. Richez; J. Boetzel; N. Floquet; K. Koteshwar; J. Stevens; B. Badet; M.-A. Badet-Denisot, *Protein Expr. Purif.* **2007**, *54*, 45–53.
285. Q. K. Huynh; E. A. Gulve; T. Dian, *Arch. Biochem. Biophys.* **2000**, *379*, 307–313.
286. M.-A. Denisot; F. Le Goffic; B. Badet, *Arch. Biochem. Biophys.* **1991**, *288*, 225–230.
287. A. Teplyakov; G. Obmolova; M.-A. Badet-Denisot; B. Badet; I. Polikarpov, *Structure* **1998**, *6*, 1047–1055.
288. M. N. Isupov; G. Obmolova; S. Butterworth; M.-A. Badet-Denisot; B. Badet; I. Polikarpov; J. A. Littlechild; A. Teplyakov, *Structure* **1996**, *4*, 801–810.
289. B. Golinelli-Pimpaneau; B. Badet, *Eur. J. Biochem.* **1991**, *201*, 175–182.
290. W. C. Winkler; A. Nahvi; A. Roth; J. A. Collins; R. R. Breaker, *Nature* **2004**, *428*, 281–286.
291. J. C. Cochrane; S. V. Lipchock; S. A. Strobel, *Chem. Biol.* **2007**, *14*, 97–105.
292. M. Mandal; R. R. Breaker, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 451–463.
293. Y. Li; C. Roux; S. Lazereg; J.-P. LeCaer; O. Laprévote; B. Badet; M.-A. Badet-Denisot, *Biochemistry* **2007**, *46*, 13163–13169.
294. Y. Hu; L. Riesland; A. J. Paterson; J. E. Kudlow, *J. Biol. Chem.* **2004**, *279*, 29988–29993.
295. Q. Chang; K. Su; J. R. Baker; X. Yang; A. J. Paterson; J. E. Kudlow, *J. Biol. Chem.* **2005**, *280*, 21981–21987.
296. H. Graack; U. Cinque; H. Kress, *Biochem. J.* **2001**, *360*, 401–412.
297. A. Teplyakov; G. Obmolova; B. Badet; M.-A. Badet-Denisot, *J. Mol. Biol.* **2001**, *313*, 1093–1102.
298. S. Mouilleron; B. Golinelli-Pimpaneau, *Protein Sci.* **2007**, *16*, 485–493.
299. S. Mouilleron; M.-A. Badet-Denisot; B. Golinelli-Pimpaneau, *J. Mol. Biol.* **2008**, *377*, 1174–1185.
300. J. Raczynska; J. Olchwy; P. V. Konariev; D. I. Svergun; S. Milewski; W. Rypniewski, *J. Mol. Biol.* **2007**, *372*, 672–688.
301. J. Olchowy; R. Jedrzejczak; S. Milewski; W. Rypniewski, *Acta Crystallogr.* **2005**, *F61*, 994–996.
302. N. Floquet; S. Mouilleron; R. Daher; B. Maigret; B. Badet; M.-A. Badet-Denisot, *FEBS Lett.* **2007**, *581*, 2981–2987.
303. S. J. Temple; C. P. Vance; J. S. Gantt, *Trends Plant Sci.* **1998**, *3*, 51–56.
304. M. A. Vanoni; B. Curti, *Arch. Biochem. Biophys.* **2005**, *433*, 193–211.
305. M. A. Vanoni; L. Dossena; R. H. H. van den Heuvel; B. Curti, *Photosyn. Res.* **2005**, *83*, 219–238.
306. R. H. H. van den Heuvel; B. Curti; M. A. Vanoni; A. Mattevi, *Cell. Mol. Life Sci.* **2004**, *61*, 669–681.
307. M. A. Vanoni; D. E. Edmondson; M. Rescigno; G. Zanetti; B. Curti, *Biochemistry* **1991**, *30*, 11478–11484.
308. M. A. Vanoni; L. Nuzzi; M. Rescigno; G. Zanetti; B. Curti, *Eur. J. Biochem.* **1991**, *202*, 181–189.
309. S. Ravasio; B. Curti; M. A. Vanoni, *Biochemistry* **2001**, *40*, 5533–5541.
310. P. Agnelli; L. Dossena; P. Colombi; S. Mulazzi; P. Morandi; G. Tedeschi; A. Negri; B. Curti; M. A. Vanoni, *Arch. Biochem. Biophys.* **2005**, *436*, 355–366.
311. M. A. Vanoni; D. E. Edmondson; G. Zanetti; B. Curti, *Biochemistry* **1992**, *31*, 4613–4623.
312. M. A. Vanoni; E. Verzotti; G. Zanetti; B. Curti, *Eur. J. Biochem.* **1996**, *236*, 937–946.
313. M. A. Vanoni; F. Fischer; S. Ravasio; E. Verzotti; D. E. Edmondson; W. R. Hagen; G. Zanetti; B. Curti, *Biochemistry* **1998**, *37*, 1828–1838.
314. D. B. Knaff; M. Hirasawa, *Biochim. Biophys. Acta* **1991**, *1056*, 93–125.
315. J. A. Leigh; J. A. Dodsworth, *Annu. Rev. Microbiol.* **2007**, *61*, 349–377.
316. S. Ravasio; L. Dossena; E. Martin-Figueroa; F. J. Florencio; A. Mattevi; P. Morandi; B. Curti; M. A. Vanoni, *Biochemistry* **2002**, *41*, 8120–8133.
317. C. Binda; R. T. Bossi; S. Wakatsuki; S. Arzt; A. Coda; B. Curti; M. A. Vanoni; A. Mattevi, *Structure* **2000**, *8*, 1299–1308.
318. R. H. H. van den Heuvel; D. Ferrari; R. T. Bossi; S. Ravasio; B. Curti; M. A. Vanoni; F. J. Florencio; A. Mattevi, *J. Biol. Chem.* **2002**, *277*, 24579–24583.
319. R. H. H. van den Heuvel; D. I. Svergun; M. V. Petoukhov; A. Coda; B. Curti; S. Ravasio; M. A. Vanoni; A. Mattevi, *J. Mol. Biol.* **2003**, *330*, 113–128.
320. M. S. Murray; R. P. Holmes; W. T. Lowther, *Biochemistry* **2008**, *47*, 2439–2449.
321. Y. Lindqvist, *J. Mol. Biol.* **1989**, *209*, 151–166.
322. Z. X. Xia; F. S. Mathews, *J. Mol. Biol.* **1990**, *212*, 837–863.
323. R. Pickersgill; D. Smith; K. Worboys; J. Jenkins, *J. Biol. Chem.* **1998**, *273*, 24660–24664.
324. M. V. Petoukhov; D. I. Svergun; P. V. Konarev; S. Ravasio; R. H. H. van den Heuvel; B. Curti; M. A. Vanoni, *J. Biol. Chem.* **2003**, *278*, 29933–29939.
325. M. V. Petoukhov; D. I. Svergun, *Curr. Opin. Struct. Biol.* **2007**, *17*, 562–571.
326. M. Cotteville; E. Larquet; S. Jonic; M. V. Petoukhov; G. Caprini; S. Paravisi; D. I. Svergun; M. A. Vanoni; N. Boisset, *J. Biol. Chem.* **2008**, *283*, 8237–8249.
327. B. van Breukelen; A. Barendregt; A. J. R. Heck; R. H. H. van den Heuvel, *Rapid Commun. Mass Spectrom.* **2006**, *20*, 2490–2496.
328. V. M. Coiro; A. Di Nola; M. A. Vanoni; M. Aschi; A. Coda; B. Curti; D. Roccatano, *Protein Sci.* **2004**, *13*, 2979–2991.
329. A. Amadei; A. B. M. Linssen; H. J. C. Berendsen, *Proteins* **1993**, *17*, 412–425.
330. A. Amadei; M. A. Ceruso; A. Di Nola, *Proteins* **1999**, *36*, 419–424.
331. S. Hayward, *Proteins* **1999**, *36*, 425–435.
332. L. Dossena; B. Curti; M. A. Vanoni, *Biochemistry* **2006**, *46*, 4473–4485.
333. S. K. Boehlein; S. M. Schuster; N. G. J. Richards, *Biochemistry* **1996**, *35*, 3031–3037.
334. N. Verma; K. Kumar; G. Kaur; S. Anand, *Crit. Rev. Biotechnol.* **2007**, *27*, 45–62.

335. N. G. J. Richards; M. S. Kilberg, *Annu. Rev. Biochem.* **2006**, *75*, 629–654.
336. N. Su; X.-Y. Pan; M. Zhou; R. C. Harvey; S. P. Hunger; M. S. Kilberg, *Pediatr. Blood Cancer* **2008**, *50*, 274–279.
337. R. Chakrabarti; S. M. Schuster, *Int. J. Pediatr. Haematol./Oncol.* **1997**, *4*, 597–611.
338. J. A. Gutierrez; X.-Y. Pan; L. Koroniak; J. Hiratake; M. S. Kilberg; N. G. J. Richards, *Chem. Biol.* **2006**, *13*, 1339–1347.
339. H. M. Lam; K. T. Coschigano; I. C. Oliveira; R. Melo-Oliveira; G. M. Coruzzi, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1996**, *47*, 569–593.
340. M. B. Herrera-Rodriguez; J. M. Maldonado; R. Perez-Vicente, *J. Plant Physiol.* **2006**, *163*, 1061–1070.
341. V. D. Dang; M. Valens; M. Bolotin-Fukuhara; B. Daignan-Fornier, *Mol. Microbiol.* **1996**, *22*, 681–692.
342. M. A. Scofield; W. S. Lewis; S. M. Schuster, *J. Biol. Chem.* **1990**, *265*, 12895–12902.
343. L. J. Reitzer; B. Maganasik, *J. Bacteriol.* **1982**, *151*, 1299–1313.
344. I. M. Andrulis; J. Chen; P. N. Ray, *Mol. Cell Biol.* **1987**, *7*, 2435–2443.
345. S. Hongo; T. Sato, *Biochim. Biophys. Acta*, **1983**, *742*, 484–489.
346. C. A. Luehr; S. M. Schuster, *Arch. Biochem. Biophys.* **1985**, *237*, 335–346.
347. P. M. Mehlhaff; S. M. Schuster, *Arch. Biochem. Biophys.* **1991**, *284*, 143–150.
348. P. M. Mehlhaff; C. A. Luehr; S. M. Schuster, *Biochemistry* **1985**, *24*, 1104–1110.
349. R. S. Markin; C. A. Luehr; S. M. Schuster, *Biochemistry* **1981**, *20*, 7226–7232.
350. M. Ciustea; J. A. Gutierrez; S. E. Abbatiello; J. R. Eyler; N. G. J. Richards, *Arch. Biochem. Biophys.* **2005**, *440*, 18–27.
351. V. Fresquet; J. B. Thoden; H. M. Holden; F. M. Raushel, *Bioorg. Chem.* **2004**, *32*, 63–75.
352. A. R. Tesson; T. S. Soper; M. Ciustea; N. G. J. Richards, *Arch. Biochem. Biophys.* **2003**, *413*, 23–31.
353. S. K. Boehlein; J. D. Stewart; E. S. Walworth; R. Thirumoorthy; N. G. J. Richards; S. M. Schuster, *Biochemistry* **1998**, *37*, 13230–13238.
354. S. E. Rognes, *Phytochemistry* **1975**, *14*, 1975–1982.
355. S. Hongo; T. Sato, *Arch. Biochem. Biophys.* **1985**, *238*, 410–417.
356. K. K. Li; W. T. Beeson, IV; I. Ghiviriga; N. G. J. Richards, *Biochemistry* **2007**, *46*, 4840–4849.
357. H. Cedar; J. H. Schwartz, *J. Biol. Chem.* **1969**, *244*, 4122–4127.
358. N. Nakatsu; H. Kato; J. Oda, *Nat. Struct. Biol.* **1998**, *5*, 15–19.
359. S. K. Hinchman; S. Henikoff; S. M. Schuster, *J. Biol. Chem.* **1992**, *267*, 144–149.
360. P. Bork; E. V. Koonin, *Proteins* **1994**, *20*, 347–355.
361. T. Karlberg; R. Collins; S. van den Berg; A. Flores; M. Hammarström; M. Högbom; L. Holmberg-Schiavone; J. Uppenberg, *Acta Crystallogr.* **2008**, *D64*, 279–286.
362. C. T. Lemke; P. L. Howell, *Structure* **2001**, *9*, 1153–1164.
363. J. D. Mougous; D. H. Lee; S. C. Hubbard; M. W. Schelle; D. J. Vocadlo; J. M. Berger; C. R. Bertozzi, *Mol. Cell* **2006**, *21*, 109–122.
364. M. Sun; T. S. Leyh, *Biochemistry* **2006**, *45*, 11304–11311.
365. M. T. Miller; B. O. Bachmann; C. A. Townsend; A. C. Rosenzweig, *Nat. Struct. Biol.* **2001**, *8*, 684–689.
366. B. O. Bachmann; C. A. Townsend, *Biochemistry* **2000**, *39*, 11187–11193.
367. H. J. McNaughton; J. E. Thirkettle; Z. Zhang; C. J. Schofield; S. E. Jensen; B. Barton; P. Greaves, *Chem. Commun.* **1998**, 2325–2326.
368. D. G. Waterman; M. Ortiz-Lombardía; M. J. Fogg; E. V. Koonin; A. A. Antson, *J. Mol. Biol.* **2006**, *356*, 97–110.
369. E. G. Mueller; P. M. Palenchar, *Protein Sci.* **2000**, *8*, 2424–2427.
370. M. T. Miller; B. O. Bachmann; C. A. Townsend; A. C. Rosenzweig, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14752–14757.
371. K. H. Baggaley; A. G. Brown; C. J. Schofield, *Nat. Prod. Rep.* **1997**, *14*, 309–333.
372. B. O. Bachmann; R. Li; C. A. Townsend, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9082–9086.
373. R. E. Hurd; B. K. John, *J. Magn. Reson.* **1991**, *91*, 648–653.
374. H. Schnizer; S. K. Boehlein; J. D. Stewart; N. G. J. Richards; S. M. Schuster, *Meth. Enzymol.* **2002**, *354*, 260–271.
375. H. Schnizer; S. K. Boehlein; J. D. Stewart; N. G. J. Richards; S. M. Schuster, *Biochemistry* **1999**, *38*, 3677–3682.
376. R. G. Yount, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1975**, *43*, 1–56.
377. M. Ibba; D. Söll, *Annu. Rev. Biochem.* **2000**, *69*, 617–650.
378. M. Ibba; D. Söll, *Science* **1999**, *286*, 1893–1897.
379. S. Bilokapic; T. Maier; D. Ahel; I. Gruic-Sovulj; D. Söll; I. Weygand-Durasevic; N. Ban, *EMBO J.* **2006**, *25*, 2498–2509.
380. O. Kotik-Kogan; N. Moor; D. Tworowski; M. Safro, *Structure* **2005**, *13*, 1799–1807.
381. A. C. Dock-Bregeon; B. Rees; A. Torres-Larios; G. Bey; J. Caillet; D. Moras, *Mol. Cell* **2004**, *16*, 375–386.
382. J. M. Avis; A. R. Fersht, *Biochemistry* **1993**, *32*, 5321–5326.
383. J. Ling; H. Roy; M. Ibba, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 72–77.
384. G. Archontis; T. Simonson; D. Moras; M. Karplus, *J. Mol. Biol.* **1998**, *275*, 823–846.
385. M. E. Budiman; M. H. Knaggs; J. S. Fetrow; R. W. Alexander, *Proteins* **2007**, *68*, 670–689.
386. A. Schön; H. Hottinger; D. Söll, *Biochimie* **1988**, *70*, 391–394.
387. D. L. Tumbula; H. D. Becker; W.-Z. Chang; D. Söll, *Nature* **2000**, *407*, 106–110.
388. M. Ibba; A. W. Curnow; D. Söll, *Trends Biochem. Sci.* **1997**, *22*, 39–42.
389. C. Balg; J. L. Huot; J. Lapointe; R. Chênevert, *J. Am. Chem. Soc.* **2008**, *130*, 3264–3265.
390. J. L. Huot; C. Balg; D. Jahn; J. Moser; A. Emond; S. P. Blais; R. Chênevert; J. Lapointe, *Biochemistry* **2007**, *46*, 13190–13198.
391. S. F. Ataide; M. Ibba, *ACS Chem. Biol.* **2006**, *1*, 285–297.
392. L. Feng; K. Sheppard; D. Tumbula-Hansen; D. Söll, *J. Biol. Chem.* **2005**, *280*, 8150–8155.
393. K. Aghaiypour; A. Wlodawer; J. Lubkowski, *Biochemistry* **2001**, *40*, 5655–5664.
394. A. L. Swain; M. Jaskolski; D. Housset; J. K. Rao; A. Wlodawer, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1474–1478.
395. K. Sheppard; P. Akochy; J. C. Salazar; D. Söll, *J. Biol. Chem.* **2007**, *282*, 11866–11873.
396. J. Hyeok Kwak; K. Shin; J. Woo; M. Kyuong Kim; S. Kim, Jr.; S. Hyun Eon; K. Hong, *Mol. Cell* **2002**, *14*, 374–381.
397. S. Namgoong; K. W. Hong; S. Y. Lee, *J. Microbiol. Biotechnol.* **2001**, *11*, 251–258.
398. G. Raczniak; H. D. Becker; B. Min; D. Söll, *J. Biol. Chem.* **2001**, *276*, 45862–45867.
399. J. C. Salazar; R. Zuniga; G. Raczniak; H. Becker; D. Söll; O. Orellana, *FEBS Lett.* **2001**, *500*, 129–131.
400. A. W. Curnow; D. L. Tumbula; J. T. Pelaschier; B. Min; D. Söll, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12838–12843.
401. H. D. Becker; B. Min; C. Jacobi; G. Raczniak; J. Pelaschier; H. Roy; S. Klein; D. Kern; D. Söll, *FEBS Lett.* **2000**, *476*, 140–144.

402. C. Francklyn, *Proc. Natl. Acad. Sci.* U.S.A. **2003**, *100*, 9650–9652.
403. M. Bailly; M. Blaise; B. Lorber; H. D. Becker; D. Kern, *Mol. Cell* **2007**, *28*, 228–239.
404. F. C. Wedler; P. D. Boyer, *J. Biol. Chem.* **1972**, *247*, 984–992.
405. D. Jahn; Y.-C. Kim; Y. Ishino; M.-W. Chen; D. Söll, *J. Biol. Chem.* **1990**, *265*, 8059–8064.
406. A. Nakamura; M. Yao; S. Chimnaronk; N. Sakai; I. Tanaka, *Science* **2006**, *312*, 1954–1958.
407. E. Schmitt; M. Panvert; S. Blanquet; Y. Mechulam, *Structure* **2005**, *13*, 1421–1433.
408. K. Y. Horiuchi; M. R. Harpel; L. Shen; Y. Luo; K. C. Rogers; R. A. Copeland, *Biochemistry* **2001**, *40*, 6450–6457.
409. M. Petřek; P. Košinová; J. Koča; M. Otyepka, *Structure* **2007**, *15*, 1357–1363.
410. S. Sekine; O. Nureki; A. Shimada; D. G. Vassylyev; S. Yokoyama, *Nat. Struct. Biol.* **2001**, *8*, 203–206.
411. S. Eiler; A. C. Dock-Bregeon; L. Moulinier; J.-C. Thierry; D. Moras, *EMBO J.* **1999**, *18*, 6532–6541.
412. A. Schön; C. G. Kannangara; S. Gough; D. Söll, *Nature* **1988**, *331*, 187–190.
413. F. J. LaRiviere; A. D. Wolfson; O. C. Uhlenbeck, *Science* **2001**, *294*, 165–168.
414. K. Sheppard; D. Söll, *J. Mol. Biol.* **2008**, *377*, 831–844.
415. S. Namgoong; K. Sheppard; R. L. Sherrer; D. Söll, *FEBS Lett.* **2007**, *581*, 309–314.
416. D. Tumbula-Hansen; L. Feng; H. Toogood; K. O. Stetter; D. Söll, *J. Biol. Chem.* **2002**, *277*, 37184–37190.
417. L. N. Kinch; N. V. Grishin, *Curr. Opin. Struct. Biol.* **2002**, *12*, 400–408.
418. A. Javelle; D. Lupo; P. Ripoche; T. Fulford; M. Merrick; F. K. Winkler, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5040–5045.
419. R. N. Fong; K.-S. Kim; C. Yoshihara; W. B. Inwood; S. Kustu, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18706–18711.
420. S. Khademi; J. O'Connell, III; J. Remis; Y. Robles-Calmenares; L. J. W. Miercke; R. M. Stroud, *Science* **2004**, *305*, 1587–1594.
421. W. P. Jencks; J. Carriuolo, *J. Am. Chem. Soc.* **1960**, *82*, 1778–1786.
422. W. D. Cornell; P. Cieplak; C. I. Bayly; I. R. Gould; K. M. Merz, Jr.; D. M. Ferguson; D. C. Spellmeyer; T. Fox; J. W. Caldwell; P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
423. A. D. MacKerell, Jr.; D. Bashford; M. Bellott; R. L. Dunbrack, Jr.; J. D. Evanseck; M. J. Field; S. Fischer; J. Gao; H. Guo; S. Ha; D. Joseph-McCarthy; L. Kuchnir; K. Kuczera; F. T. K. Lau; C. Mattos; S. Michnick; T. Ngo; D. T. Nguyen; B. Prodhom; W. E. Reiher, III; B. Roux; M. Schlenkrich; J. C. Smith; R. Stote; J. Straub; M. Watanabe; J. Wiorkiewcz-Kuczera; D. Yin; M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
424. Y. Fan; L. Lund; L. Yang; F. M. Raushel; Y.-Q. Gao, *Biochemistry* **2008**, *47*, 2935–2944.
425. A. E. Nixon; S. J. Benkovic, *Protein Eng.* **1999**, *13*, 323–327.
426. A. E. Nixon; M. Ostermeier; S. J. Benkovic, *Trends Biotechnol.* **1998**, *16*, 258–264.
427. A. E. Nixon; M. S. Warren; S. J. Benkovic, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1069–1073.
428. S. N. Reuland; A. P. Vlasov; S. A. Krupenko, *Protein Sci.* **2006**, *15*, 1076–1084.
429. S. W. Michnick; P. H. Ear; E. N. Manderson; I. Remy; E. Stefan, *Nat. Rev. Drug Discov.* **2007**, *6*, 569–582.
430. J. D. Stevenson; S. J. Benkovic, *J. Chem. Soc. Perkin Trans. 2* 1483–1493.
431. J. N. Pelletier; K. M. Arndt; A. Plückthun; S. W. Michnick, *Nat. Biotechnol.* **1999**, *17*, 683–690.
432. S. J. Benkovic; G. G. Hammes; S. Hammes-Schiffer, *Biochemistry* **2008**, *47*, 3317–3321.
433. S. An; R. Kumar; E. D. Sheets; S. J. Benkovic, *Science* **2008**, *320*, 103–106.
434. K. Jørgensen; A. Vinther Rasmussen; M. Morant; A. Holm Nielsen; N. Bjarnholt; M. Zagrobelny; S. Bak; B. Lindberg Møller, *Curr. Opin. Plant Biol.* **2005**, *8*, 280–291.
435. L. A. Nogaj; S. I. Beale, *J. Biol. Chem.* **2005**, *280*, 24301–24307.
436. I. Morgunov; P. A. Srere, *J. Biol. Chem.* **1998**, *273*, 29540–29544.

Biographical Sketches

Nigel G. J. Richards was born in Leicester, England, and received his B.Sc. degree in chemistry from Imperial College, University of London (1980). He then moved to the University of Cambridge where he worked on the synthesis of aromatic natural products using novel alkyne chemistry under the direction of Ralph A. Raphael. After receiving his Ph.D., he carried out postdoctoral research with W. Clark Still at Columbia University while holding a Harkness Fellowship from the Commonwealth Fund of New York (1983–85).

During this period he was a member of the initial team that developed the MACROMODEL software package. In 1985 he assumed the post of lecturer in biological chemistry at Southampton University, and subsequently moved to the University of Florida in 1991. In recent years, his research has focused on the bioinorganic chemistry of Fe(III)- and Mn(II)-dependent enzymes together with studies of structure–function relationships in glutamine-dependent amidotransferases. His group is internationally known for their integrated use of experimental and computational methods to solve problems in mechanistic and structural enzymology. He is currently professor of chemistry at the University of Florida, where he was named a Distinguished Teaching Scholar in 2007.

Robert N. Humkey was born in Lubbock, TX, and grew up in the small town of Versailles, KY. In 1998, he began his undergraduate studies in biochemistry at the University of San Diego, and was awarded his B.S. degree in 2002. During his senior year, an internship with a local pharmaceutical company introduced him to research and he moved to the University of Florida to pursue a doctorate in computational biochemistry (2003). As a UF Alumni Fellow, he has been developing computational tools for use in discovering potent inhibitors for the enzyme asparagine synthetase, which have potential use as clinical agents for the treatment of acute lymphoblastic leukemia.

Kai Li was born in Zouping, China, and received his B.S. degree from Nankai University (1995). He then moved to Beijing where he became a graduate student in plant medicine at the Chinese Academy of Sciences. After obtaining his M.S. degree in 1998, he worked in the Chinese Institute of Veterinary Medicine and developed analytic methods for measuring drug residues in meat, milk, and eggs. After a short time spent at SUNY Albany, Kai moved to the University of Florida where he obtained a Ph.D. degree (2007) under the supervision of Professor Richards. During his doctoral research, he developed a new NMR-based assay for evaluating the efficiency of ammonia translocation in glutamine-dependent amidotransferases, and also showed the

important role of asparagine in regulating the undesired glutaminase activity of asparagine synthetase in cells. He is presently carrying out postdoctoral studies at the University of North Carolina with an emphasis on RNA biochemistry.

Megan E. Meyer was born in Lakewood, OH, and graduated from Seton Hall University with a B.S. in biochemistry in 2005. While at Seton Hall University, she worked on carbohydrate chemistry under the direction of Dr. Cecilia Marzabadi, and was a 'star' pitcher on the NCAA softball team. Choosing science over sport, she joined the graduate program in chemistry at the University of Florida in August 2005, and is now working on the design and characterization of novel asparagine synthetase inhibitors with the guidance of Professor Richards. These compounds have potential for use in improved clinical protocols against acute lymphoblastic leukemia. In addition to developing a series of new assays for use in library screening experiments, Megan is also involved in experiments to obtain crystallographic 'snapshots' of asparagine synthetase as it proceeds through its catalytic cycle.

Tania C. Córdova de Sintjago was born in Caracas, Venezuela, and was awarded her M.S. (1982) and Ph.D. (1991) degrees in chemistry from the Universidad Central de Venezuela (UCV). She then carried out postdoctoral studies in computational chemistry and drug discovery, working under the direction of Dr. Judith Hempel (Biosym Technologies, Inc). This work was supported by an INTEVEP fellowship (Petroleos de Venezuela). She presently holds the rank of professor in the Department of Chemistry at UCV, where her research interests include physical organic chemistry, reaction mechanism, theoretical calculations on gas-phase organic reactions, enzyme catalysis, and QSAR. She has participated as invited lecturer and discussion leader in a number of international events. During her sabbatical year at the University of Florida, Dr. Córdova de Sintjago carried out a series of computational investigations into the structure and catalytic mechanism of glutamine-dependent amidotransferases.

8.07 Fatty Acid Biosynthesis and Oxidation

Huaning Zhang, Carl A. Machutta, and Peter J. Tonge, Stony Brook University, Stony Brook, NY, USA

8.07.1 Introduction

This chapter focuses on the catalytic transformations that result in the cyclic biosynthesis and breakdown of fatty acids. These metabolic pathways will serve as a paradigm for three classes of chemical reactions: carbon–carbon bond formation and cleavage, oxidation and reduction, and hydration–dehydration. The most extensively studied reactions are those involved in microbial fatty acid biosynthesis (Type II fatty acid synthase (FAS-II)) and mammalian fatty acid β-oxidation. In both pathways, the reactions are catalyzed by separate enzymes that have been cloned and overexpressed, thus providing a ready source of material for structural and mechanistic studies. In contrast, mammalian fatty acid biosynthesis and microbial fatty acid breakdown are catalyzed by multifunctional enzymes (MFEs) that have historically been less amenable to analysis.

The cyclic series of reactions that result in the synthesis and breakdown of fatty acids are shown in **Figure 1**. Fatty acid biosynthesis, exemplified by the pathway from *Escherichia coli*, is initiated by the condensation of malonyl-acyl carrier protein (ACP) with acetyl-CoA by the β-ketoacyl-ACP synthase (KAS) FabH to generate β-ketobutyryl-ACP. The β-keto group is subsequently reduced by the NADPH-dependent β-ketoacyl-ACP reductase (KAR) FabG followed by the elimination of H_2O catalyzed by FabZ or FabA. The crotonyl-ACP that is formed is reduced by the NADH-dependent FabI and subsequent rounds of elongation are initiated by the action of the KASs FabD and FabF. Enzymes in the fatty acid β-oxidation pathway catalyze the reverse reactions, a major difference being that the acyl groups are carried by CoA rather than ACP. Fatty acid breakdown is initiated by the action of the flavoenzyme acyl-CoA dehydrogenase (ACD), which oxidizes saturated acyl-CoAs to enoyl-CoAs. This reaction is followed by a hydration, catalyzed by enoyl-CoA hydratase (ECH), followed by a second oxidation, catalyzed by β-hydroxyacyl-CoA dehydrogenase (HAD), to generate a β-ketoacyl-CoA. The final step in the pathway is catalyzed by thiolase and involves carbon–carbon bond cleavage leading to the formation of acetyl-CoA and a saturated acyl-CoA that is two carbon atoms shorter.

In discussions on the mechanisms of the enzymes involved in each pathway, there will be a particular focus on three superfamilies: enzymes that share the thiolase fold and catalyze carbon–carbon bond formation and cleavage; reactions catalyzed by NAD(P)-dependent enzymes in the fatty acid biosynthesis pathway involve proteins that are members of the short-chain dehydrogenase reductase (SDR) superfamily; and finally there are mechanistic parallels between the hydration and dehydration reactions in each pathway with a particular focus on the crotonase superfamily.

8.07.2   Thiolase Enzyme Homologs

### 8.07.2.1   Fatty Acid Biosynthesis: Carbon–Carbon Bond Formation

Carbon–carbon bond formation is a reaction of fundamental importance to the cellular metabolism of all living systems and includes alkylation reactions involving one and five carbon fragments as well as carboxylation reactions. In addition, a very common method of generating carbon–carbon bonds in biology includes the reactions of enolates and their equivalents (such as enamines) with aldehydes, ketones, keto acids, and esters. Reactions in which the enolate derives from an acyl thioester are Claisen condensations, whereas the remainder are classified as aldol reactions.

8.07.2.2   Claisen Condensation Reactions and the Thiolase Superfamily

Claisen condensation reactions are performed by enzymes that are members of the thiolase superfamily based on a three-dimensional fold first characterized in a degradative thiolase from *Saccharomyces cerevisiae*.[1,2] These enzymes primarily form dimers, with each subunit sharing a common superfamily topology,[3] whereas tetramers are observed only in the biosynthetic thiolase subfamily.[4] The monomer is composed of three domains, two core domains, each consisting of a mixed five-stranded β-sheet covered on each face by α-helices, and a loop domain. The two core domains pack together so that the overall fold of the monomer is a five-layered $\alpha\beta\alpha\beta\alpha$ structure. This is shown in **Figure 2(a)** for the KAS II enzyme from *Escherichia coli* (ecFabF). All members of the family have at least one catalytically essential cysteine that becomes covalently modified during the reaction. This cysteine is in the N-terminal domain and lies at the N-terminus of an α-helix (**Figures 2(b)–2(d)**), whereas all other catalytic residues are normally contained within the C-terminal domain.

Members of the superfamily may be subdivided based on either function or structural homology and play critical roles in polyketide and isoprenoid biosyntheses, fatty acid biosynthesis, and fatty acid degradation. Wierenga and coworkers[4] have discussed structural features that group thiolase family members into three subfamilies. In the first group are KAS I and KAS II, which initiate each cycle of elongation in fatty acid biosynthesis by condensing a C2 carbon unit derived from malonyl-ACP with the growing fatty acid (**Figure 3**). In addition to the catalytic cysteine, the two other conserved active site residues in the KAS I and KAS II enzymes are the histidine residues (**Figure 2(c)**).

Although most bacteria possess only one elongating KAS, which is generally more similar to KAS II than to KAS I, organisms such as *E. coli* and *Mycobacterium tuberculosis* possess two elongating KAS enzymes. In *E. coli* the
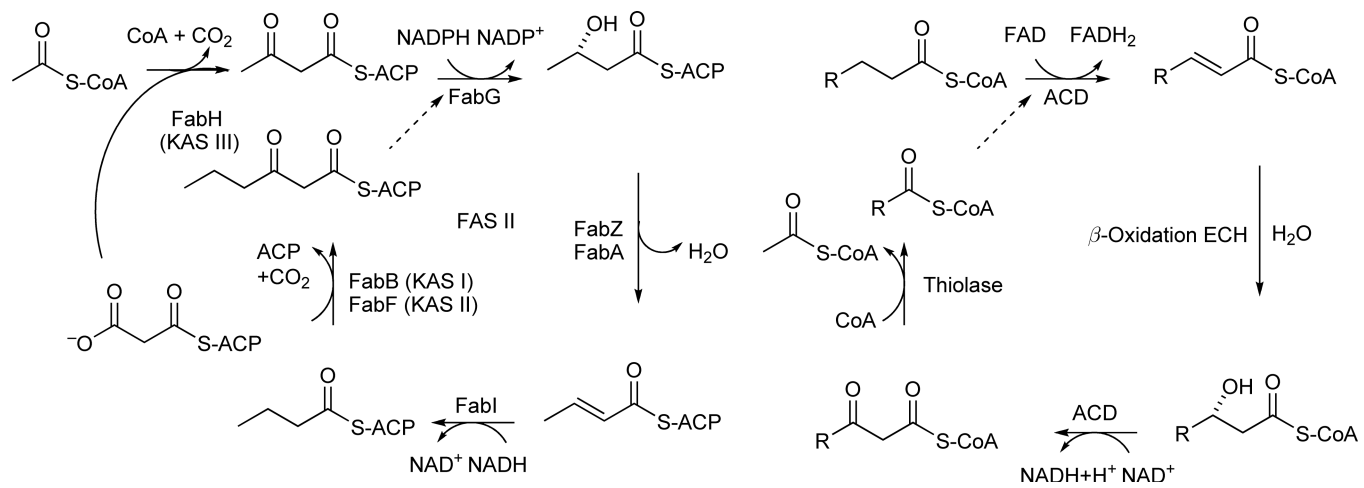
**Figure 1** Fatty acid biosynthesis and oxidation.
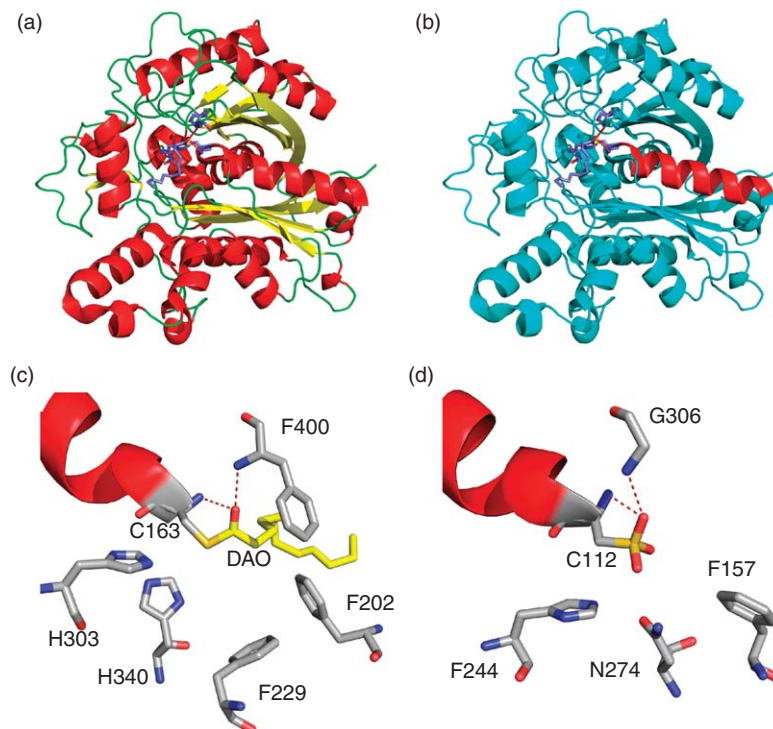
Figure 2 Structures of ecFabF (KAS II) and ecFabH (KAS III). (a) Structure of the ecFabF monomer showing the classic thiolase fold (helices, red; sheets, yellow; loops, green). (b) Structure of the ecFabF monomer (cyan) showing the active site α-helix (red). (c) Active site of ecFabF showing the catalytic residues and oxyanion hole. The catalytic cysteine (C163) is acylated with a C_{12} fatty acid (DAO, yellow) and lies at the N-terminus of an α-helix. (d) Active site of ecFabH showing the catalytic residues and oxyanion hole. Note that C112 in ecFabH has been oxidized to a sulfonic acid. The figures were made using PyMOL[37] and the structures of ecFabF from 2gfy.pdb[26] and ecFabH from 1mzs.pdb.[28]

KAS I and KAS II enzymes are FabB and FabF, whereas in *M. tuberculosis* they are KasA and KasB. FabB is essential for cell viability and required for elongating saturated fatty acids, whereas FabF is nonessential and widely held as the enzyme required for the temperature-dependent regulation of fatty acid biosynthesis.[5–7] In *M. tuberculosis*, KasA and KasB have overlapping substrate specificity with KasB possibly being responsible for the synthesis of very long-chain fatty acids (C54). Both KasA and KasB are important for the growth of *M. tuberculosis in vivo.*[8–10]

The second group contains the Type III KAS, which is the initiating enzyme for fatty acid biosynthesis, as well as the bacterial and plant Type III polyketide synthase (PKS III), such as chalcone synthase and stilbene synthase. 3-Hydroxy-3-methylglutaryl (HMG)-CoA synthase also falls into this category, in which all enzymes have a catalytic triad of Cys-His-Asn (**Figure 2(d)**). The final group of enzymes is the biosynthetic and degradative thiolases that have a Cys-Asn-His catalytic triad together with a second cysteine in the active site.

8.07.2.3   Mechanistic Challenges: KAS I–KAS III

The mechanistic challenges faced by the thiolase enzymes include acyl transfer to the active site cysteine, stabilization of oxyanion transition states during each step of the reaction, and generation of the required carbanion nucleophile in the second substrate. In each case the catalytic cysteine lies at the N-terminus of an α-helix (**Figure 2(b)**), and, by analogy to proposals advanced in the discussions of enzyme catalysis and protein structure,[11–16] it has been suggested that the helix dipole plays a central role in increasing the nucleophilicity of the cysteine and also providing assistance via transition state stabilization.[17] For enzymes such as KAS I–KAS III that utilize malonyl-ACP as the second substrate, an acetyl-ACP carbanion must be generated by
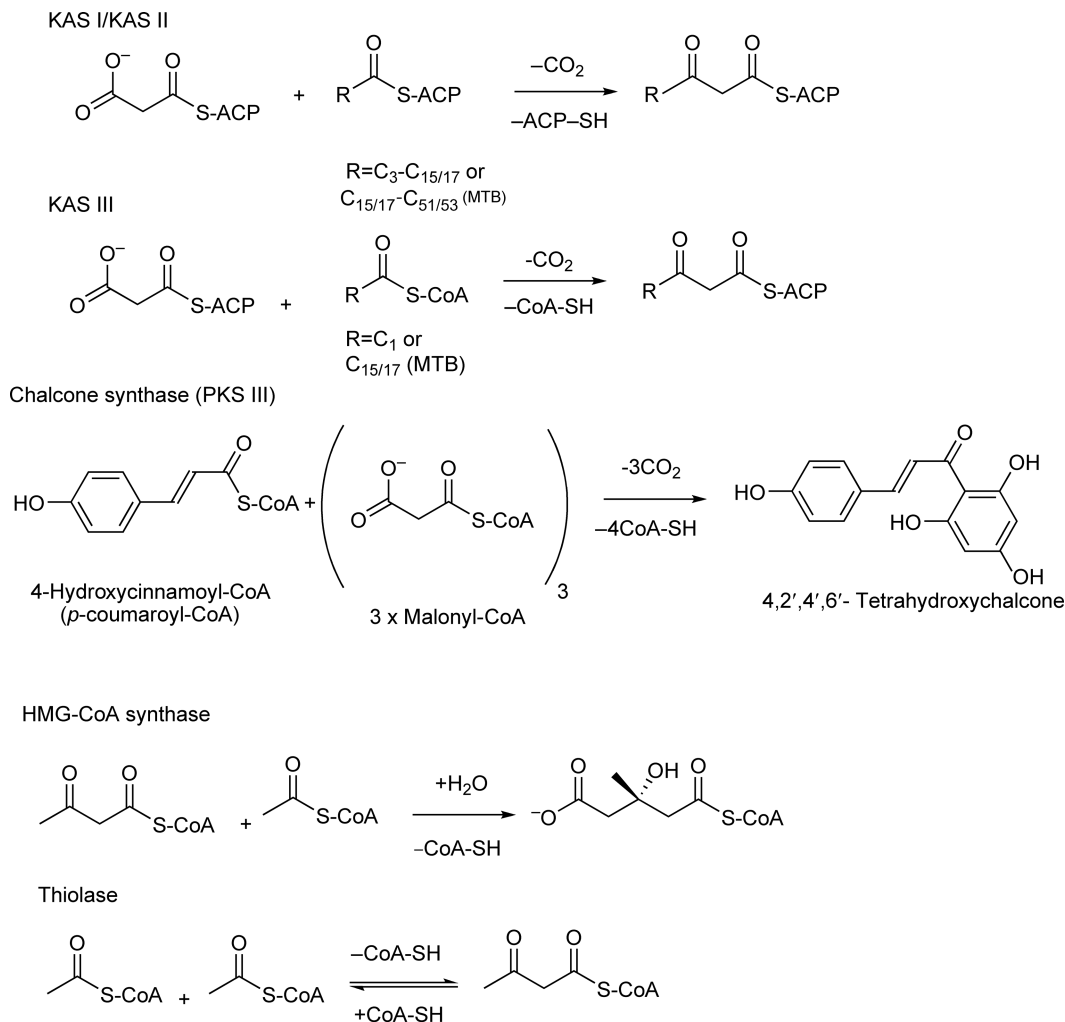
Figure 3   Reactions in the thiolase superfamily.

decarboxylation. In KAS I and KAS II, this is achieved by the two histidine residues whereas in KAS III the catalytic His and Asn are thought to be involved (**Figure 2**).

In the first half of the ping-pong bi–bi reaction, the active site cysteine is acylated by either an acyl-ACP (KAS I and KAS II) or an acetyl-CoA (KAS III), or in the case of the *M. tuberculosis* KAS III (FabH) a long-chain acyl-CoA. Attack of the active site cysteine thiolate on the donor acyl thioester is aided by the dipole of the active site helix and an oxyanion hole composed of two backbone NH groups. The additional catalytic residues (His, His or His, Asn) are thought to primarily function in the decarboxylation of malonyl-ACP and stabilization of the acetyl-ACP carbanion that is formed. The carbanion subsequently attacks the acyl-enzyme thioester leading to the formation of the β-ketoacyl-ACP product.

The proposed reaction mechanism for the KAS III enzymes is shown in **Figure 4**. Given the similarities of the KAS enzymes with cysteine proteases, a potentially surprising aspect of this mechanism is that the histidine adjacent to the cysteine is not involved in the initial acylation reaction, for example, by functioning as a general base during attack of the cysteine on the acyl-ACP or acyl-CoA. However, Rock and coworkers have examined the mechanism of the KAS enzymes in detail and have shown that mutation of the conserved histidine and asparagine residues in ecFabH (H244A and N274A) does not affect the rate of acylation.[17] Instead, the histidine and asparagine (or His/His) residues function primarily in the condensation half of the reaction in which a C–C bond is formed between the malonyl-ACP C2 carbon and the acyl-enzyme carbonyl group. In principle,
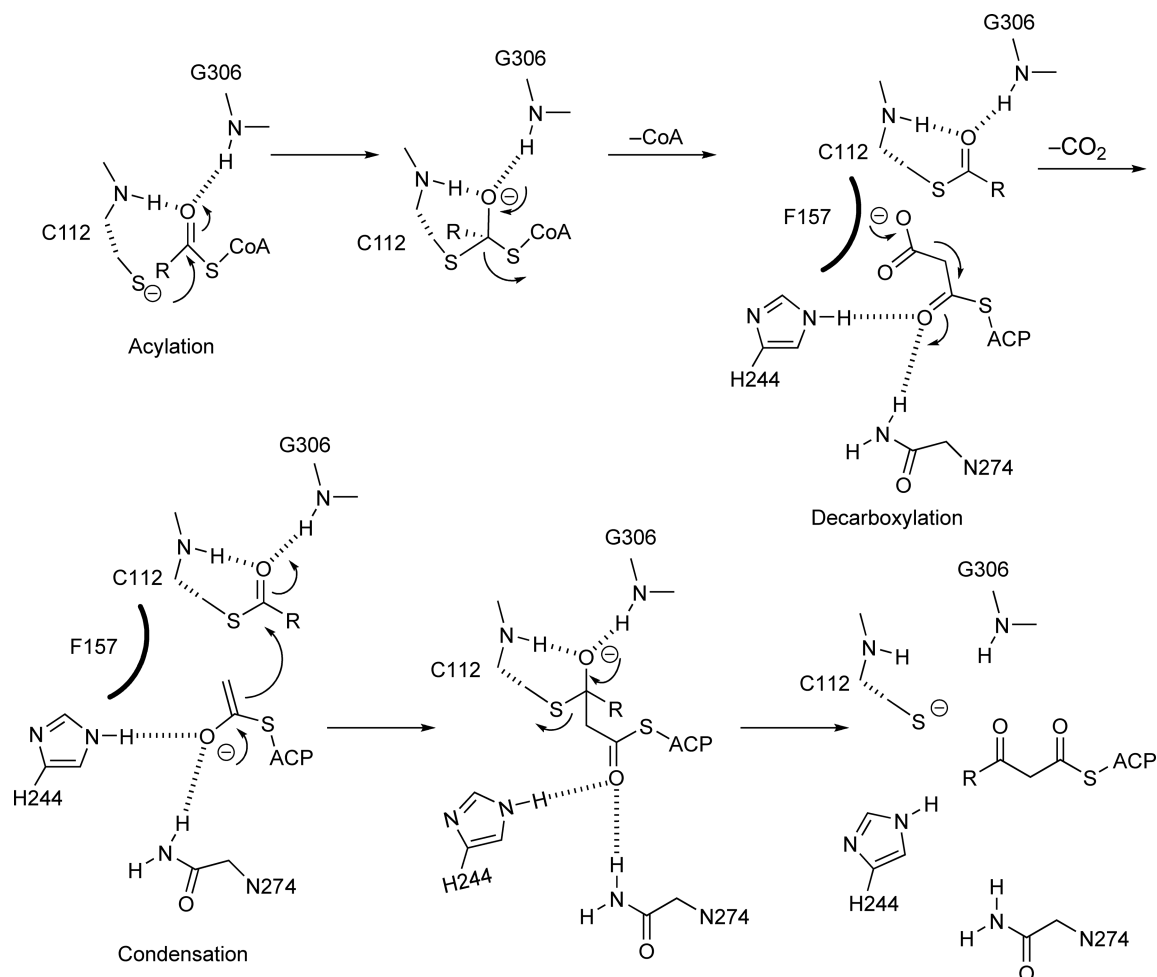
Figure 4 Reaction mechanism for the KAS III enzymes. For the KAS I and KAS II enzymes, N274 is replaced by a histidine (H340 in ecFabF).

decarboxylation and C–C bond formation can proceed through a concerted reaction; however, it is now generally thought that malonyl-ACP decarboxylates to generate a resonance-stabilized carbanion that subsequently reacts with the acyl-enzyme (**Figure 4**). This hypothesis is supported by the observation that mutagenesis (C112S in ecFabH) or modification of the catalytic cysteine actually accelerates the rate of decarboxylation.[7,17] In the mechanism shown in **Figure 4**, both the histidine and asparagine are hydrogen bonded to the malonyl-ACP thioester carbonyl and stabilize the accumulation of negative charge on the carbonyl during decarboxylation and carbanion formation. This process is thought to be assisted by F157, which destabilizes the negative charge on the malonyl carboxylate.

The mechanism for the KAS I and KAS II enzymes is thought to parallel that proposed for the KAS III enzymes, with H340 (ecFabF numbering) fulfilling the role played by N274 in KAS III. Mutagenesis of either histidine does not severely impact acylation, supporting the idea that the active site cysteine is already in the nucleophilic (deprotonated) form.[18] However, there is some debate as to the role played by the additional histidine (H303 in ecFabF). Rock and coworkers have shown that both the N^{δ}-1 and N^{ϵ}-1 of H304 (*Streptococcus pneumoniae* numbering; spFabF) are hydrogen bonded to structured water molecules (WATs) and have proposed that N^{ϵ}-1 acts as a base to assist the attack of water on the malonyl carboxylate, leading to carbanion formation through the elimination of bicarbonate. von Wettstein-Knowles and coworkers[18] also suggest that H304 (H298 in ecFabB, the KAS I from *E. coli*) acts as a base, but instead directly deprotonates the malonyl

carboxylate in the course of catalyzing the decarboxylation reaction. The proposed mechanism of decarboxylation for KAS I/KAS II differs fundamentally to that for KAS III (and also thiolase homologs such as chalcone synthase), where the negative charge on the malonyl group is destabilized by hydrophobic groups such as F157 in ecFabH. Interestingly, we note that KAS I and KAS II enzymes have two conserved phenylalanines close to the position occupied by F157 in KAS III, raising the possibility that F202 and F229 (**Figure 2(c)**) could be involved in destabilizing the negative charge on the malonyl carboxylate. Finally, there has also been discussion in the literature concerning other conserved residues including a lysine (K335) in the elongation enzymes that is close to H340 and participates in a hydrogen-bonding network with H303. Mutagenesis of K335 likely affects both the positioning and ionization state(s) of the histidine(s) and principally impacts malonyl-ACP decarboxylation and condensation. K335 mutations stabilize the KAS acyl-enzyme intermediate leading to its use in the preparation of stable complexes for structural studies.[19] Stable acyl-enzymes have also been prepared by mutation of the active site cysteine to a serine.[18]

## 8.07.2.4   Substrate and Inhibitor Recognition by the KAS Enzymes

The FAS-II pathway is a validated target for antimicrobial drug discovery,[20,21] and inhibitor development programs have focused primarily on the KAS enzymes[22] as well as the enoyl-ACP reductase that catalyzes the last step in the elongation reaction.[21,23] Inhibitors of the KAS enzymes include the natural products cerulenin,[24] thiolactomycin,[25] and platensimycin,[26] as well as synthetic molecules such as the alkyl-CoA disulfides developed by Reynolds and coworkers[27] and the substituted indoles developed at GlaxoSmithKline (GSK) (**Figure 5**).[28] The interaction of these compounds with the KAS enzymes has been studied using both kinetic and structural methods, providing information not only for additional inhibitor design but also shedding light on molecular features important for substrate recognition.

The active sites of KAS enzymes are composed of two channels, one that binds the acyl group and the other that accommodates the incoming pantetheine from ACP or CoA. This can clearly be seen in **Figures 6(a) and 6(b)** in which the structures of ecFabF complexed with dodecanoic acid and platensimycin have been overlaid. Platensimycin is a recently discovered natural product isolated from *Streptomyces platensis* that inhibits the KAS I and KAS II enzymes, binding more tightly to the acyl-enzyme form of the enzyme. The dihydroxybenzoate moiety interacts with the two active site histidines, binding in the location normally occupied by the malonyl group, whereas the remainder of the molecule extends into the pantetheine-binding channel.[26] Although platensimycin is a selective KAS I–KAS II inhibitor, the Merck group has also isolated the related compound platencin (**Figure 5**), which inhibits KAS III in addition to KAS I and KAS II.[29]

In addition to platensimycin and platencin, other natural product KAS inhibitors include thiolactomycin and cerulenin. Thiolactomycin binds to KAS I–KAS II in a similar location to that occupied by the
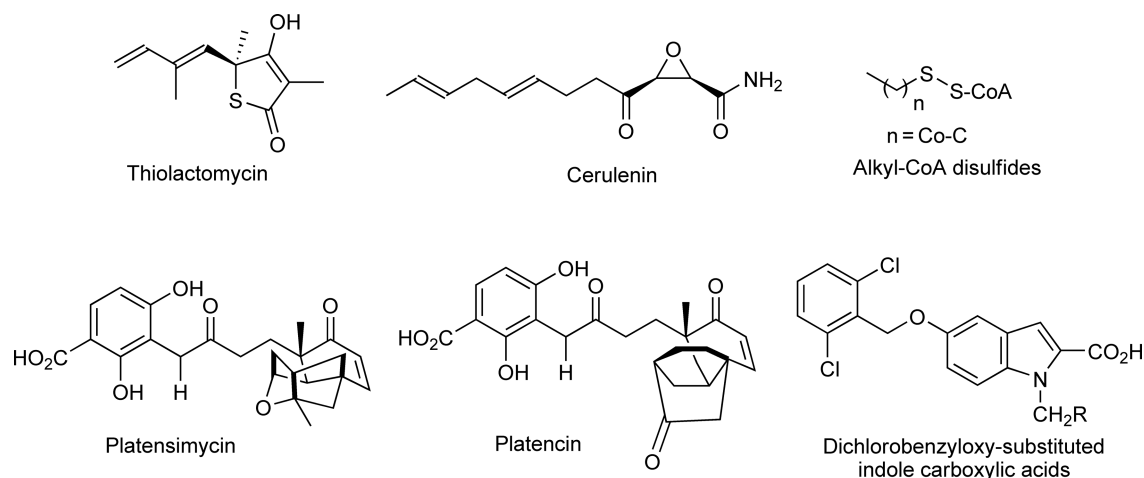
Figure 5   KAS inhibitors.
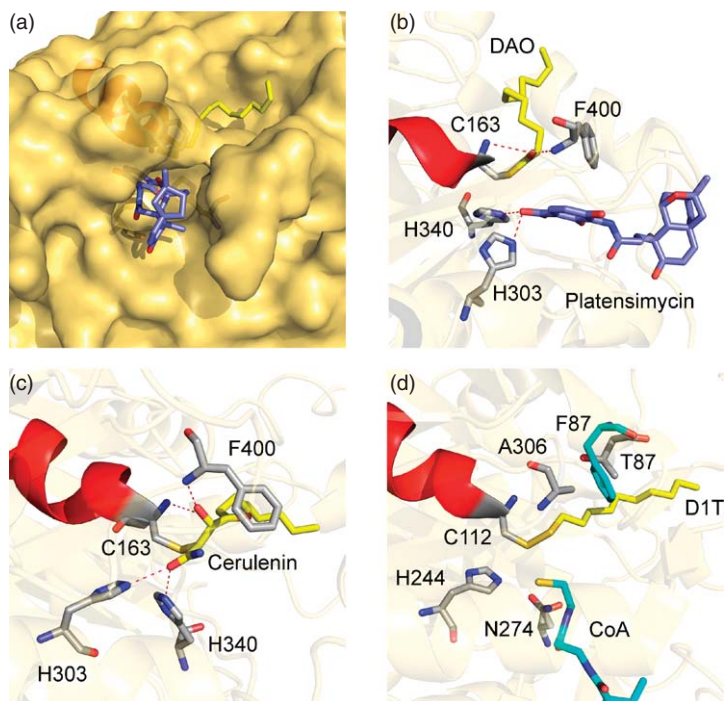
Figure 6 Substrate and inhibitors bound to KAS I–KAS III. (a) Surface representation of ecFabF in which the catalytic cysteine (C163) is acylated with a C_{12} fatty acid (DAO, yellow; 2gfy.pdb[26]). The structure of ecFabF C163Q complexed with platensimycin (2gfx.pdb[26]) has been superimposed in order to show the inhibitor (slate) bound in the pantetheine-binding channel. (b) Active site representation of the structures from a. The aromatic ring of platensimycin binds where the malonyl group is expected to bind. The C_{10} acyl group is located in the acyl-binding channel. (c) Structure of cerulenin bound to ecFabF taken from 1b3n.pdb.[31] (d) The active site of ecFabH inhibited by methanethiol and complexed with CoA (2eft.pdb[27]) superimposed on the structure of the FabH from *M. tuberculosis* (mtFabH) inhibited by a C_{10} alkyl disulfide (D1T, yellow; 2qo0.pdb[34]). Residues in gray are for mtFabH whereas molecules in cyan are for ecFabH. Note that F87 in ecFabH is important for chain length specificity and ensures that only acetyl-CoA can acylate C112. In mtFabH, F87 is replaced by a threonine permitting longer acyl chains to bind. The figure was made using PyMOL.[37]

dihydroxybenzoyl group of platensimycin, and is thus a competitive inhibitor of malonyl-ACP.[30] Although thiolactomycin and platensimycin are noncovalent KAS inhibitors, cerulenin, a mycotoxin produced by the fungus *Cephalosporium caerulens*, is a covalent inhibitor and alkylates the active site cysteine through an epoxide ring-opening reaction.[30,31] Both the hydroxyl groups generated from the epoxide oxygen as well as the cerulenin amide group make hydrogen-bonding interactions in the active site, whereas the nonadiene inhibitor tail occupies the hydrophobic acyl group channel (**Figure 6(c)**).[30,31] Cerulenin inhibits the KASs in both the Type I and Type II FASs. However, while cerulenin inhibits the KAS I and KAS II enzymes in the FAS-II pathway, it does not inhibit the KAS III enzymes since the nonadiene tail cannot be accommodated in the acyl-binding pocket of these enzymes, which are specific for acetyl groups.

As noted above, a major difference between the priming (KAS III) and elongation (KAS I and KAS II) enzymes is that the substrate for acylation is acetyl-CoA and not an acyl-ACP. In this regard, the *M. tuberculosis* (MTB) KAS III (FabH) is unique in that it has evolved specificity for longer chain substrates (C_{16}/C_{18}) rather than using the typical acetyl-CoA primer.[32,33] Mycobacteria have a requirement for very long-chain (C_{54}–C_{56}) fatty acids that are used to synthesize mycolic acids, C_{60}–C_{90} components of the cell wall. Consequently, mycobacteria utilize a Type I FAS (FAS-I), such as that found in eukaryotes, for the *de novo* synthesis of C_{16}/C_{18} fatty acyl-CoAs that are then elongated by the mycobacterial FAS-II system. Crystal structures from Wright, Reynolds, and coworkers have elucidated the molecular basis for substrate specificity in the mycobacterial KAS III, and have shown that a phenylalanine limits the size of the acyl-binding pocket in ecFabH while this residue is replaced by a threonine in mtFabH.[27,34] In the course of their studies, these authors have

developed several classes of FabH inhibitors including a series of alkyl disulfides.[27,35,36] In **Figure 6(d)**, the structure of ecFabH complexed with CoA has been overlaid with that of mtFabH inhibited by a C_{10} alkyl disulfide, showing how F87 in ecFabH would prevent binding of longer chain acyl groups. Finally, the recent structure of the MTB KasA enzyme in complex with a PEG molecule provides additional insight into the ability of the mycobacterial KAS enzymes to bind very long chain fatty acids.[38]

8.07.2.5   PKS III and HMG-CoA Synthase

Type I and Type II PKSs catalyze multiple rounds of reactions by catalytic modules encoded either by a single polypeptide (PKS I) or on separate polypeptides (PKS II) by analogy to FAS-I and FAS-II. In contrast, PKS IIIs are dimers of KASs that catalyze multiple condensation reactions in one active site and include chalcone synthase, stilbene synthase, and 2-pyrone synthase[32,39–41] (see Chapters 1.05, 1.07, and 1.04). In the case of chalcone synthase, three consecutive condensation reactions each utilizing malonyl-CoA, followed by a cyclization reaction, lead to the formation of 4, 2′,4′,6′-tetrahydroxychalcone from 4-hydroxycinnamoyl-CoA (**Figure 3**). Recruitment of a reductase leads to the formation of a product lacking the 6′-hydroxy group, a reaction that requires an intermediate in the synthesis of chalcone to dissociate from the synthase active site.

The structure and mechanism of chalcone synthase have been extensively studied by Noel and coworkers.[32,40–42] In the first step of the reaction, the 4-hydroxycinnamoyl group is transferred to the active site cysteine (C164) aided by H303. In contrast to the KAS enzymes, a proline replaces the residue that normally contributes a backbone NH to the oxyanion hole, and consequently the typical oxyanion hole that also includes the backbone NH of the catalytic cysteine is not present in PKS III enzymes. Instead this role is played by the side chain residues of the conserved H303 and N336. Since these residues also participate in decarboxylation of malonyl-CoA, the acyl-enzyme thioester must move out of the oxyanion hole to permit the condensation reactions to occur. The initial steps in the chalcone synthase reaction are shown in **Figure 7** based on Jez *et al.*[42]

The active site of chalcone synthase is shown in **Figure 8(a)**. Two structures have been superimposed, one with malonyl-CoA bound to the C164A mutant and the other with naringenin (NAR) as the ligand. NAR is a product analog, formed from chalcone through a cyclization reaction catalyzed by chalcone isomerase. Thus, NAR has structural features derived from both 4-hydroxycinnamoyl-CoA as well as from the cyclization of the three malonyl groups. Surprisingly, there are few hydrogen-bonding interactions between the ligand and the enzyme, and the authors have proposed that the surface topology of the pocket controls the stereochemistry and regiochemistry of chalcone biosynthesis. The incoming malonyl group that is being delivered to the active site through the 16 Å long CoA-binding tunnel is also shown. Like all PKS III enzymes, the active site of chalcone synthase contains a conserved proline (P375) in addition to C164, H303, and N336. P375 abolishes the oxyanion hole present in other members of the thiolase family that is formed by two backbone side-chain NH groups, one of which is always derived from the active site cysteine. The active site also contains two phenylalanine residues (F215 and F265), one of which (F215) is also found in other thiolase active sites. Noel and coworkers have proposed that F215 and F265 act as gatekeeper residues that limit the access of solvent to the active site during the reaction.[32,43]

The structural and kinetic data on chalcone synthase and related PKS IIIs have been used to develop a model that explains how these enzymes control loading, condensation, and cyclization reactions in one active site. The attention of readers is drawn to an excellent review by Austin and Noel in which the structure and mechanism of plant and bacterial PKS III homologs have been compared and contrasted.[32] Several noteworthy points from their analyses include a discussion on the bacterial PKS III enzymes that contain multiple cysteine residues in their active sites, such as tetrahydroxynaphthalene synthase and DpgA, a PKS III involved in dihydroxyphenylglycine synthesis, and the possibility that some bacterial PKS III enzymes may use ACP as the substrate carrier.

The KAS III–PKS III subfamily of thiolases also includes HMG-CoA synthase, an enzyme that catalyzes the formation of HMG-CoA from acetyl-CoA and acetoacetyl-CoA (AcAc-CoA). The synthesis of HMG is the first committed step in the mevalonate pathway and HMG-CoA synthase is thus a target for the development of novel antibacterials as well as cholesterol-lowering drugs.[44,45] HMG-CoA synthase formally catalyzes an aldol condensation in which an acetyl group attached to the catalytic cysteine attacks the C3 (keto) group of AcAc-CoA. Formation of the acetyl-Cys carbanion is assisted by a glutamate acting as a base whereas the tetrahedral intermediate that results is protonated by a histidine to generate HMG-CoA (**Figure 9**). This
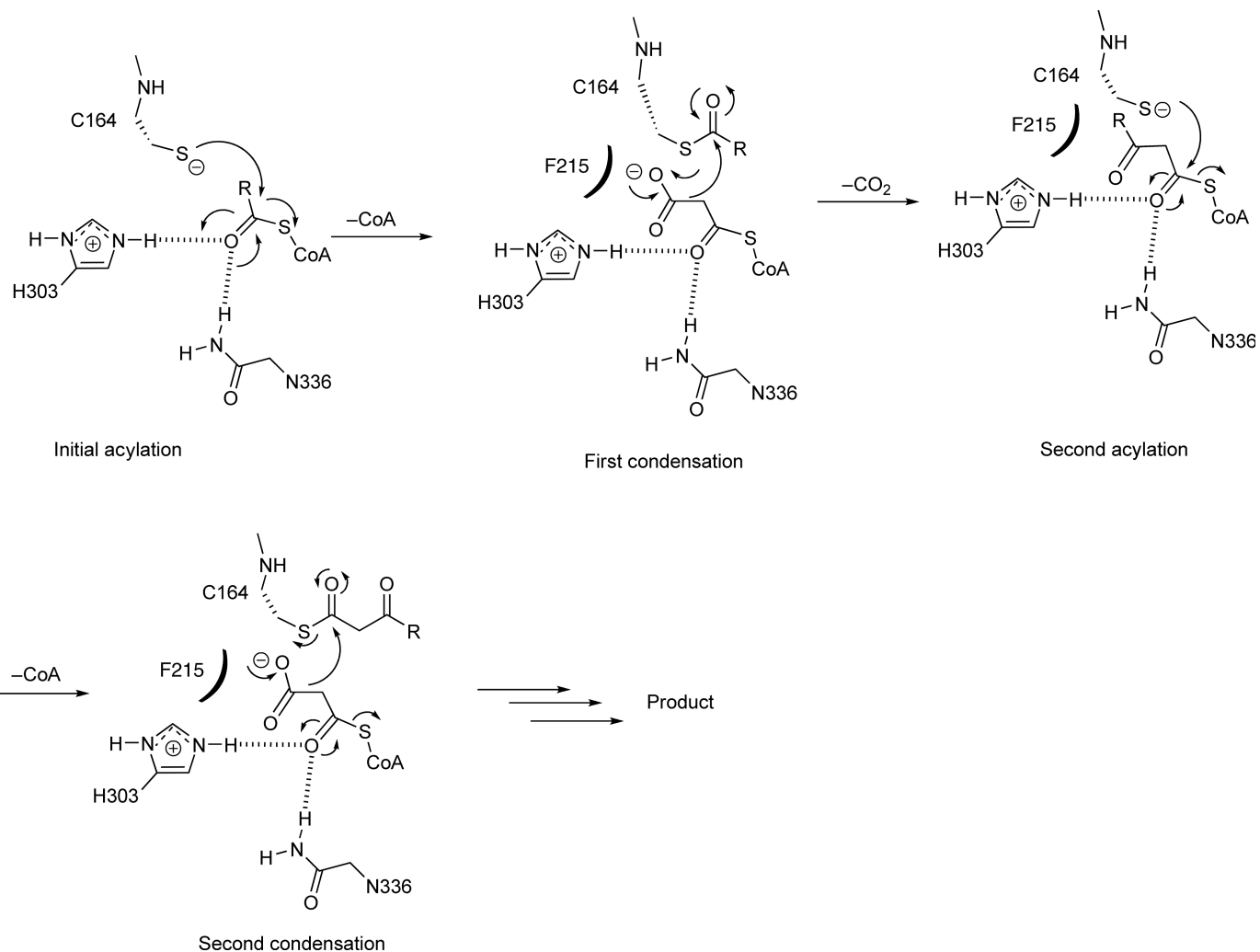
Figure 7 Chalcone synthase mechanism. This shows a portion of the reaction mechanism in which His303 and Asn336 form the oxyanion hole for the acylation and condensation reactions. Steps after the second condensation are not shown but involve a third condensation and cyclization. Note that the decarboxylation and C–C bond formation are shown as concerted. By analogy to the KAS enzymes, this step could also be stepwise with decarboxylation preceding carbanion formation.
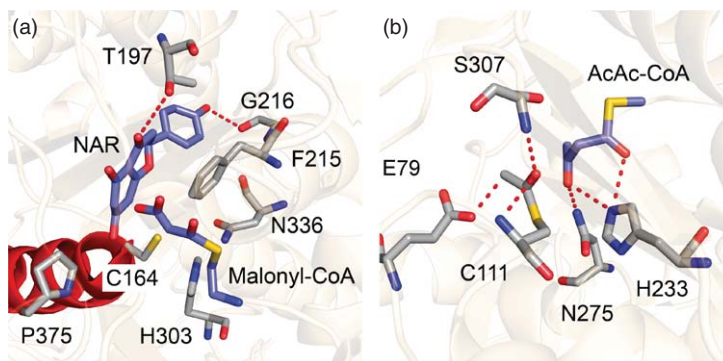
Figure 8 Chalcone and HMG-CoA synthase. (a) Chalcone synthase from *Alfalfa* complexed with naringenin (NAR, slate) (1cgk.pdb) superimposed on the structure of the C164A mutant complexed with malonyl-CoA (slate) (1cml.pdb).[43] Residues shown in gray are from the NAR–enzyme complex. (b) HMG-CoA synthase from *S. aureus* complexed with acetoacetyl-CoA (AcAc-CoA). The active site cysteine in this structure is acetylated (1xpk.pdb[44]).

reaction differs from other thiolase family members in that the carbanion is located on the acyl-enzyme rather than on the noncovalently bound second substrate.

The active site of the *Staphylococcus aureus* HMG-CoA synthase is shown in **Figure 8(b)**. This structure is one of a series determined by Harrison and coworkers using cryocrystallography in which they were able to visualize different intermediates on the reaction pathway.[44] One of the interesting aspects of this work was the discovery that the reaction is reversible and indeed the snapshots they have obtained are of the reaction running backward with HMG-CoA, first acylating the active site cysteine and then subsequently being cleaved to yield AcAc-CoA and an acetyl-enzyme intermediate. In **Figure 8(b)** it can be seen that E79 is perfectly positioned to abstract a proton from the acetyl-cysteine methyl group, the carbonyl of which is hydrogen bonded in the oxyanion hole. In addition, AcAc-CoA is hydrogen bonded to H233 and N275 and is positioned for attack by the acetyl methyl group.

8.07.2.6   Biosynthetic and Degradative Thiolases

Although peroxisomal degradative thiolases are dimers,[46] biosynthetic and degradative thiolases are primarily tetrameric enzymes as shown in **Figure 10**.[47] These enzymes exclusively utilize CoA-based substrates, which are accommodated in a well-defined binding site, and access the active site through a tunnel (**Figure 10**). Although degradative thiolases have a broad substrate specificity for various chain-length β-ketoacyl-CoAs, biosynthetic thiolases exclusively catalyze the synthesis of AcAc-CoA from two molecules of acetyl-CoA. The active site contains four conserved catalytic residues that include two cysteine residues, a histidine and an asparagine (**Figure 10**). The first cysteine, C89, in the biosynthetic thiolase from *Zoogloea ramigera*, is the residue that is acylated during the ping-pong bi–bi kinetic reaction, the neighboring H348 acting as a general base in the initial transfer of the acetyl group from acetyl-CoA to C89 with the transition state stabilization provided by the oxyanion hole formed by the backbone NH groups of C89 and G380 (**Figures 10 and 11**). Subsequently, a second molecule of acetyl-CoA binds and the second cysteine (C378) acts as a base to abstract a proton from the acetyl methyl group, thus generating the required carbanion for the Claisen condensation reaction. The negative charge that develops during proton abstraction by C378 is stabilized by a second oxyanion hole formed by the side chain of H348 and a WAT that is hydrogen bonded to the conserved asparagine (N316).

In the degradative thiolases, the reaction is the reverse of that described for the biosynthetic enzymes (**Figure 8**), and indeed both enzyme classes freely catalyze reversible reactions. The enzymes differ in their catalytic efficiency, and the k_{cat} value for biosynthetic thiolase is ~100-fold larger ($810\,s^{-1}$) than for the degradative thiolase ($8\,s^{-1}$).[47–49] A series of kinetic studies including isotope exchange, kinetic isotope effects, and site-directed mutagenesis have resulted in a detailed understanding of both the forward and reverse reactions catalyzed by thiolases. These studies have included the use of thiol-specific inactivators and labeled ^{14}C-acetyl-CoA to demonstrate that C89 is the active site nucleophile in the biosynthetic thiolase from
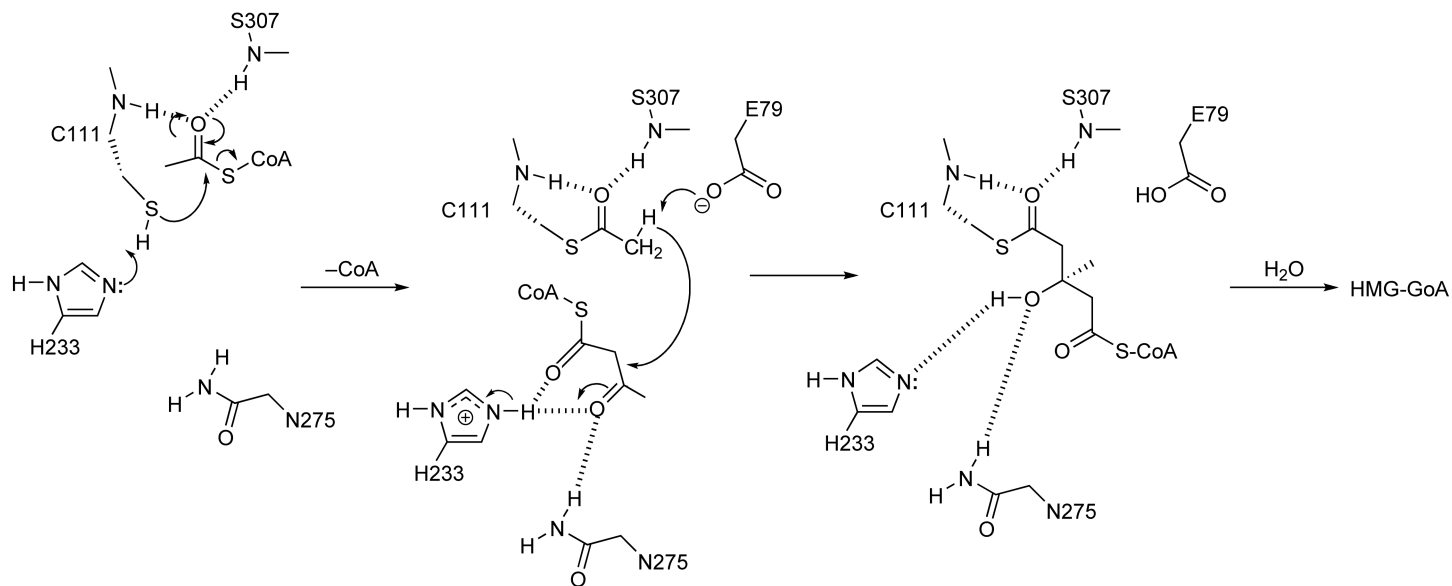
Figure 9 HMG-CoA synthase mechanism. H233 is shown acting as a general base in the initial acylation step that then protonates the THI formed following the attack of the acetyl group on AcAc-CoA. The oxyanion hole for acylation and deacylation of C111 is formed by the backbone NH of S307. Analysis of structural data indicates that the C111 NH is 3.1 Å from the acetyl carbonyl oxygen, suggesting that this group could also function in transition state stabilization.
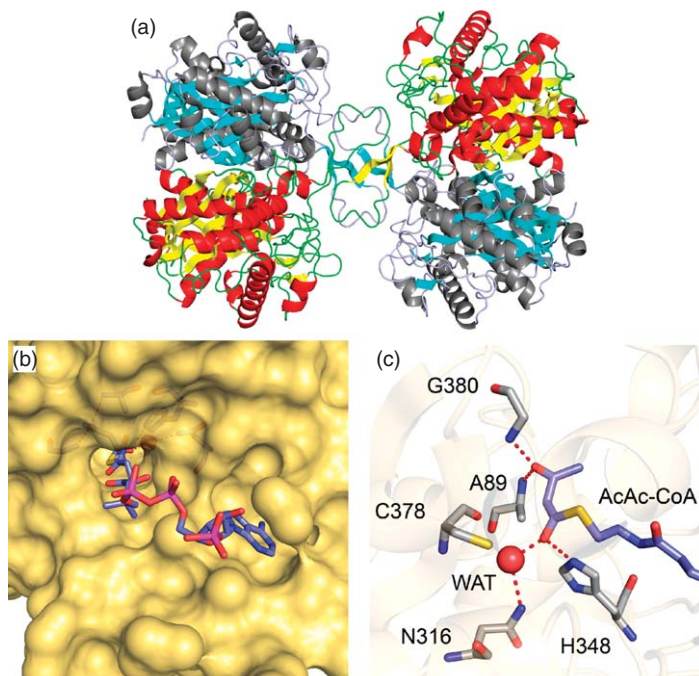
Figure 10 The biosynthetic thiolase from *Z. ramigera*. (a) Thiolase structure showing the tetramerization motif. (b) Surface representation showing the CoA-binding site and the tunnel that provides access for the (acyl)-pantetheine to the active site. (c) The AcAc-CoA group surrounded by active site residues. Note that in this structure the catalytic cysteine (C89) has been mutated to an alanine.

Z. ramigera[48] and that C378 is the base following the observation that the replacement of this residue to a less acidic serine resulted in an enzyme with 0.1% of wild-type activity.[50–52] Finally, it has been shown that the rate-limiting step for the biosynthetic thiolases is the breakdown of the acyl-enzyme, whereas acylation is rate limiting for the degradative thiolases.[47–50,53]

8.07.3   Oxidoreductases in Fatty Acid Biosynthesis and Breakdown

Both the biosynthetic and degradative fatty acid cycles contain two oxidoreductases each. In the biosynthetic pathway, the β-ketoacyl-ACP formed by the KAS enzymes is reduced by an NADPH-dependent reductase, encoded by the *fabG* gene in *E. coli*. Following a dehydration step, the resulting enoyl-ACP is reduced by an enoyl-ACP reductase, encoded by the *fabI* gene in *E. coli*. FabI is an NADH-dependent reductase, and both FabI and FabG are members of the SDR superfamily. Not all bacteria utilize FabI as their enoyl-ACP reductase, and currently, three other enzymes that include FabV, FabL, and FabK are known. Both FabV and FabL are also members of the SDR family; however, the flavin-dependent enoyl-ACP reductase FabK is not.

In the degradative pathway there are two oxidoreductases: a flavin-dependent ACD, which catalyzes the initial oxidation step in the pathway, and an NADH-dependent HAD that catalyzes the penultimate step in the pathway. The mechanisms of these two enzyme classes will be briefly mentioned; however, this section will primarily focus on enzymes that are members of the SDR family.

8.07.3.1   The Short-Chain Dehydrogenase Reductase Family

Most NAD(P)-dependent oxidoreductases, dehydratases, and epimerases are members of either the aldo-keto reductase (AKR) superfamily or the SDR superfamily. Proteins in the AKR family share a common $(\alpha/\beta)_8$-barrel fold and are involved in the metabolism of many important compounds including aldehydes, sugars, and
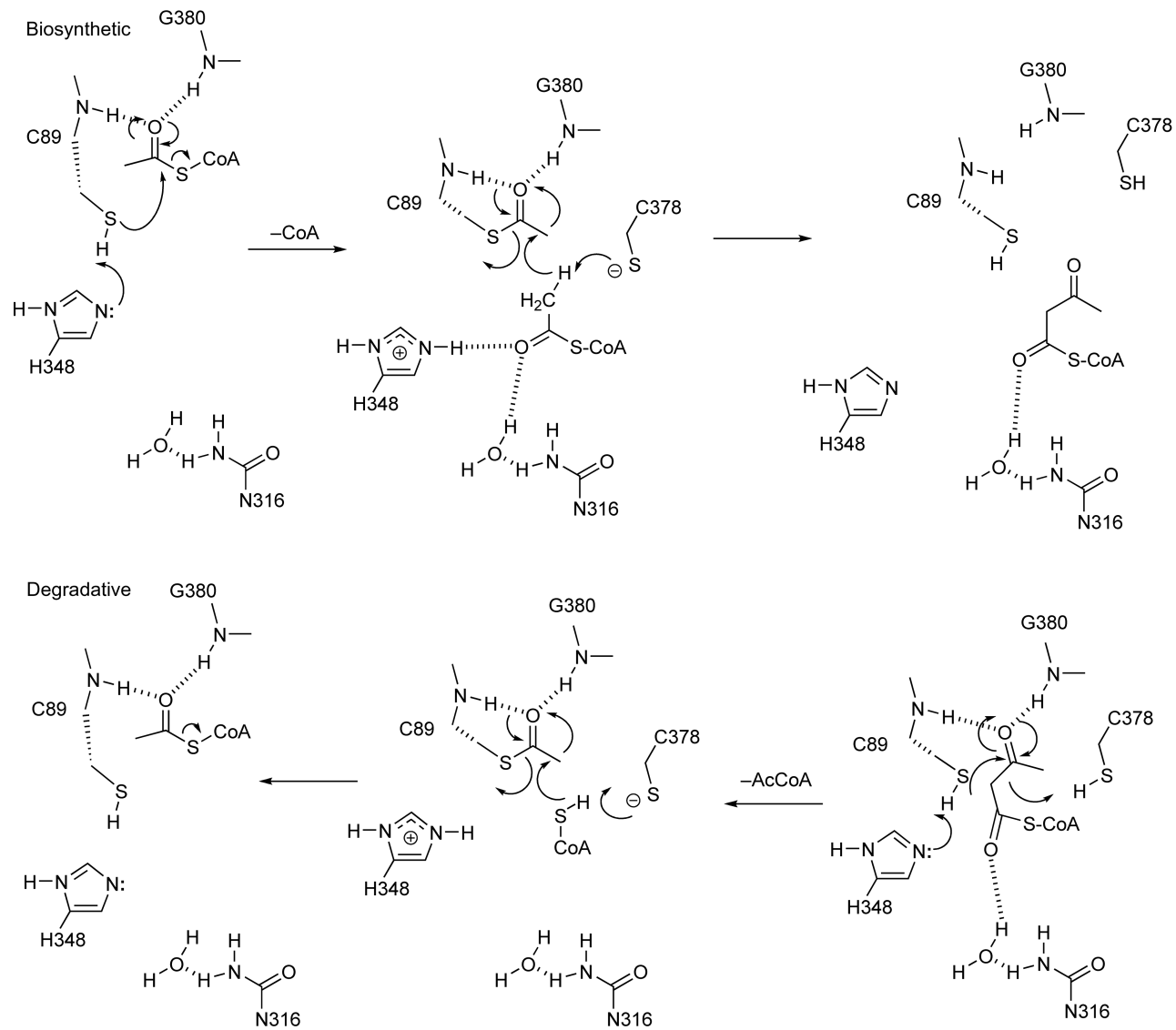
**Figure 11** The reaction mechanism for biosynthetic and degradative thiolases.

steroids. The AKR superfamily has been extensively studied by Penning and coworkers,[54–58] and the reader's attention is drawn to several reviews. SDR enzymes also operate on a wide range of substrates, including steroids and sugars, but utilize a Rossmann fold consisting of a central β-sheet flanked by α-helices to bind the cofactor.[59–62]

The classification of enzymes into the SDR family started with efforts to distinguish short-chain alcohol dehydrogenases from their zinc-dependent medium- and long-chain counterparts.[63] SDR enzymes are single domain proteins of $\sim$250 residues with typically a GXXXGXG nucleotide binding motif in the N-terminus and form either homodimers or homotetramers. All SDR enzymes contain a conserved tyrosine residue in their active site, which is thought to play a principal role in catalysis, similar to that played by the metal ion in the zinc-dependent alcohol dehydrogenases. In many cases, a conserved lysine is also present in the active site, with the tyrosine and lysine organized in an YXXXK or YXXMXXXK motif. The two enzyme subfamilies discussed below differ in the identity of the third conserved residue in the active site triad. The KARs are part of an SDR subfamily in which the third residue is a serine. 7α-Hydroxysteroid dehydrogenase (7α-HSDH) and UDP-galactose-4-epimerase are examples of proteins in this subfamily.[64–66] In contrast, the FabI enoyl-ACP reductases have either a tyrosine or a phenylalanine as the third catalytic residue.

8.07.3.1.1 β-Ketoacyl-ACP reductases: YKS SDR enzymes

The structure of 7α-HSDH from *E. coli* is shown in **Figure 12**. Notable features of this structure include the characteristic Rossmann fold of the SDR enzymes and the change in the position of the substrate-binding loop (residues 195–210) that occurs upon formation of the ternary complex. Variability in the sequence of the substrate-binding loop is likely a primary factor in the substrate specificity of SDR enzymes. A close-up of the active site highlighting interactions between the three active site residues, Y159, K163, and S146, and the substrate and NAD$^+$ cofactor is also shown.[64]

The role of the three conserved residues have been studied in 7α-HSDH as well as several other family members including UDP-galactose-4-epimerase (Y149, K153, S124). There is a general consensus that the conserved tyrosine plays a critical role in catalysis.[65,67] UDP-galactose-4-epimerase replacement of Y149 with a phenylalanine reduces k_{cat} 10 000-fold, whereas replacement of S124 also has a dramatic ($\sim$3000-fold) affect on activity. These data resulted in the proposed mechanism shown in **Figure 13** in which both Y149 and S124 function as acid–base catalysts in protonating/deprotonating the substrate oxygen. As can be seen in **Figure 12**, K163 in 7α-HSDH is hydrogen bonded to the 2$'$ and 3$'$-ribose hydroxyl groups of the cofactor, and there is a general consensus that the conserved lysine plays a primary role in cofactor binding while also modulating the pK_a of the active site tyrosine. In UDP-galactose-4-epimerase the pK_a of Y149 is 6.07, about 4 units lower than that of a tyrosine in solution, which is an effect attributed in part to the neighboring serine as well as to the positive electrostatic field created by K153 and the nicotinamide ring of NAD$^+$.[65,67]
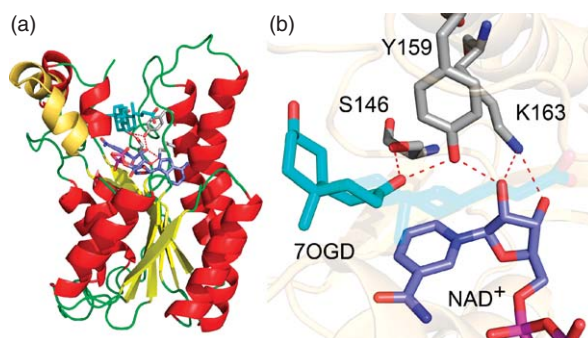
Figure 12   Structure of 7α-HSDH from *E. coli*. (a) Overall structure of 7α-HSDH in complex with NAD$^+$ and the product 7-oxoglycochenodeoxycholic acid (7OGD). The substrate-binding loop in the binary enzyme NAD$^+$ structure is shown in light orange. (b) Interaction of the three conserved active site residues and the two ligands. The pdb files used were 1ahh and 1ahi, and the figure was made with PyMOL.[37]
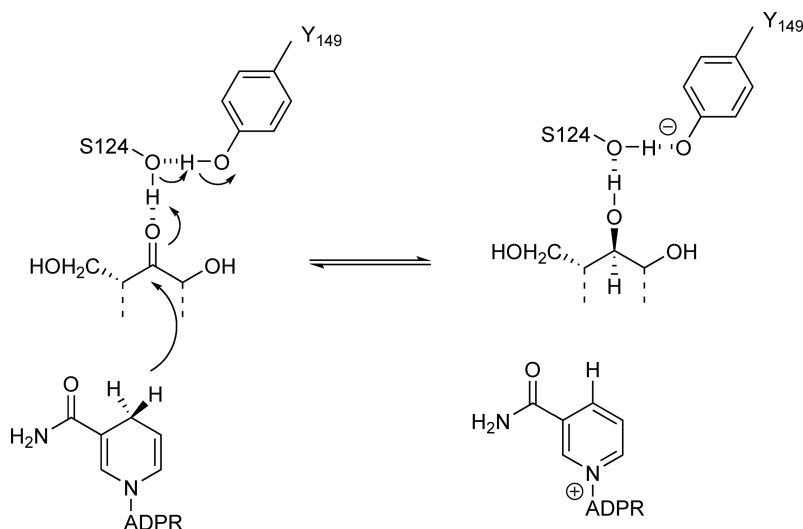
Figure 13 Mechanism of UDP-galactose-4-epimerase.

Structural and mechanistic studies have also focused on the FAS-II KARs, from several organisms including *E. coli*, *Brassica napus*, *M. tuberculosis*, and *S. pneumoniae*.[68–73] Currently, there are no ternary complex structures of these homotetrameric enzymes; however, X-ray studies have identified a conformational change caused by cofactor binding that results in a reorganization of the catalytic residues in the *E. coli* FabG (ecFabG), leading to the creation of a hydrogen-bonding network, which is thought to be important for protonation of the substrate and to account for the negative cooperativity observed for this enzyme.[74] These enzymes catalyze the NADPH-dependent reduction of β-ketoacyl-ACPs to the corresponding 3(*R*)-hydroxyacyl-ACP product in which the *pro*-4*S* hydrogen is transferred from the cofactor. Kinetic studies on the *Plasmodium falciparum* FabG support an ordered mechanism[73] as suggested for ecFabG[70] in which NADPH binds first to the enzyme.[74] Unlike enzymes such as UDP-galactose-4-epimerase in which the catalytic tyrosine has a pK_a value around neutral, pH dependence studies suggest that this residue must be protonated for activity in the FabG enzymes.[75]

A detailed mechanistic study utilizing kinetic isotope effect measurements has been described on the enzymes from *S. pneumoniae* (spFabG) and *M. tuberculosis* (MabA) with AcAc-CoA as the substrate.[68,75,76] In contrast to the reports on other FabGs, both sets of studies support a random rather than an ordered kinetic mechanism for spFabG and MabA. Studies on spFabG are consistent with a random bi–bi rapid-equilibrium kinetic mechanism, in which either substrate can bind first to the free enzyme, and a stepwise mechanism that proceeds through an enolate intermediate (**Figure 14**).

Interestingly, in contrast to spFabG, kinetic isotope effect measurements on MabA point to a concerted mechanism in which both hydride transfer and protonation occur in the same transition state (**Figure 15**).[75,76] The dependence of $^D(V/K)$ for one substrate on the concentration of the second substrate for MabA supports a steady-state kinetic mechanism in which either substrate can bind to the free enzyme and where chemistry is only partially rate limiting. This contrasts with the mechanism of spFabG in which a rapid-equilibrium mechanism has been proposed and where chemistry is almost completely rate limiting.[68] Finally, although MabA is a tetramer at high concentration and when crystallized,[71] at low concentration the protein dissociates into a dimer with a dissociation constant of 22 µmol l^{-1}.[77] Basso and coworkers have shown that NADPH binds to MabA with positive cooperativity, in contrast to the negative cooperativity observed for ecFabG,[70,74] and have used pre-steady-state kinetics to conclude that two conformations of MabA are present in solution with different affinities for NADPH.[76]

8.07.3.1.2 Inhibition of the β-ketoacyl-ACP reductases

Most microorganisms possess only a single KAR, and several authors have pointed out these enzymes are therefore potentially good targets for drug discovery. It is consequently surprising that potent inhibitors of this enzyme class have not been reported so far. Tasdemir *et al.*[78] have identified several flavonoids that inhibit the
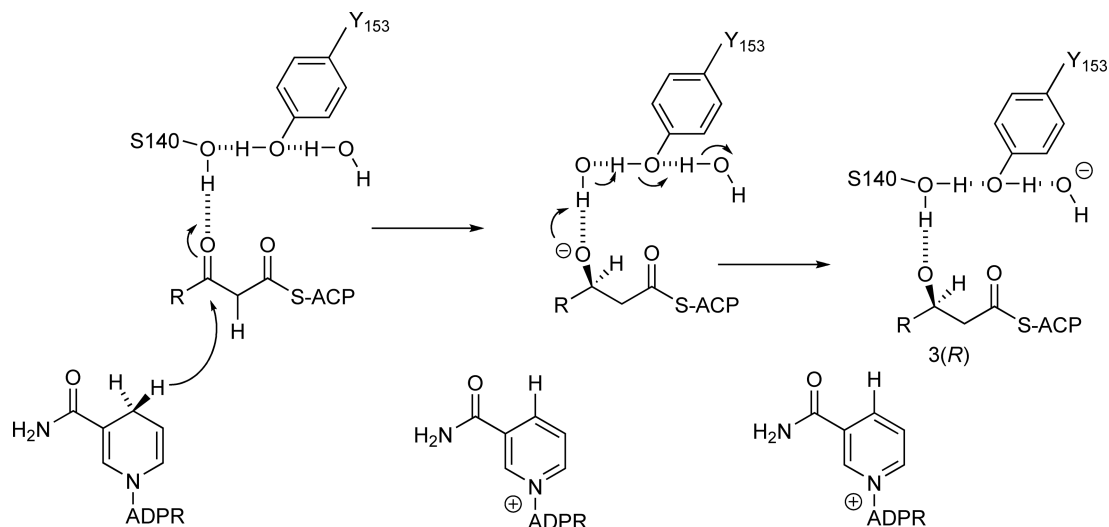
Figure 14 Stepwise mechanism for *S. pneumoniae FabG.* No structure exists for substrate or product fatty acid bound to FabG and so it is not known precisely how the catalytic serine and tyrosine interact with the ligand. In this structure, Ser140 is shown hydrogen bonded to the β-keto carbonyl, which is in the keto tautomer. Thrall, Meek, and coworkers have suggested that hydride transfer occurs to the enol tautomer of the substrate.[69]
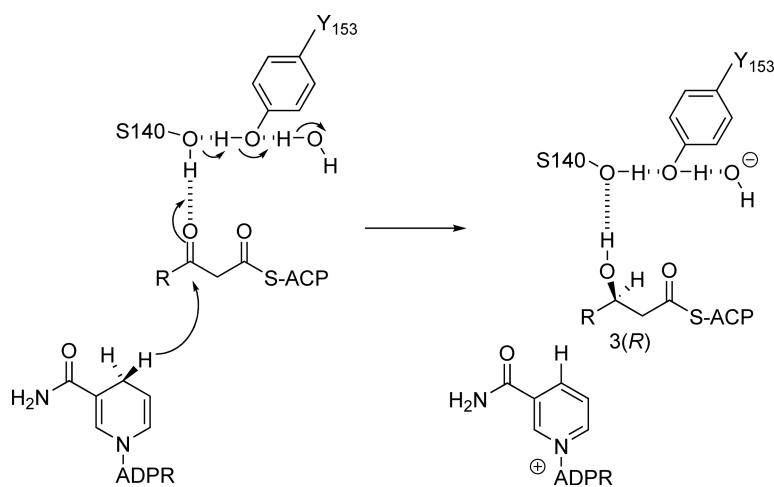
Figure 15 Concerted mechanism for MabA.

P. falciparum FabG with low micromolar IC_{50} values, whereas Quemard and coworkers[79] have reported that one or more products generated from the reaction of activated isoniazid with $NADP^+$ inhibit MabA also with low micromolar IC_{50} values.

8.07.3.1.3   Enoyl-ACP reductases in the SDR family

Enoyl-ACP reductases catalyze the terminal reaction in the fatty acid elongation cycle (**Figure 1**). Currently, four known families of enoyl-ACP reductases have been identified in microorganisms, encoded by the *fabI*, *fabL*, *fabK*, and *fabV* genes.[80–84] FabK is a flavin-containing triosephosphate isomerase (TIM)-barrel enoyl-ACP reductase,[85] first identified in *S. pneumoniae*,[83] whereas FabI, FabV, and FabL are all members of the SDR family. FabI is the most intensively studied enoyl-ACP reductase, first identified in *E. coli* and *Salmonella typhimurium* as the product of the *envM* gene.[80,86,87] In these initial studies, it was demonstrated that treatment with the

antimicrobial diazaborine compound resulted in the same phenotype as a temperature-sensitive *envM* mutation, and subsequently resistance to diazaborine was shown to be due to mutations in the *fabI* (*envM*) gene. Since then many studies have validated FabI as an essential component of the FAS-II pathway and a sensitive target for drug discovery.[20,21,88–90]

8.07.3.1.4   The FabI enoyl-ACP reductase

FabI catalyzes the NADH- or NADPH-dependent reduction of α,β-unsaturated enoyl-ACPs in which the *pro*-4*S* hydrogen of the cofactor is transferred as a hydride to the C3 carbon of the substrate.[91] Both hydride transfer and protonation occur on the same face of the double bond (*syn: si* face at C3 and *re* face at C2), yielding a product in which the 2*R* and 3*S* hydrogens are added during the reaction.[92] Mechanistic studies on the FabI enzyme from *M. tuberculosis* (InhA) are consistent with a stepwise mechanism, in which hydride transfer generates an enolate intermediate that is subsequently protonated to generate the product.[91]

The structures of the FabI enzymes have been determined from several different microbial sources including *B. napus*, *E. coli*, *M. tuberculosis*, *P. falciparum*, and *Helicobacter pylori*.[69,93–97] The conserved tyrosine and lysine are present in the YXXMXXXK or YXXXMXXXK motif, whereas the catalytic serine found in the SDR dehydrogenases (see Section 8.07.3.1) is replaced with a tyrosine in most enoyl-ACP reductases but a phenylalanine in mycobacterial FabIs. Conformational changes identified in structural studies reveal ligand-induced changes in position/ordering of the substrate-binding loop and in the orientation of the conserved tyrosine. **Figure 16** compares the active sites of InhA in binary and ternary complexes where it can be seen that binding of a substrate to the active site (C16-*N*-acetylcysteamine) results in a rotation of 60° about the Cα–Cβ bond of Y158.[94,98] In the absence of a substrate, Y158 is pointing away from the active site and the conformation adopted by Y158 in the ternary InhA complex is thought to be that required for catalysis. Interestingly, the orientation of Y158 in the ternary C16:NAD$^+$ complex is similar to that seen in other binary and ternary FabI structures, where the alternative orientation of the conserved tyrosine is not observed.[97,99,100] The rotation of Y158 in InhA has been proposed as a potential explanation for the observed inverse solvent isotope effect on the reaction.[101]

8.07.3.1.5   FabI mechanism and role of active site residues

The mechanism of InhA has been explored in detail using kinetic isotope effects and site-directed mutagenesis. As noted above, use of deuterated cofactor indicates that InhA is a B-side-specific reductase, whereas the isotope effects point to a ternary complex mechanism that is not strictly ordered but in which there is a preference for cofactor binding first to the enzyme.[91] Hydride transfer results in the formation of an enolate intermediate that is subsequently protonated to yield a product (**Figure 17**). Based on structural studies and by analogy to the mechanisms of other SDR enzymes, it has generally been assumed that Y158 plays a central role in catalysis, either by supplying the proton that results in the collapse of the enolate intermediate or in
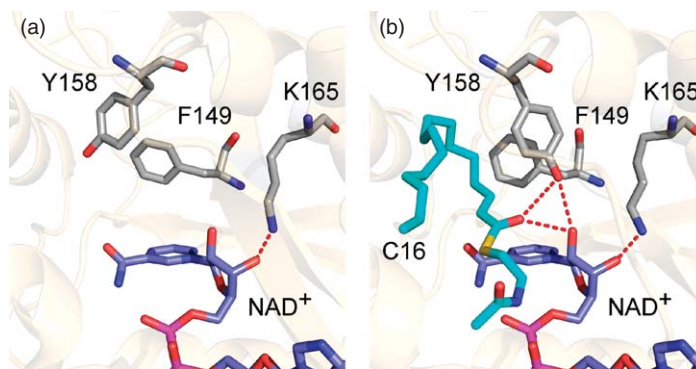
Figure 16   Binary and ternary complexes of InhA. This figure shows the repositioning of the conserved tyrosine in InhA (Y158) that occurs when both cofactor and substrate are bound to the enzyme. (a) Structure of InhA complexed with NAD$^+$. (b) Structure of InhA complexed with NAD$^+$ and hexadecenoyl-*N*-acetylysteamine (C16). The conserved lysine and phenylalanine residues (K165 and F149) are also shown. The figure was made with PyMOL[37] using coordinates from the pdb files 1eny and 1bvr.[94,98]
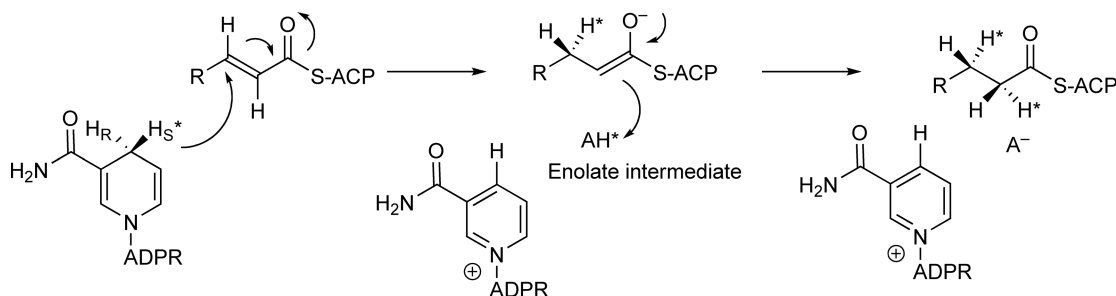
Figure 17 Stepwise mechanism for InhA.

providing electrophilic stabilization of the enolate intermediate. However, while site-directed mutagenesis clearly points to a role for K165 in cofactor binding, data for Y158 are less conclusive. The observation that the Y158S InhA mutant has wild-type activity indicates that Y158 does not function formally as a proton donor in the reaction. In addition, the fact that the Y158F InhA mutant has a k_{cat} value that is only 24-fold lower than wild-type InhA further suggests that Y158 does not play a major role in stabilizing the accumulation of negative charge on the substrate carbonyl in the transition states for the reaction.[101] In this regard, Anderson and coworkers[92,102] have suggested that additional transition state stabilization may not be required in the reactions catalyzed by InhA, and dienoyl-CoA reductase, and have pointed out that the positively charged NAD^+ can also stabilize the transition state. In addition, in contrast to ECH (see Section 8.07.4.2.3) and medium-chain ACD (MCAD) (see Section 8.07.3.2.1), there is no evidence that the active site of InhA (or dienoyl-CoA reductase) polarizes the carbonyl group of substrate analogs or products.[102,103]

One question that still remains concerns the donor responsible for protonating the enolate intermediate. In the structure of the C16 substrate bound to InhA in the presence of NAD^+, the thioester carbonyl is in the *s-cis* conformation so that the *si* face of the C3 carbon is oriented toward the cofactor. In this conformation, the *re* face of the C2 carbon is also oriented toward the cofactor, making it difficult to see how the enzyme could stereospecifically protonate the substrate without a significant structural reorganization. However, the 2′-hydroxyl of the cofactor ribose is on the correct face of the substrate[98] and could in principle be the source of the proton.[102]

There has also been discussion concerning the role of the third conserved residue in the enoyl-ACP reductase active site: a tyrosine in most FabIs but a phenylalanine in InhA. Sacchettini and coworkers[98] have noted that the position of the F149 side chain controls access to the active site from a hydrogen-bonded network of ordered WATs that leads away from the substrate-binding site. F149 is close to the nicotinamide ring of NAD[98] and Raman spectroscopy has been used to show that this residue is important for binding NADH in the correct conformation for B-side hydride transfer.[103] The C–D stretching frequencies (v_{C-D}) for *R*- and *S*-NADD bound to wild-type and F149A InhA were measured using Raman spectroscopy (**Figure 18**). For wild-type InhA, v_{C-D} decreases $11\,cm^{-1}$ upon binding 4(*S*)-NADD, whereas v_{C-D} increases $34\,cm^{-1}$ for 4(*R*)-NADD, leading to the conclusion that the nicotinamide ring adopts a bound conformation in which the 4(*S*)C–D bond is in a pseudoaxial orientation (**Figure 19**). In contrast to the wild-type enzyme, Raman spectra of NADD bound to F149A InhA resemble those of NADD in solution, leading to the conclusion that the F149A mutant is no longer able to optimally position the cofactor for hydride transfer. Kinetic studies demonstrate that this enzyme-induced modulation in cofactor structure is directly linked to catalysis, since replacement of F149 with an alanine causes a 30-fold decrease in k_{cat} and 2-fold increase in $^D(V/K_{NADH})$, and that the catalysis of substrate reduction by InhA results, in part, from correct orientation of the cofactor in the ground state.

### 8.07.3.1.6 FabI cofactor specificity

With regard to cofactor specificity, it is interesting that saFabI is NADPH-dependent, whereas the FabI homologs from *E. coli* (ecFabI), *Bacillus subtilis*, *Haemophilus influenzae*, and *M. tuberculosis* are all NADH-dependent enoyl-ACP reductases.[80,84,104,105] According to the ecFabI–NAD^+ crystal structure, Q40 is very close ($2.8\,\text{Å}$) to the 2′-hydroxyl group of the NAD^+ adenosine moiety (**Figure 20**). Although the crystal
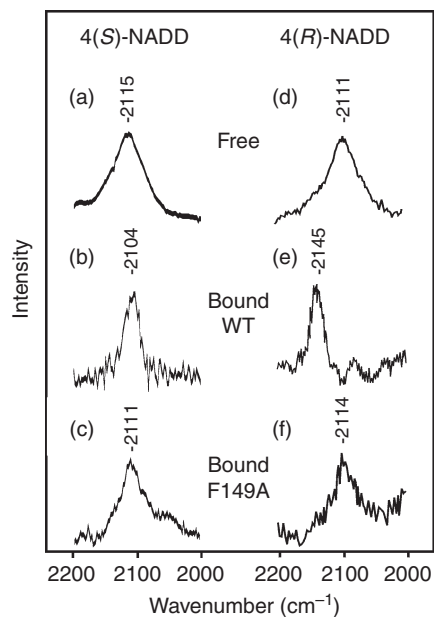
Figure 18   Raman difference spectra in the C–D stretching region. (a) 4(S)-NADD in aqueous solution, (b) in a binary complex with wild-type InhA, and (c) in a binary complex of with F149A-InhA. (d) 4(R)-NADD in aqueous solution, (e) in a binary complex with wild-type InhA, and (f) in a binary complex of with F149A-InhA. Reproduced with permission from A. F. Bell; C. F. Stratton; X. Zhang; P. Novichenok; A. A. Jaye; P. A. Nair; S. Parikh; R. Rawat; P. J. Tonge, *J. Am. Chem. Soc.* **2007**, *129*, 6425.

Model compound III
(C7 = O *anti* to N1)
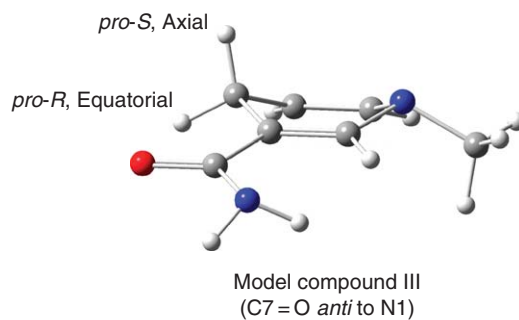
Figure 19   A nicotinamide analog in which the *pro-S* hydrogen is axial. Reproduced with permission from A. F. Bell; C. F. Stratton; X. Zhang; P. Novichenok; A. A. Jaye; P. A. Nair; S. Parikh; R. Rawat; P. J. Tonge, *J. Am. Chem. Soc.* **2007**, *129*, 6425.
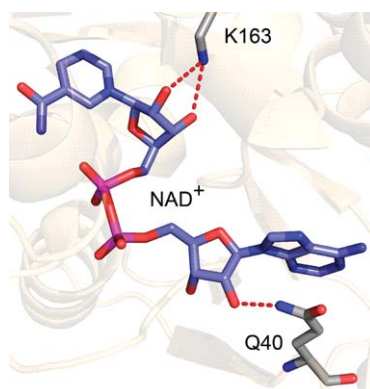
Figure 20   NADH bound to ecFabI. The 2′-adenosine hydroxyl is hydrogen bonded to Q40 in ecFabI. The figure was made with PyMOL[37] using coordinates from the pdb file 1qsg.[99]

structure for saFabI is currently unavailable, sequence alignment of the FabI proteins from different organisms indicates that two positively charged residues, R40 and K41 from saFabI, appear close to the position of Q40 in ecFabI. To investigate the importance of R40 and K41 in the interaction of NADPH with saFabI, two single mutants, R40Q and K41N, were constructed and showed at least a 50-fold decrease in k_{cat}/K_m for NADPH, whereas k_{cat}/K_m for NADH increased by 5–7-fold. Introduction of both mutations into saFabI resulted in an additional 10-fold decrease in the value of k_{cat}/K_m for NADPH, proving that both R40 and K41 are involved in interactions with the 2′-phosphate of NADPH. However, the fact that cofactor specificity was not completely reversed in the saFabI double mutant indicates that other factors are critical for optimal binding of NADH.[105] FabL from *B. subtilis* is also NADPH-dependent. Interestingly, BsuFabL has a R37 in the corresponding position to K41 in saFabI, which could be a factor for determining the specificity of FabL for NADPH compared to NADH.

8.07.3.1.7   *ACP binding and the role of active site residues in other FabIs*

Most structural and mechanistic studies on FabI enzymes, as well as other components of the FAS-II pathway, have exploited enoyl thioesters of *N*-acetylcysteamine or CoA because of the ease of synthesis of these substrates. However, the natural substrate for these enzymes is a thioester of ACP, a small acidic protein that carries the growing fatty acid. For InhA, Blanchard and coworkers have shown that the enzyme will reduce both CoA and ACP-based substrates with similar k_{cat} values but with a greatly increased K_m value of the CoA substrate compared to the corresponding enoyl-ACP.[91] For example, the K_m value for 2-octenoyl-CoA is ~100-fold larger than that for 2-octenoyl-ACP. As expected for an enzyme that normally operates on C_{18+} fatty acids, substrates with increasing chain lengths have lower K_m values for reduction by InhA. In addition, although most FabIs will reduce crotonyl-CoA (C4)[80,106] albeit with large K_m values, InhA is unable to reduce this substrate even up to a substrate concentration of 8 mmol l^{-1}.[80,106] The FabI enzyme from *S. aureus* also has a preference for longer chain fatty acids (C12 compared to C4), and for this enzyme increasing substrate chain length results in an increase in k_{cat}, rather than a decrease in K_m, suggesting that remote interactions between the longer acyl chain and the enzyme modulate the precise orientation of the catalytic groups in the active site in *S. aureus*.[105] Finally, there have been several reports that substrate inhibition is observed at high substrate concentrations when enoyl-CoAs are used. For InhA, no substrate inhibition is observed when the corresponding ACP substrate is used (unpublished data) supporting the proposal that the CoA substrate binds nonspecifically to the enzyme possibly in the NADH-binding site.[91,101]

Because of the central role of ACP in fatty acid biosynthesis, as well as in systems such as PKSs, there has been a long-standing interest in understanding the molecular basis for ACP recognition by target enzymes. Rock and coworkers examined the interaction of ACP with the FAS-II enzymes FabG and FabH using a combination of site-directed mutagenesis and molecular modeling, and concluded that the acidic ACP helix α2 interacts with a basic patch of residues close to the substrate-binding loop.[81,107,108] Tonge and coworkers[109] subsequently determined the structure of ACP bound to ecFabI using a combination of X-ray crystallography and molecular dynamics simulations. Although the X-ray structure clearly showed the location of two ACP molecules on the ecFabI tetramer, the absence of electron density for portions of the structure prevented a detailed analysis of the complex particularly since the ACP phosphopantetheine could not be observed in the structure. Subsequently, computational methods were used to build and refine the productive ACP–FabI complex (**Figure 21**), which was validated by site-directed mutagenesis.

This structure clearly showed the involvement of ecFabI residues K201, R204, and K205 at the FabI–ACP interface (**Figure 22**), as predicted from the previous studies on FabG and FabH. Interestingly, the acyl-phosphopantetheine was shown to enter the active site through the minor portal, contrary to the expectations based on the structure of the C16 substrate bound to InhA.[98] In addition, in the ACP–ecFabI structure the substrate carbonyl is hydrogen bonded to Y146 and not Y156 as might have been expected from previous structural studies on both substrate and inhibitors bound to InhA and other FabIs (**Figure 22**). In support of the ACP–ecFabI structure, site-directed mutagenesis indicated a more important role for Y146 in catalysis compared to Y156. Although the Y156F ecFabI mutant had wild-type activity, the k_{cat}/K_m value for reduction of dodecenoyl-ACP was reduced 50-fold in the Y146F mutant compared to the wild-type enzyme. Given that Y146 is replaced by a phenylalanine in InhA, and given that InhA does not have a basic patch of residues in the same position as that found in other FabIs, the current belief is that substrate recognition differs fundamentally
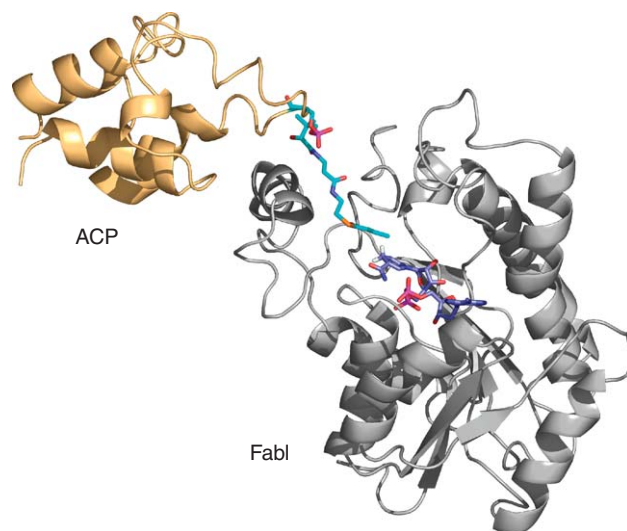
Figure 21   The structure of ACP bound to ecFabI. The phosphopantetheine is colored cyan and enters the active site of FabI through the minor portal. The figure was made with PyMOL.[37]
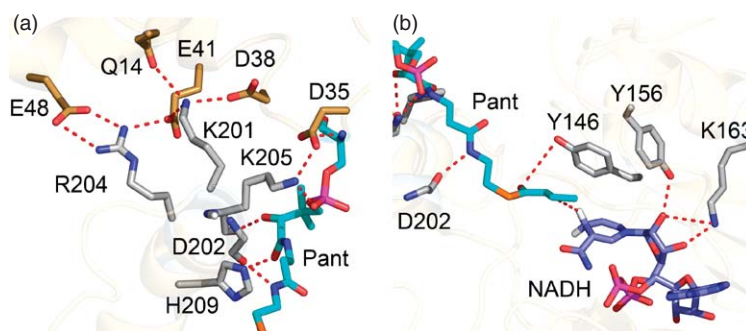
Figure 22   The structure of ACP bound to ecFabI. (a) Interactions between ACP (gold) and ecFabI (gray) at the interface between the two proteins. The phosphopantetheine is colored cyan. Residues that were mutated in ecFabI include K201, R204, and K205. (b) Interactions in the active site of the complex. Note that the substrate is bound in the *s-trans* conformation to present the correct face of the double bond to the cofactor. The figure was made with PyMOL.[37]
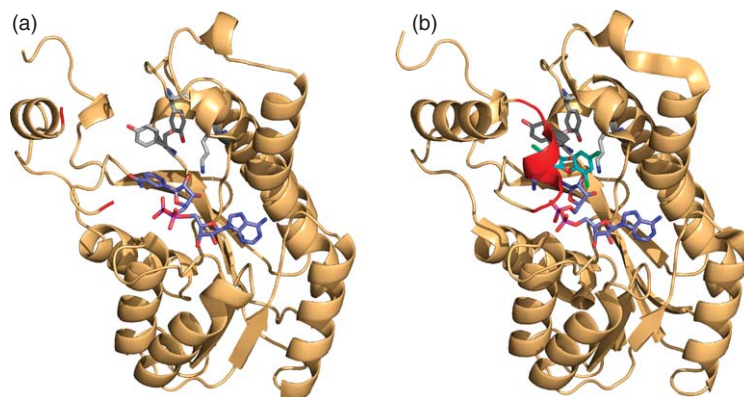
between InhA and other FabIs. In this regard, it is noteworthy that the ACP protein in *M. tuberculosis* (AcpM) is significantly larger than other ACPs.

8.07.3.1.8   Inhibition of FabI

As a validated drug target, several drug discovery programs have focused on developing inhibitors of the FabI enzyme in different organisms including *M. tuberculosis*, *S. aureus*, and *P. falciparum*. FabI inhibitors may be grouped into those that covalently modify the cofactor and those that bind noncovalently to the enzyme in the presence of the cofactor. Examples of compounds that form covalent adducts with NAD^+ include the diazaborines and the frontline tuberculosis drug isoniazid, whereas the noncovalent inhibitors include diphenyl ether-based compounds such as triclosan and 5-octyl-2-phenoxyphenol (8PP), as well as FabI inhibitors developed from screening programs (**Figure 23**).[21]

The diazaborines were the first class of FabI inhibitors identified. Structural studies with the FabI enzymes from *E. coli* and *B. napus* reveal that these compounds form a covalent adduct with the NAD^+ 2′-ribose hydroxyl group and that complex formation leads to ordering of the substrate-binding loop, a loop of amino acids that covers the active site in the ternary complex (**Figure 24**).[93,113] The size of this loop is thought to correlate with

Figure 23 FabI inhibitors. The diazaborines and isoniazid covalently modify the NAD(P) cofactor. Activation of isoniazid requires the mycobacterial enzyme KatG. Diphenyl ether-based inhibitors and two compounds developed from screening programs are also shown: a benzofuran analog of the naphthopyridone series developed by GSK[110,111] and a 2-pyridone developed by Crystal Genomics.[112]

Figure 24 FabI inhibitors that covalently modify the NAD cofactor. (a) Thienodiazaborine bound to ecFabI. (b) The INH–NAD adduct bound to InhA. The figures were made using PyMOL[37] and the pdb files 1dfh and 1zid.[93,114]

the chain length specificity of the enzyme, and in InhA the substrate-binding loop is considerably longer than in other FabIs.[98] Isoniazid also forms a covalent adduct with the NAD^+ cofactor, and this isoniazid (INH)–NAD adduct is a potent inhibitor of InhA (**Figures 23 and 24**)[91,94,114–118] (see Chapter 8.20). The X-ray structure of INH–NAD bound to InhA reveals that F149 has flipped up from its position in the binary enzyme complex so that the aromatic ring forms a stacking interaction with the acyl-pyridine group of the inhibitor.[114]

The diphenyl ether triclosan, an antimicrobial additive in many personal care products, is a potent inhibitor of the FabI enzymes from *E. coli*, *S. aureus*, and *P. falciparum* (see Lu and Tonge,[21] and references therein). Similar to the diazaborine inhibitors, binding of triclosan to ecFabI in the presence of NAD^+ results in ordering of the substrate-binding loop (**Figure 25**).[99] Kinetic studies have shown that triclosan is a slow-onset inhibitor of

Figure 25   Loop ordering caused by triclosan binding to ecFabI. (a) Structure of the ecFabI monomer in complex with NAD^+ (slate). Active site residues are shown in gray. The ends of the disordered substrate-binding loop are shown in red. (b) Structure of the ecFabI:NAD^+:triclosan ternary complex showing the ordered substrate-binding loop. The figures were made using PyMOL[37] and the pdb files 1eny and 1qsg.[94,99]

ecFabI ($K_i^* = 7$ pmol l^{-1})[119–121] and it has been proposed that the slow step in formation of the final enzyme–inhibitor complex (EI*) involves ordering of the substrate-binding loop.[23] In support of this hypothesis, triclosan is only a rapid reversible inhibitor of InhA ($K_i = 0.2$ µmol l^{-1})[122] and the InhA substrate-binding loop is disordered in the InhA:NAD^+:triclosan ternary complex.[23] In contrast, the INH–NAD adduct is a slow-onset inhibitor of InhA ($K_i^* = 0.8$ nmol l^{-1})[118] and structural studies reveal that the loop is ordered in the complex of this inhibitor bound to InhA.[114]

Slow-onset inhibitors are expected to have long residence times on their targets thus improving their *in vivo* antimicrobial activity.[21,123,124] Based on this knowledge, a series of diphenyl ether-based inhibitors have been designed with the objective of causing loop ordering upon binding to InhA. Using structure-based design, a series of alkyl-diphenyl ether InhA inhibitors were synthesized, the most potent of which (8PP) inhibits InhA with a K_i value of 1 nmol l^{-1} (**Figure 23**).[21,23] These compounds have minimum inhibitory concentration (MIC) values of 1–2 µg ml^{-1} against both drug-sensitive and isoniazid-resistant strains of *M. tuberculosis*.[23] Isoniazid resistance arises primarily from mutations in KatG, the mycobacterial catalase-peroxidase that activates isoniazid, and the data are thus consistent with the hypothesis that InhA inhibitors that do not require KatG activation should be active against drug-resistant strains of *M. tuberculosis*. Efforts are underway to improve the potency of the compounds, probe their *in vivo* mechanism of action, and explore chemical space about the diphenyl ether pharmacophore.[125–127] Although the alkyl diphenyl ethers were designed to promote loop ordering in InhA, structural studies indicate that the substrate-binding loop remains disordered when 8PP is bound to InhA (**Figure 26**) and consistent with the observation that 8PP is not a slow-onset enzyme
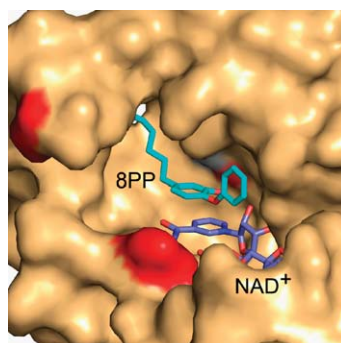
Figure 26   Structure of 5-octyl-2-phenoxyphenol (8PP) bound to InhA. 8PP binds to InhA without causing ordering of the substrate-binding loop the ends of which are colored red. The figure was made using PyMOL[37] and the pdb file 2b37.[23]

inhibitor.[23] A selection of these compounds also have excellent antibacterial activity against other Gram-positive and Gram-negative pathogens, such as *S. aureus*, *Francisella tularensis*, and *Burkholderia pseudomallei*, raising the possibility that they could be used to develop novel broad-spectrum chemotherapeutics.[105] Recent studies on the FabI enzyme from *F. tularensis* support the importance of inhibitor residence time, rather than the thermodynamic affinity of the inhibitor for the enzyme, in modulating *in vivo* antibacterial activity.[124]

8.07.3.2   Acyl-CoA Dehydrogenase and Hydroxyacyl-CoA Dehydrogenase

In addition to the SDR enzymes in the FAS-II pathway, two other oxidoreductases are present in the fatty acid oxidation pathway: ACD and HAD. ACDs are flavin-dependent enzymes that catalyze the first step in the pathway, whereas HAD uses NAD^+ to oxidize β-hydroxyacyl-CoAs to the corresponding β-ketoacyl-CoAs.

8.07.3.2.1   Acyl-CoA dehydrogenases

The ACDs constitute a family of flavin-containing enzymes with at least nine members that catalyze the α,β-oxidation of fatty acyl-CoA thioesters[128,129] (see Chapter 7.03). Interest in the structure and mechanism of ACDs stems in part from their potential role in diseases such as sudden infant death syndrome.[130] Most ACDs are homotetramers[128] binding one molecule of flavin adenine dinucleotide (FAD) per subunit, and MCAD, from the mitochondrial β-oxidation pathway, is the most heavily studied family member. Several reviews on the structure and mechanism of the ACDs have appeared, and the reader's attention is drawn to Thorpe and Kim,[131] Kim and Miura,[128] and Ghisla and Thorpe.[129]

   In the structure of octanoyl-CoA bound to MCAD, the thioester carbonyl is hydrogen bonded to the backbone NH of E376 as well as the FAD 2′-ribose hydroxyl (**Figure 27**).[132] The substrate is bound so that the *pro*-3*R* hydrogen is oriented toward the FAD N5 atom whereas the side chain of E376 is close to the substrate's *pro*-2*R* hydrogen. This orientation is in agreement with the known stereochemistry for the reaction in which E376 abstracts the *pro*-2*R* hydrogen with concomitant transfer of the *pro*-3*R* hydrogen to the flavin as a hydride (**Figure 28**). Subsequently, the flavin is reoxidized by electron-transferring flavoprotein (ETF). Kinetic isotope effects are consistent with a concerted mechanism in which both proton abstraction and hydride transfer occur in the same transition state,[133–135] although theoretical studies point to a stepwise mechanism that involves the formation of an enolate intermediate.[136] Regardless of whether the reaction is concerted or stepwise, an important feature of the MCAD-catalyzed reaction is the ability of the enzyme to stabilize a carbanionic transition state, as seen through the use of substrate analog reporters of the active site environment (see below) and the ability of the enzyme to catalyze the exchange of the α-proton with solvent.[137,138]

   Mechanistic challenges faced by the enzyme include the mismatch in redox potentials for the free enzyme ($-145\,mV$) and substrate ($-40\,mV$), and the basicity of the substrates' α-protons ($\sim pK_a \sim 20$). Much information has been gained by analyzing the interaction of the enzyme with chromophoric substrate analogs, such as 3-thiaoctanoyl-CoA,[139] *p*-substituted phenylacetyl-CoAs,[140] 3-indolepropionyl-CoA,[141] and 2,4-hexadienoyl (HD)-CoA,[142,143] and FAD analogs, such 5-deaza-FAD[133,144,145] and 2′-deoxyFAD.[137,146] These studies have revealed that interactions between enzyme and substrate, such as those formed with the oxyanion hole, result in
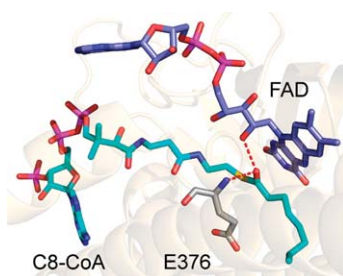

Figure 27   Structure of octanoyl-CoA bound to MCAD. The substrate carbonyl group (C8-CoA, cyan) is hydrogen bonded to the backbone NH of E376 (gray) and the 2-hydroxyl of the FAD ribose (slate). The figure was made using PyMOL[37] and the pdb file 3mde.[132]
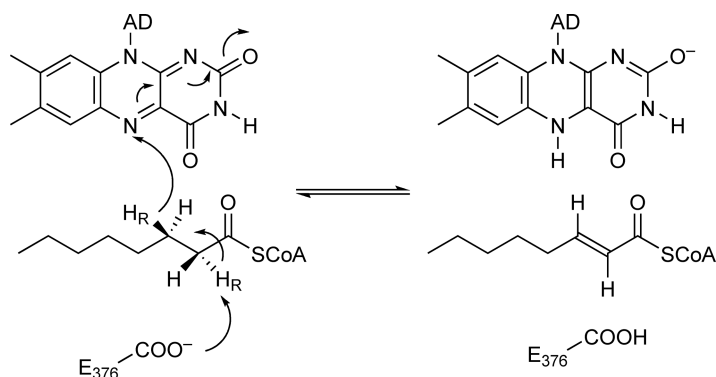
Figure 28   Substrate oxidation catalyzed by MCAD.

a large change in redox potential,[147,148] so that the reaction is now thermodynamically favorable, and a large increase in α-proton acidity.[140]

The complex formed between the reduced enzyme and the oxidized enoyl-CoA product is characterized by a charge-transfer band at 570 nm.[149] The nature of this species, together with that of other stable enzyme–ligand complexes, such as that formed between the enzyme and the product analog HD-CoA, have been probed by Raman, resonance Raman, and NMR spectroscopies.[137,142,150–152] These spectroscopic experiments have provided direct evidence for the ability of the enzyme to polarize the ground state of the ligand and experiments with 2′-deoxyFAD provide a direct link between ligand polarization and reactivity. Raman spectra of HD-CoA bound to MCAD indicate that the HD enone vibrational band decreases 27 cm^{-1} upon binding to the enzyme (**Figure 29**). In contrast, the 2′-deoxyFAD-MCAD is only able to cause a 10 cm^{-1} decrease in the enone vibrational frequency. The changes in enone frequency result from catalytically relevant
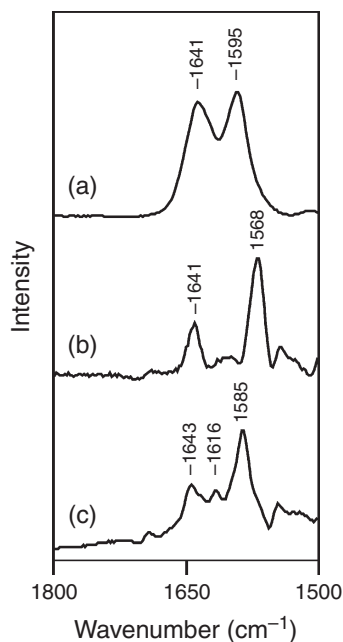
Figure 29   Raman spectra of HD-CoA bound to MCAD. Raman difference spectra of HD-CoA were obtained using 752 nm excitation. (a) Free in solution. (b) Bound to wild-type MCAD. (c) Bound to 2′-deoxyFAD-MCAD. Reproduced with permission from J. Wu; A. F. Bell; L. Luo; A. W. Stephens; M. T. Stankovich; P. J. Tonge, *Biochemistry* **2003**, *42*, 11846.
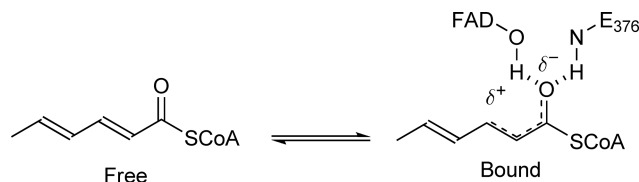
**Figure 30**   Polarization of HD-CoA by MCAD.

stress on the ligand, which is a consequence of placing the ligand in an environment that is complementary to the transition state for the reaction rather than the ground state.[137] The decrease in the ability of the 2′-deoxyFAD-MCAD enzyme to cause ligand polarization indicates that the 2′-hydroxyl may contribute up to $17 \, \text{kJ} \, \text{mol}^{-1}$ in ground state destabilization, a value that is remarkably similar to the $15 \, \text{kJ} \, \text{mol}^{-1}$ increase in activation energy for the reaction caused by the removal of this hydrogen bond. The enzyme-induced changes in the electronic structure of the ligand result in an increase in electron density at C2 and a decrease at C3, thereby promoting α-proton abstraction and hydride transfer (**Figure 30**). The Raman studies thus provide a direct link between ligand polarization and reactivity.[137]

The structure of HD-CoA bound to MCAD has also been studied using ^{13}C NMR and ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) spectroscopy. Interestingly, large 13–14 ppm decreases (upfield changes) in chemical shift are observed for the C1 and C3 carbons of HD-CoA upon binding to MCAD, instead of the downfield shifts expected from the decrease in electron density at C1 and C3 predicted from the Raman data. The apparent discrepancy between the Raman and NMR data was rationalized by proposing that the HD-CoA chemical shifts were influenced by the isoalloxazine ring current resulting from placing the enzyme-bound flavin in a magnetic field.[142]

8.07.3.2.2   *Hydroxyacyl-CoA dehydrogenase*

L-HAD catalyzes the NAD^{+}-dependent oxidation of β-hydroxyacyl-CoAs to the corresponding β-ketoacyl-CoAs. The most heavily studied isoform of HAD is a dimeric enzyme (EC 1.1.1.35) with broad substrate that functions in the mitochondrial fatty acid β-oxidation pathway.[153–156] This enzyme was originally known as short-chain HAD (SCHAD). Peroxisomes contain two MFEs (MFE1 and MFE2) with HAD activity. MFE1 is homologous to the mitochondrial HAD,[157] whereas the HAD component of MFE2 is a member of the SDR superfamily.[158] Like the FAS-II KARs that are also SDR family members, MFE2 operates on D-hydroxyacyl-CoAs (3R) rather than the corresponding L-isomers (3S). Thus, although HAD and the FAS-II KARs catalyze the interconversion of β-hydroxy and β-ketoacyl groups, these two classes of enzyme differ fundamentally in the stereochemistry of the reaction and in their catalytic mechanisms: The KAR enzymes use the S-Y-K triad found in SDR enzymes whereas HAD uses a catalytic histidine-serine-glutamate triad to catalyze the reaction.

The X-ray structure of HAD reveals that the enzyme has two domains: an N-terminal domain that binds the cofactor and that has a characteristic $\beta\alpha\beta$ Rossmann fold, and a C-terminal domain that is involved in subunit dimerization. The two domains are connected by a flexible linker, and the binding of substrate in a cleft between the domains causes a shift in the relative orientation of the two domains. Structures of binary enzyme–substrate and enzyme–cofactor complexes have been determined together with a stable ternary complex in which oxidized cofactor (NAD^{+}) and product (AcAc-CoA) are both bound to the enzyme (**Figure 31**). These studies reveal that the β-keto group of the product is hydrogen bonded to S137 and H158, with a conserved glutamate (E170) hydrogen bonded to H158. The β-keto group is also hydrogen bonded to N208, whereas the thioester carbonyl (C1) is hydrogen bonded to the backbone NH of N161. The AcAc group is positioned so that the *re* face of the C3 carbonyl is oriented toward the cofactor. This is the geometry expected if the 3S-hydroxyacyl group is bound in the active site with the proton oriented toward the NAD^{+}. Site-directed mutagenesis supports the importance of both H158 and E170 in the reaction catalyzed by HAD, and a mechanism has been proposed in which H158 acts as a general acid to protonate the carbonyl oxygen following
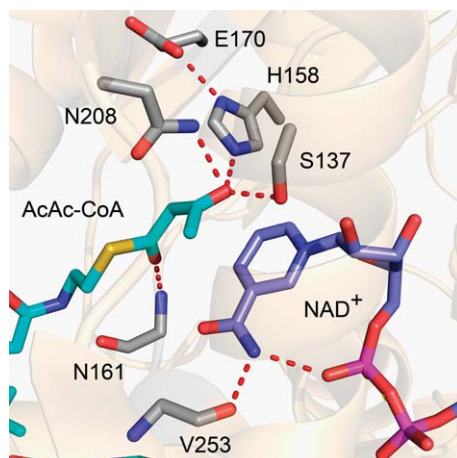
Figure 31   Interactions in the active site of HAD. AcAc-CoA is shown in cyan and NAD$^+$ in slate. Red dashes represent hydrogen bonds. The figure was made using PyMOL[37] and the pdb file 1f0y.[153]
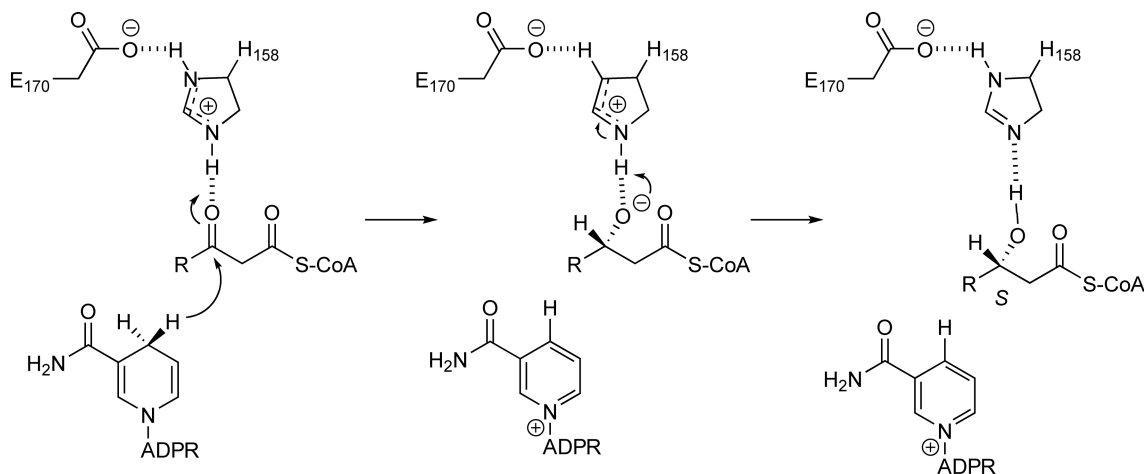
Figure 32   Mechanism of the HAD reaction.

hydride transfer to the C3 carbon (**Figure 32**). The other residues around the C3 carbonyl are thought to be involved in stabilizing the accumulation of negative charge on the oxygen in the transition state.

8.07.4   Hydratases and Dehydratases

As elaborated above, the major difference in the FAS-II and fatty acid oxidation cycles is the stereochemistry of the β-hydroxyacyl group that is generated during each round of elongation or degradation. In the biosynthetic pathway, the KAR enzymes such as FabG generate a 3R-(D)-hydroxyacyl-ACP, whereas in the oxidation pathway the HAD enzyme accepts 3S-(L)-hydroxyacyl-CoA as the substrate. Consequently, the enzymes involved in each pathway that catalyze the dehydration/hydration reactions must also have the opposite stereochemistry. In *E. coli* the dehydration of 3(R)-hydroxyacyl-ACPs to the corresponding 2-*trans*-enoyl-ACPs is catalyzed by the FabZ and FabA enzymes, whereas in the mammalian β-oxidation pathway the hydration of enoyl-CoAs is performed by ECH (crotonase). In this section, we will briefly review the FabZ and FabA enzymes, and will then discuss ECH together with other members of the crotonase superfamily.

8.07.4.1 Dehydration in the FAS-II Pathway: FabZ and FabA

Although both FabZ and FabA catalyze the dehydration of 3(R)-hydroxyacyl-ACPs, FabA also catalyzes a second reaction in which the 2-*trans*-enoyl-ACP product is isomerized to a 3-*cis*-enoyl-ACP.[159–164] The isomerization reaction is an essential step in the biosynthesis of unsaturated fatty acids and FabA, together with the KAS FabB, is found in Gram-negative bacteria that produce unsaturated fatty acids. Although FabA has a preference for fatty acids up to C10 in length, FabZ has a broad substrate specificity and is much more widely distributed than FabA. Structures of FabA from *E. coli* (ecFabA),[165] as well as FabZ from *Pseudomonas aeruginosa* (paFabZ),[166] *P. falciparum*,[167] and *H. pylori* (hpFabZ),[168] have been determined, and there is a continuing interest in understanding the catalytic mechanisms and in developing inhibitors of these enzymes.[78,168–170]

The structures of both FabZ and FabA are characterized by a $\alpha + \beta$ 'hotdog' fold, in which six antiparallel β-sheets with topology 1/2/4/5/6/3 wrap around a central five or six turn α-helix located between $\beta2$ and $\beta3$ (**Figure 33**). FabA is a symmetric dimer, in which each active site is located at the interface between the two dimers, whereas FabZ is hexameric. In FabA the putative active site residues are an aspartate (D84′) and a histidine (H70), each contributed by one subunit, and located about halfway down a substrate-binding tunnel. In the structure of ecFabA inhibited by the mechanism-based inactivator 3-decynoyl-*N*-acetylcysteamine, the thioester carbonyl is hydrogen bonded to the backbone NH of A105.[165] This interaction may play a role in stabilizing enolate intermediate(s) and transition states on the reaction pathway. Alternatively, the substrate could bind so that the carbonyl group is hydrogen bonded to residues at the N-terminus of the central α-helix, raising the possibility that the helix dipole could be involved in stabilizing the accumulation of negative charge on the substrate during the reaction. Rando and Bloch[171] have studied the mechanism of ecFabA, and have shown that dehydration to the 2-enoyl product precedes isomerization to the 3-enoyl species (**Figure 34**). The active sites of the FabZ enzymes are all similar to FabA except that the aspartate in FabA is replaced by a glutamate (E63 in paFabZ). Since the side chains of E63 and D84 occupy very similar positions in FabZ and FabA, it is not immediately clear why FabZ does not also catalyze substrate isomerization. However, Rock and coworkers have suggested that differences in the shape of the substrate tunnel between the two enzymes result in a subtle repositioning of the substrate in FabZ so that the 3-*cis* product cannot be formed.[166]

8.07.4.2 Hydratases in the Fatty Acid β-Oxidation Pathway: Enoyl-CoA Hydratase and the Crotonase Superfamily

Enzymes that catalyze the heterolytic cleavage of a C–H bond α to a carboxylic acid must overcome the mechanistic challenge posed by the weak acidity of the α-proton.[172] In hydration–dehydration reactions involving α,β-unsaturated carboxylic acids, this challenge is commonly met in two ways[173]: either by converting the carboxylic acid to a CoA or ACP thioester or by using metal ions to neutralize the negative charge on the carboxylate. Mitochondrial ECH,[174] β-hydroxydecanoyl thioester dehydratase (FabA),[175] and the mammalian Type I FAS[176] are all enzymes that catalyze the *syn* hydration–dehydration of α,β-unsaturated
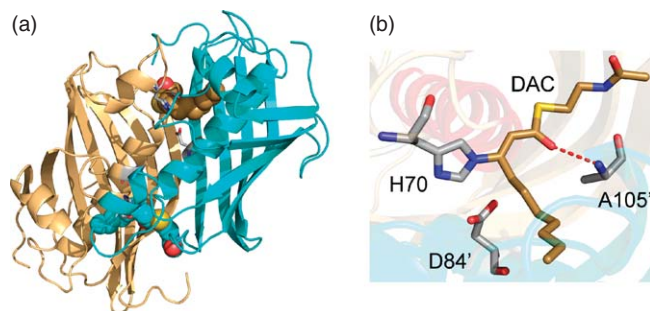
Figure 33 Structure of ecFabA inhibited by 3-decynoyl-NAC. (a) Cartoon of the the ecFabA dimer with the inhibitor (DAC) shown as spheres. (b) One of the active sites in ecFabA showing the two putative catalytic residues. The figure was made using PyMOL[37] and the pdb file 1mka.[165]
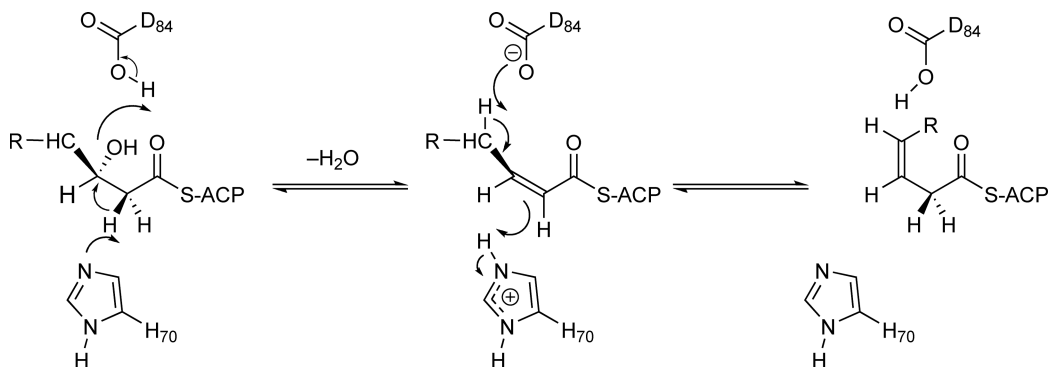
Figure 34 Mechanism of FabA. In this mechanism, H70 acts as a base to abstract the C2 proton whereas D84 protonates the leaving group in the first half of the reaction. Subsequently, D84 acts as a base to abstract a proton from C4 leading to the formation of the 3-*cis*-enoyl-ACP product following reprotonation of C2 by H70. As proposed by Kimber *et al.*,[166] the proximity of C4 to D84 is essential for the isomerization reaction and in FabZ C4 may not be close enough to the carboxylate of E63 for the latter to function as a base. Note that the elimination is *syn* and involves abstraction of the *pro-2S* proton from the substrate.[175] Reproduced with permission from M. Leesong; B. S. Henderson; J. R. Gillig; J. M. Schwab; J. L. Smith, *Structure* **1996**, *4*, 253.

thioesters, whereas aconitase,[177] fumarase,[178] and enolase,[179] are examples of enzymes that bind the free carboxylic acid of the substrate and utilize metal ions to catalyze *anti*-addition–elimination reaction. Mohrig *et al.*[173] have shown that hydration of crotonyl-*N*-acetylcysteamine occurs with a fourfold preference for the *anti* compared to the *syn* pathway, in agreement with the expectation that the *anti* pathway is favored in nonenzymatic reactions, due to the eclipsed geometry of substituents required in *syn* addition–elimination reactions. It is perhaps surprising that enzymes such as ECH stereospecifically catalyze *syn* reactions, and it has been concluded that the stereochemical path adopted by each enzyme class appears to be based on historical contingency rather than deriving from specific mechanistic advantages presented by one pathway over the other.[173,180]

Mammalian cells possess several enzymes capable of hydrating 2-enoyl-CoA thioesters. In the mitochondrial fatty acid β-oxidation cycle, which is the principal focus of this review, the hydration of enoyl-CoAs is catalyzed by ECH, an enzyme that was originally named crotonase. The mitochondrial ECH, which catalyzes the *syn*-hydration of *trans*-2-enoyl-CoAs to the corresponding 3(S)-hydroxyacyl-CoAs, is sometimes abbreviated as ECH1 because of the identification of a second enzyme (ECH2) that catalyzes the formation of 3(R)-hydroxyacyl-CoAs. ECH2 is found in peroxisomes as a component of the R-specific MFE2 that also has dehydrogenase activity. In addition, a second MFE is also found in peroxisomes that is specific for 3(S)-hydroxyacyl-CoAs (MFE1). The hydratase domain of MFE2 (ECH2) is characterized by a hotdog fold and is structurally homologous to the FAS-II FabA and FabZ dehydratases. In contrast, the hydratase domain of MFE1 is a member of the crotonase superfamily (see Kunau *et al.*,[181] Poirier *et al.*,[182] and Bhaumik *et al.*,[183] and references therein). In this section of the review, we will concentrate on the reaction catalyzed by the mitochondrial ECH.

8.07.4.2.1 *Enoyl-CoA hydratase*

Mitochondrial ECH catalyzes the hydration of *trans*-2-crotonoyl-CoA, with k_{cat} and k_{cat}/K_m values of $1790\,s^{-1}$ and $3.6 \times 10^8\,mol^{-1}\,l\,s^{-1}$, respectively, which is close to the diffusion-controlled limit for the encounter of enzyme and substrate.[184] Although the enzyme also efficiently hydrates substrates with longer acyl groups, the catalytic efficiency decreases as the length of the acyl chain is increased.[185]

There has been significant discussion over the mechanism of the reaction and whether the addition of water is concerted or stepwise. The general expectation is that β-elimination reactions such as that catalyzed by ECH will operate by a stepwise mechanism in which abstraction of the α-proton results in the formation of a carbanion that subsequently eliminates the β-leaving group (**Figure 35**). Stabilization of this carbanion will be a critical component of the reaction,[186] and Gerlt, Gassman, and coworkers[187,188] have proposed that
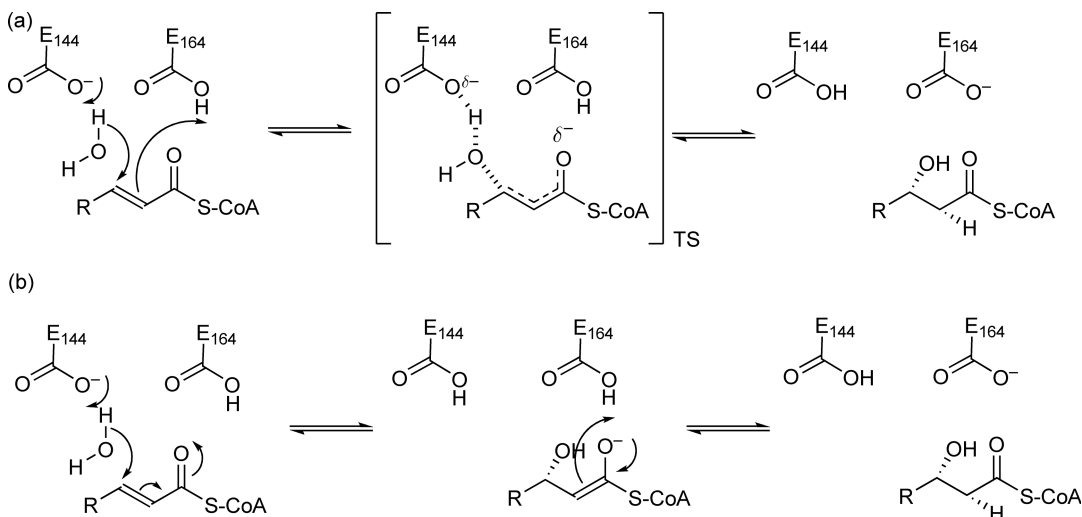
Figure 35   Mechanism of the ECH reaction. (a) Concerted mechanism that proceeds via a carbanionic transition state. (b) E1cb reaction proceeding via the formation of an enolate intermediate.

electrophilic catalysis of such reactions involves protonation of the adjacent carbonyl group to lower the pK_a of the α-proton. Interestingly, based on the use of rigorous kinetic isotope effects, Bahnson and Anderson[189,190] concluded that the reaction is concerted. Bahnson *et al.*[191] have, however, since suggested that the mechanism could be stepwise based on the structure of a substrate analog bound to the enzyme and the presence of a tightly bound WAT in the active site. The proposed four-membered cyclic transition state would also account for the observed *syn* stereochemistry of the reaction. Whether concerted or stepwise, it is clear that the enzyme has evolved to stabilize the accumulation of negative charge on the carbonyl oxygen of the substrate as would occur in either an E1cb mechanism or the concerted addition of water that proceeded through a carbanionic transition state. The ability of the enzyme to catalyze α-proton exchange with solvent is an indication that the active site can stabilize the enolate formed by proton abstraction.[192] The importance of α-proton acidity has also been directly demonstrated by studying the hydration of crotonyl-oxyCoA in which the thioester of the normal substrate crotonyl-CoA has been replaced with an oxyester.[193] This single atom S to O replacement increases the pK_a value of the α-protons $\sim$10 000-fold, and leads to a 300-fold reduction in k_{cat}. Studies with substrate analogs have revealed much concerning the role of active site residues and the ability of the enzyme to cause and stabilize charge rearrangement (see Section 8.07.4.2.3).

8.07.4.2.2   Structure of enoyl-CoA hydratase and the role of conserved residues

The structure of ECH in complex with AcAc-CoA was initially determined by Engel *et al.*[194] Structures of the enzyme with the ligands AcAc-CoA,[195] 4,4′-dimethylamino-cinnamoyl-CoA,[191] and 2,4-HD-CoA are now available.[196] The enzyme is a homohexamer with six active sites, organized as a dimer of trimers. CoA ligands enter the active site through a tunnel that points toward the intertrimer space (**Figure 36**). The carbonyl group of the ligand is hydrogen bonded in an oxyanion hole formed by the backbone NH groups of G141 and A98, and G141 lies at the N-terminus of a short two-turn α-helix, raising the possibility that the dipole from this helix contributes to transition state stabilization. Finally, the active site reveals the presence of two glutamates, E144 and E164, to which a WAT is hydrogen bonded. The WAT is positioned on the *si* face of the C3 carbon and the substrate is bound in the *s-cis* conformation so that the *re* face of C2 is also oriented toward the two glutamates. This conformation is consistent with the stereochemistry of the reaction.

   The role of the active site residues has been probed by site-directed mutagenesis. E164 was initially proposed to be involved in the ECH reaction,[192] based on the observation that the homologous residue in 3,2-enoyl-CoA isomerase E165 was important for catalysis.[197] It was subsequently shown by Anderson and coworkers[192] that the replacement of E164 with a glutamine reduced the catalytic activity by over a 1000-fold and resulted in an enzyme that was unable to catalyze the exchange of the α-proton with solvent, whereas
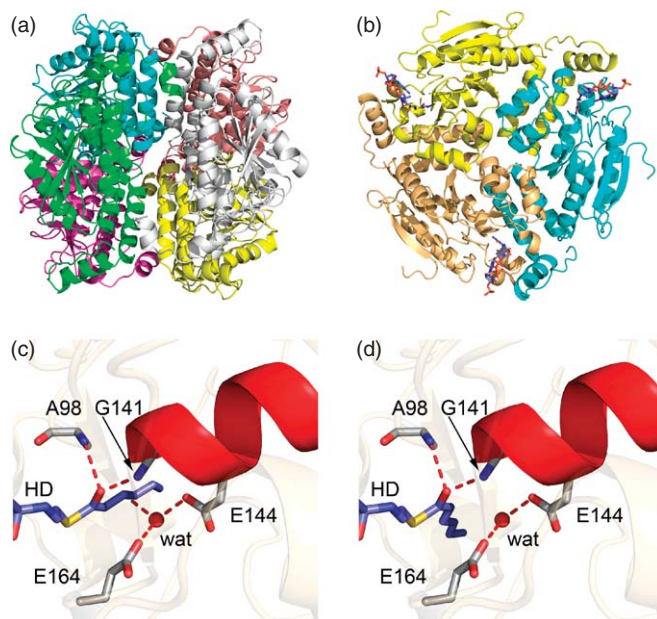
Figure 36 Structure of ECH. (a) View showing the overall hexamer, which is a dimer of trimers. (b) View of one of the trimers showing the ligands pointing toward the trimer interface. (c) Active site showing the catalytic residues and interactions between the HD ligand and the enzyme. The ligand is bound in the *s-cis* conformation. (d) Model of the ligand bound in the *s-trans* conformation obtained by 180° rotation of the C3=C2–C1=O single bond in the HD ligand from c. The figure was made using PyMOL[37] and the pdb file 1mj3.[196]

Muller-Newan *et al.*[198] subsequently demonstrated that the E164Q mutant had a k_{cat} value that was reduced more than 100 000-fold compared to the wild type.

As will be described below, ECH is the prototypical member of the crotonase superfamily. The first structure determined of a crotonase family member was 4-chlorobenzoyl-CoA dehalogenase[199] (see Chapter 8.04). This structure was used to develop a model of the ECH active site from which it was predicted that E144, a second conserved glutamate, might also be located in the active site.[200] Subsequently, it was demonstrated that E144 also played a critical role in substrate hydration, replacement of this residue with a glutamine reducing k_{cat} by 7700-fold, and also eliminating the ability of the enzyme to catalyze α-proton exchange.[184] The identification of both E164 and E144 as catalytic residues, together with subsequent X-ray crystallographic studies, provided the platform for a detailed analysis of the catalytic mechanism from which it was concluded that both residues act in concert to catalyze substrate hydration.[184]

Although ECH stereospecifically catalyzes the exchange of the *pro-2R* α-proton with solvent, NMR experiments with high concentrations of enzyme revealed that the enzyme could also catalyze the exchange of the second α-proton with solvent (*pro-2S*). Subsequently, it was shown that exchange of the *pro-2S* hydrogen occurred in parallel with the formation of the incorrect 3(*R*)-hydroxybutyryl-CoA enantiomer. Although the formation of the 3(*R*) enantiomer occurs through the incorrect (and catalytically inefficient) hydration of *trans*-2-crotonyl-CoA, 3(*R*)-hydroxybutyryl-CoA is efficiently dehydrated by the enzyme to yield *cis*-2-crotonyl-CoA. Thus, the enzyme is capable of catalyzing the epimerization of crotonyl-CoA (**Figure 37**).[201]

The rate of formation of 3(*R*)-hydroxybutyryl-CoA from *trans*-2-crotonyl-CoA is 4×10^5-fold slower than the normal hydration reaction but at least 1.6×10^6-fold faster than the nonenzymatic reaction. Thus, the overall stereospecificity of the reaction (3(*S*)/3(*R*)) is 400 000:1. Interestingly, although the replacement of E164 with aspartate or glutamine reduces the rate of formation of 3(*S*)-hydroxybutyryl-CoA, the rate of formation of the 3(*R*) enantiomer is unaffected.[202] Thus, k_S/k_R is 1000 for E164D and 0.33 for E164Q. In contrast, mutagenesis of E144 affected the rate of formation of both product enantiomers. These observations led to the proposal that *trans*-2-crotonyl-CoA was bound in the active site in two conformations, *s-cis* or *s-trans* about the C(=O)–C2 single bond, and that the formation of 3(*R*)-hydroxybutyryl-CoA results from the addition of water to the
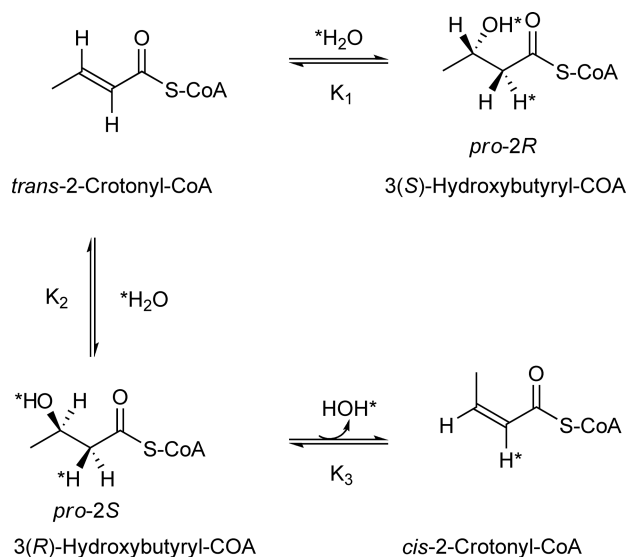
Figure 37   Formation of 3(S)- and 3(R)-hydroxybutyryl-CoA by ECH. Experimentally, the equilibrium constant K_1 is 7.5. Since the two enantiomers have the same energy, K_2 must also be 7.5. *Ab initio* studies predict that *trans*-2-crotonyl-CoA is 12 kJ mol^{-1} more stable than the *cis* isomer, giving an equilibrium constant between the two enantiomers of 0.0079 (K_4). K_3, the equilibrium constant for the dehydration of 3(R)-hydroxybutyryl-CoA to *cis*-2-crotonyl-CoA, can then be calculated from the relationship $K_3 = K_4/K_2 = 0.001$. Reproduced with permission from W. J. Wu; Y. Feng; X. He; H. S. Hofstein; D. P. Raleigh; P. J. Tonge, *J. Am. Chem. Soc.* **2000**, *122*, 3987.

s-trans conformer of *trans*-2-crotonyl-CoA. Only the *s-cis* conformer is appropriately aligned with the catalytic machinery, and thus the stereospecificity of the reaction results from the preferential hydration of one of the two bound substrate conformers, rather than the preferential binding of a single substrate conformer.

8.07.4.2.3   Raman and NMR studies of ligand polarization in enoyl-CoA hydratase

The equilibrium constant for the hydration of *trans*-2-crotonyl-CoA to 3(S)-hydroxybutyryl-CoA is 7.5 in favor of hydration. Thus, at equilibrium only a small fraction of the unsaturated acyl-CoA is present. Anderson and coworkers developed a series of unsaturated active site probes in which an aromatic or alkenyl group was introduced into the substrate in conjugation with the C2=C3 double bond. These ligands, such as cinnamoyl-CoA, 4,4′-dimethylaminocinnamoyl-CoA, and 2,4-HD-CoA, bind with micromolar affinity to the enzyme but are not hydrated. These conjugated substrate analogs have proved extremely valuable in probing electronic effects in the active site of the enzyme using techniques such as absorption, NMR, and Raman spectroscopies. Binding of these probes to ECH results in a red shift in the $\pi-\pi^*$ electronic transition of the conjugated α,β-unsaturated acyl group, indicating that the enzyme has reduced the energy gap between ground and excited states.[203] Although this effect is likely primarily due to the stabilization of the excited state, in which negative charge is expected to be transferred to the carbonyl oxygen, absorption spectroscopy does not inform on whether the enzyme has affected the energies of ground state, excited state, or both. To gain more insight into the enzyme-induced alteration in the electronic structure of these molecules, NMR and Raman spectroscopies were used to specifically probe the ground state of the substrate analogs.[203] The ^{13}C NMR studies revealed that the enzyme causes a reduction in electron density at C3 and C1, and an increase in electron density at C2, seen, for example, by +2.8, +2.8, and −3.0 chemical shift changes at C3, C1, and C2, respectively, when 4,4′-dimethylaminocinnamoyl-CoA binds to the enzyme.[204] Anderson and coworkers estimated that the energy required to cause the observed ground state electronic changes was 13 kJ mol^{-1}.[204]

Similar changes in ground state structure are observed using Raman spectroscopy. **Figure 38** shows the Raman difference spectrum of 2,4-HD-CoA free in solution and bound to ECH. Major changes are observed in the intense bands around 1510–1650 cm^{-1}, which result from an alteration in the vibrational coupling within the HD group caused by selective polarization of the C3=C2–C1(=O) enone fragment, which becomes uncoupled
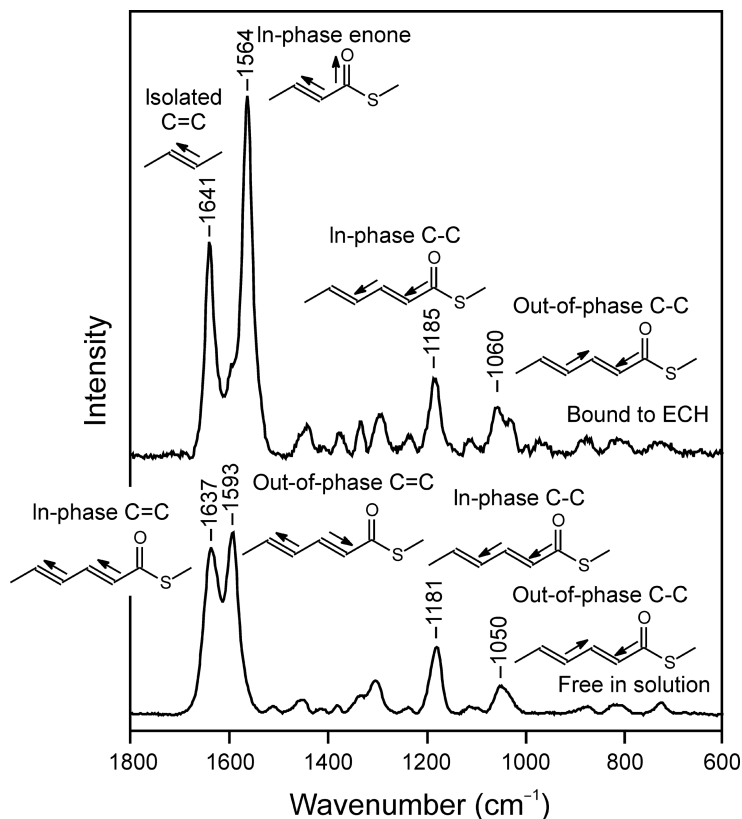
Figure 38   Raman difference spectra of HD-CoA free in solution and bound to ECH.

from the C5=C4 double bond.[205] Ligand polarization results from hydrogen bonding of the enone carbonyl in the oxyanion hole. Although the mutation of E144 or E164 did not affect ground state polarization, replacement of G141 with a proline caused a 10^6-fold reduction in k_{cat} and resulted in a complete loss of ligand polarization.[206] This result provides a direct link between ligand polarization and catalysis: the changes in ligand structure result from placing the substrate (analog) in an environment that has evolved to bind the transition state rather than the ground state for the reaction. This stresses the substrate because it is under pressure to adopt the charge distribution and structure present in the transition state.[196] In response to this stress, the substrate deforms, an effect that is classically referred to as ground state strain.[207] Based on an analysis of the Raman data, it was estimated that up to 40 kJ mol^{-1} of strain energy may be present in the ground state. This strain is a consequence of the stress applied to the substrate along the reaction coordinate and the resulting ground state destabilization may be responsible for up to 30% (10^5-fold) of the increase in k_{cat} caused by the enzyme (10^{14}-fold).[196]

Although mutagenesis of E164 and E144 does not alter the ground state polarization of the ligand, it does affect the intensities of the Raman bands. This allows fours bands to be observed in the Raman spectrum of HD-CoA bound to the enzyme (**Figure 39**). Analysis of these data is consistent with the presence of two conformers of the ligand in the active site, that are *s-cis* and *s-trans* about the C3=C2–C1=O single bond. These two conformers are also present in the spectrum of the ligand bound to the wild-type enzyme, except that the bands arising from the *s-cis* conformer are much more intense. Since the change in intensity is a function of the electronic environment and does not necessarily reflect the relative amounts of the two conformers, the data provide direct evidence for two bound ligand populations in agreement with the stereochemical data presented above. To reiterate, the stereospecificity of the reaction results from the preferential hydration of one of the two bound substrate conformers, rather than the preferential binding of a single substrate conformer. This makes sense given the difficulty in fixing the conformation of acyl group in the active site given that the latter has no useful polar groups that the enzyme could use for specific binding interactions.[196]
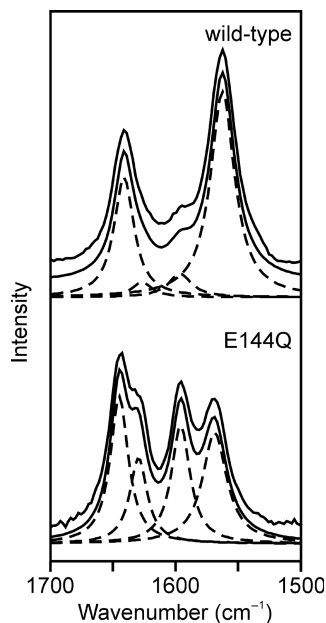
Figure 39   Raman spectra of HD-CoA bound to wild-type and E144Q ECH. Four bands can be observed in the spectrum of the ligand bound to the E144Q enzyme, in the 1550–1650 cm^{-1} region. These four bands can also be discerned in the wild-type spectrum. Reproduced with permission from A. F. Bell; Y. Feng; H. A. Hofstein; S. Parikh; J. Wu; M. J. Rudolph; C. Kisker; A. Whitty; P. J. Tonge, *Chem. Biol.* **2002**, *9*, 1247.
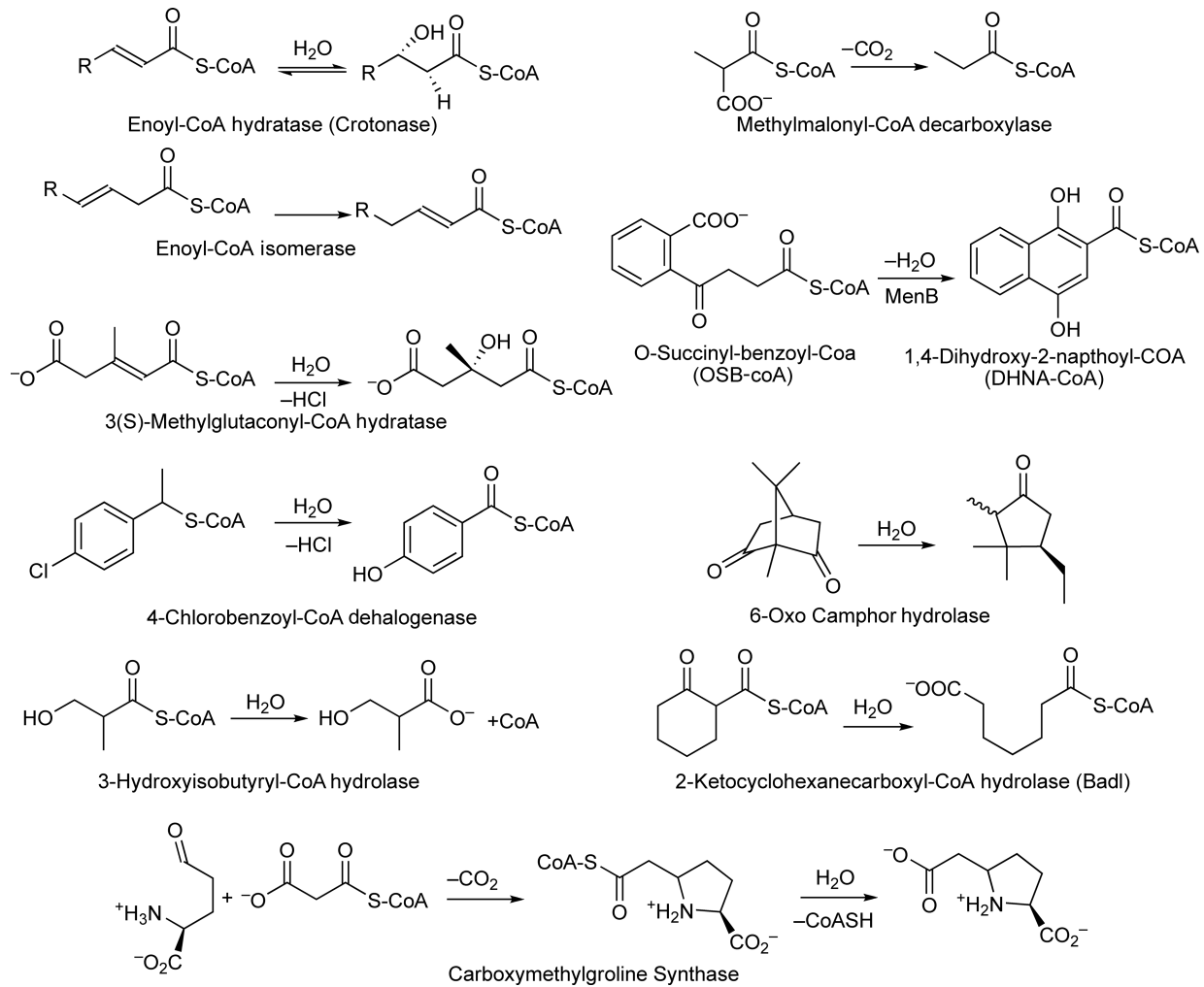
8.07.4.3   The Crotonase Superfamily

ECH or crotonase is the prototypical member of the crotonase superfamily. As noted above, sequence homology between ECH and 3,2-enoyl-CoA isomerase as well as with dihydroxynaphthoate synthase (MenB) and 4-chlorobenzoyl-CoA dehalogenase resulted in the initial proposal for a superfamily based on the crotonase scaffold.[197,198] Since then, many more members of the superfamily have been identified.[208,209] Most family members utilize substrates that are CoA thioesters, and a unifying mechanistic theme throughout the superfamily concerns the use of an oxyanion hole to stabilize carbanionic transition states. **Figure 40** shows the reactions catalyzed by a subset of family members.[197,199,210–217]

 The structure and mechanism of a number of crotonase family members have been studied in detail, revealing similarities with the mechanism of the reaction catalyzed by ECH.

8.07.4.3.1   *4-Chlorobenzoyl-CoA dehalogenase*

4-Chlorobenzoyl-CoA dehalogenase was the first member of the crotonase superfamily to be structurally characterized[199] (see Chapter 8.04). The enzyme catalyzes the dehalogenation of 4-chlorobenzoyl-CoA to 4-hydroxybenzoyl-CoA and is involved in the degradation of 4-chlorobenzoate in *Pseudomonas* sp. strain CBS-3. The reaction proceeds through the formation of a covalent Meisenheimer adduct of an active site aspartate (D145), which subsequently hydrolyzes to give a product with the assistance of H90 acting as a general base.[218] Although the dehalogenase lacks homologs of E144 and E164 in ECH, the structural studies reveal the presence of an oxyanion hole, formed by the backbone NH groups of F64 and G114, which stabilizes the accumulation of negative charge on the benzoyl carbonyl during Meisenheimer complex formation and breakdown.[199] This oxyanion hole and the adjacent α-helix cause ligand polarization in an analogous fashion to that observed for ECH. Both NMR and Raman spectroscopies provide direct evidence for catalytically relevant ground state strain.[219–221] Interestingly, studies with site-directed mutations introduced close to the benzoyl carbonyl permitted the establishment of a linear correlation between the frequency of the carbonyl bond vibration and the rate of Meisenheimer complex formation. This linear correlation is reminiscent of previous studies on serine proteases[11] and strengthens the link between ligand polarization and catalysis.[222]

**Figure 40** Reactions in the crotonase superfamily.

8.07.4.3.2   Dihydroxynaphthoyl-CoA synthase and BadI: Dieckmann and retro-Dieckmann reactions

As we have seen in the section on thiolases, Claisen condensations normally involve activation of the electrophilic carbonyl through formation of a thioester and stabilization of the attacking carbanion also as a thioester. Dihydroxynaphthoyl-CoA synthase (MenB) catalyzes an intramolecular Claisen condensation reaction in which only the nucleophilic portion of the molecule has been converted to a thioester. This reaction is a component of the menaquinone biosynthetic pathway, and most studies have focused on the enzyme from *M. tuberculosis* based on the premise that this pathway may be a valid target for the development of novel compounds that inhibit both replicating and nonreplicating bacteria.[223]

The structure of MenB is shown in **Figure 41**. In this crotonase family member, the C-terminus of one subunit crosses the trimer–trimer interface to form part of the active site of an adjacent monomer. Conserved residues in the active site include G161 and G105, which form the oxyanion hole, as well as D192, S190, and Y287. All residues are contributed by one subunit except for Y287, which is from the C-terminus of the adjacent monomer. This structure is of the MenB from *M. tuberculosis* and shows a second aspartate in the active site, D185. D192 and D185 are structurally homologous to E144 and E164 in ECH. However, even though D185 is intriguingly positioned close to the substrate, this residue is not conserved in all MenB homologs. However, site-directed mutagenesis indicates that D185, together with D192 and Y287, is essential for catalytic activity. The ligand used for the structural studies was AcAc-CoA and thus lacks the naphthoate ring structure of the product. This may be one reason why Y287 is pointing out of the active site. Additionally, a loop of amino acids close to the active site is disordered in the crystal structure (108–133 in all subunits except B: 108–125).[223]

Based on the structural data, a reaction mechanism has been proposed in which the aromatic carboxylate of the substrate (*O*-succinylbenzoate) abstracts the *pro*-2*S* hydrogen to generate a resonance-stabilized carbanion that subsequently attacks the protonated carboxyl group. Following elimination of water, the final product is formed by a thermodynamically favorable keto-enol tautomerism (**Figure 42**). The proposed mechanism is consistent with the stereochemical course of the reaction elucidated by Igbavboa and Leistner.[224]

It is interesting to compare the mechanism of the Dieckmann reaction catalyzed by MenB with the retro-Dieckmann reaction catalyzed by 2-ketocyclohexanecarboxyl-CoA hydrolase (BadI). BadI catalyzes the hydrolysis of 2*S*-ketocyclohexanecarboxyl-CoA to pimelyl-CoA during the anaerobic catabolism of benzoate by *Rhodopseudomonas palustris*.[225] Eberhard and Gerlt[215] have determined the stereochemistry of the BadI reaction and have shown that during substrate hydrolysis the *pro*-2*S* proton is incorporated into the product, and thus that the reaction proceeds through inversion of configuration (**Figure 43**). As MenB abstracts the *pro*-2*R* proton from *O*-succinylbenzoate-CoA, the two enzymes have opposite stereochemistries, which is surprising given that the sequences of MenB and BadI are more than 50% identical. The active site residues in MenB, S190, D192, and Y287 are conserved in BadI (S138, D140, and Y235), and presumably occupy similar positions in the two active sites. Although the C2 succinyl carbon of OSB must approach the aromatic carboxylate during 1,4-dihydroxy-2-naphthoyl-CoA (DHNA)-CoA formation, it is not known what the initial conformation of
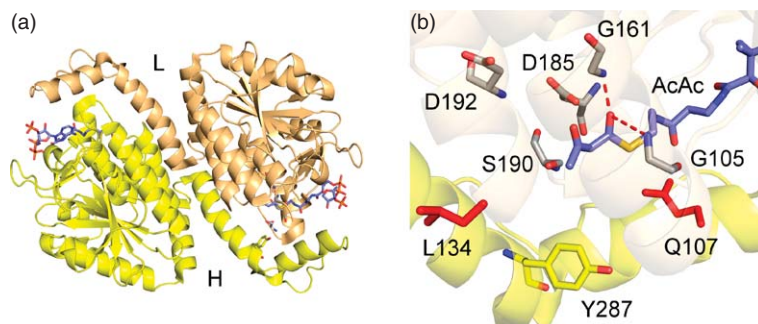

Figure 41   Structure of MenB, the dihydroxynaphthoyl-CoA synthase from *M. tuberculosis.* (a) Two monomers from the MenB hexamer showing the C-terminus of one monomer crossing the trimer–trimer interface to contribute to the active site of the adjacent monomer. (b) Active site showing the location of conserved catalytic residues from subunit L (gray) or H (yellow). The AcAc ligand is shown in slate and the residues in red are at the ends of a loop that is disordered in the crystal structure.
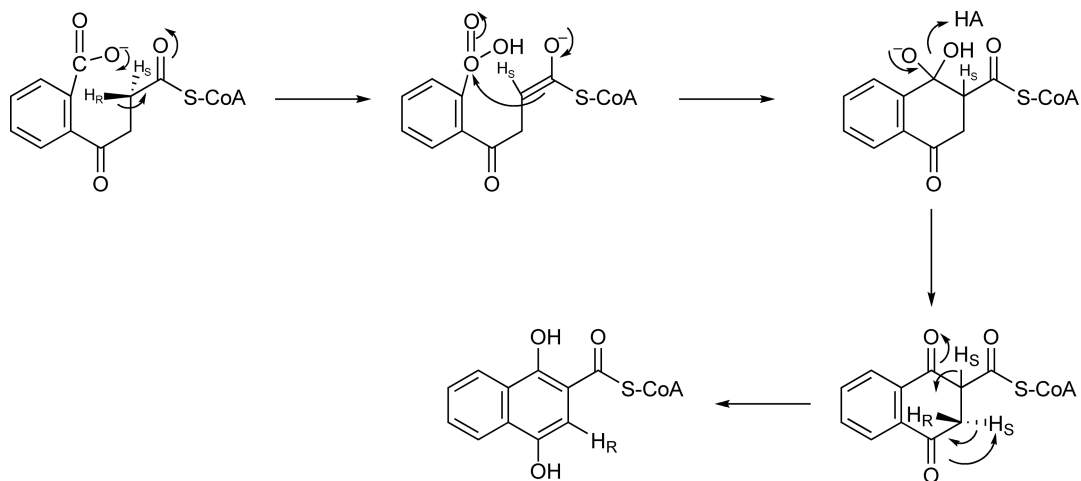
Figure 42 Mechanism of the MenB reaction.
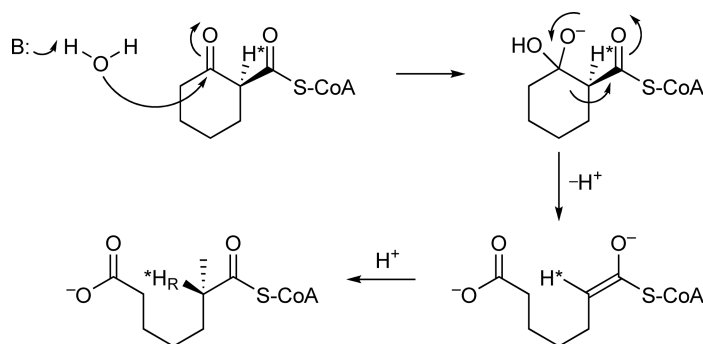
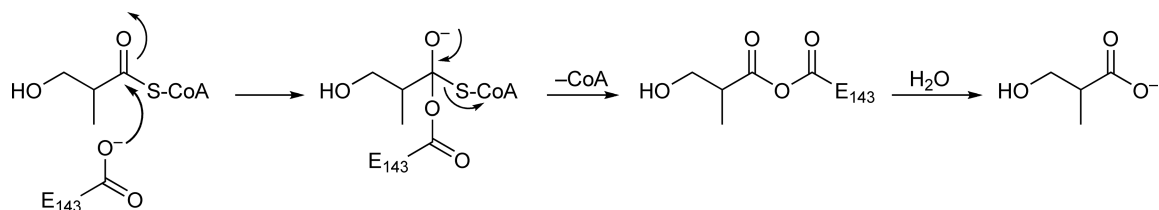**Figure 43** Mechanism of the BadI reaction.

the substrate is in the MenB active site. Further insight into the mechanisms of these two enzymes must await the determination of structures of catalytically relevant enzyme–ligand complexes.

8.07.4.3.3 3-Hydroxyisobutyryl-CoA hydrolase

During the course of their studies on the crotonase superfamily, Gerlt and coworkers have identified and characterized a number of new family members including 3(S)-methylglutaconyl-CoA hydratase and methyl-malonyl-CoA decarboxylase.[210,212] They have also studied 3-hydroxyisobutyryl-CoA hydrolase from *P. aeruginosa.* Sequence homology indicates that this enzyme contains a glutamate (E143) that is homologous to E164 in ECH. Interestingly, ^{18}O-labeling studies indicate that rather than acting as a base to catalyze the attack of water, E143 acts as a nucleophile, generating an anhydride with the substrate that subsequently hydrolyzes (**Figure 44**). The enzyme also contains the consensus GGG sequence that provides one of the hydrogen bonds in the crotonase oxyanion hole, which instead of stabilizing an enolate in this case stabilizes tetrahedral transition states during intermediate formation and breakdown.

8.07.5 Summary and Future Prospects

The biosynthesis and breakdown of fatty acids are both essential processes in cellular metabolism. Fatty acids are needed for critical structures such as cell membranes and cell walls, whereas fatty acid catabolism is a mechanism of energy generation in the cell. Although these metabolic pathways are present in all living

Figure 44 Mechanism of the 3-hydroxyisobutyryl-CoA hydrolase reaction.

organisms, the catalytic reactions performed by microbial fatty acid biosynthesis and by mammalian β-oxidation are the best understood, since these reactions are catalyzed by individual proteins that are amenable to expression, purification, and characterization. Common mechanistic themes linking enzymes in each pathway have been highlighted and have been used in this review as a foundation for discussing the reactions catalyzed by three superfamilies: the thiolase superfamily, the SDR superfamily, and the crotonase superfamily. In addition, the enzymes in both pathways are linked at an even more fundamental level, since they all catalyze reactions in which one (or in some cases both) substrates are thioesters either of ACP or of CoA. Compared to the carboxylic acids from which they are derived, thioesters are intrinsically more reactive because of an increase in both α-proton acidity and carbonyl electrophilicity. These chemical properties are enhanced by each enzyme through interactions that stabilize the accumulation of negative charge on the thioester carbonyl resulting, in many cases, in the formation and stabilization of carbanions or of carbanionic transition states.

Although further mechanistic details on the reactions catalyzed by each enzyme will continue to be elucidated, future studies are likely to rapidly expand to include the multienzyme homologs, driven in particular by recent advances in structural biology. Notably this includes the recent 3.2 Å structures of a Type I mammalian FAS,[226] which will not only shed light on the functioning of this multienzyme complex as well as related systems involved in polyketide biosynthesis,[227] but may also provide insight into the functioning of the putative noncovalent multienzyme complex thought to be formed by components of the microbial FAS-II pathway.[228–230] In addition, a major area of continued growth will involve additional validation of pathway components as targets for drug discovery and the development of potent enzyme inhibitors against these targets. The KAS and enoyl-ACP reductase enzymes in the microbial fatty acid biosynthesis pathway are the most advanced targets, and currently Affinium Pharmaceuticals are conducting Phase I clinical trials on an enoyl-ACP reductase (FabI) inhibitor for the treatment of sensitive and resistant staphylococcal infections.[231] Although other components of the FAS-II pathway are also potential targets, structures of the mammalian FAS will aid efforts to develop inhibitors of this enzyme system, based on the premise that *de novo* fatty acid synthesis is essential for tumorigenesis.[232] Although inhibitor discovery efforts will continue to be underpinned by screening programs that identify lead compounds and by the synthesis of compound libraries for lead optimization, rational structure-based approaches will also play a fundamental role in the development of novel chemotherapeutics. An understanding of the mechanistic basis for enzyme inhibition is naturally of critical importance in this process, since it not only informs on the specificity and selectivity of enzyme inhibitors, but also drives inhibitor discovery by providing insight into subtleties such as the behavior of slow-onset inhibitors. In this regard, a number of studies now support the importance of slow-onset inhibitors in the drug discovery process, since these compounds will have long residence times on the enzyme target, and thus increased activity *in vivo*.[21,123,124] Developing assays to identify compounds that dissociate slowly from the enzyme target and optimizing inhibitors for increased residence time requires a precise understanding of the slow structural change that accompanies formation of the final EI* and, more generally, detailed knowledge of the mechanism of the enzyme that is being targeted and of the structure of transitions states formed on the reaction pathway. This review highlights the progress that has been made in understanding the mechanism of enzymes involved in the metabolism of fatty acids and related compounds, to the point that we can now envisage the development of compounds that selectively antagonize these metabolic processes in the cell.

References

1. M. Mathieu; J. P. Zeelen; R. A. Pauptit; R. Erdmann; W. H. Kunau; R. K. Wierenga, *Structure* **1994**, *2*, 797.
2. M. Mathieu; Y. Modis; J. P. Zeelen; C. K. Engel; R. A. Abagyan; A. Ahlberg; B. Rasmussen; V. S. Lamzin; W. H. Kunau; R. K. Wierenga, *J. Mol. Biol.* **1997**, *273*, 714.
3. Y. Modis; R. K. Wierenga, *Structure* **1999**, *7*, 1279.
4. A. M. Haapalainen; G. Merilainen; R. K. Wierenga, *Trends Biochem. Sci.* **2006**, *31*, 64.
5. J. L. Garwin; A. L. Klages; J. E. Cronan, Jr., *J. Biol. Chem.* **1980**, *255*, 11949.
6. J. L. Garwin; A. L. Klages; J. E. Cronan, Jr. *J. Biol. Chem.* **1980**, *255*, 3263.
7. R. J. Heath; C. O. Rock, *Nat. Prod. Rep.* **2002**, *19*, 581.
8. L. Y. Gao; F. Laval; E. H. Lawson; R. K. Groger; A. Woodruff; J. H. Morisaki; J. S. Cox; M. Daffe; E. J. Brown, *Mol. Microbiol.* **2003**, *49*, 1547.
9. R. A. Slayden; C. E. Barry, 3rd, *Tuberculosis* **2002**, *82*, 149.
10. M. L. Schaeffer; G. Agnihotri; C. Volker; H. Kallender; P. J. Brennan; J. T. Lonsdale, *J. Biol. Chem.* **2001**, *276*, 47029.
11. P. J. Tonge; P. R. Carey, *Biochemistry,* **1992**, *31*, 9122.
12. D. Sali; M. Bycroft; A. R. Fersht, *Nature* **1988**, *335*, 740.
13. H. Nicholson; D. E. Anderson; S. Dao-pin; B. W. Matthews, *Biochemistry* **1991**, *30*, 9816.
14. D. J. Lockhart; P. S. Kim, *Science,* **1992**, *257*, 947.
15. T. Kortemme; T. E. Creighton, *J. Mol. Biol.* **1995**, *253*, 799.
16. W. G. Hol; P. T. van Duijnen; H. J. Berendsen, *Nature* **1978**, *273*, 443.
17. C. Davies; R. J. Heath; S. W. White; C. O. Rock, *Structure* **2000**, *8*, 185.
18. J. G. Olsen; A. Kadziola; P. von Wettstein-Knowles; M. Siggaard-Andersen; S. Larsen, *Structure* **2001**, *9*, 233.
19. K. A. McGuire; M. Siggaard-Andersen; M. G. Bangera; J. G. Olsen; P. von Wettstein-Knowles, *Biochemistry* **2001**, *40*, 9836.
20. R. J. Heath; S. W. White; C. O. Rock, *Appl. Microbiol. Biotechnol.* **2002**, *58*, 695.
21. H. Lu; P. J. Tonge, *Acc. Chem. Res.* **2008**, *41*, 11.
22. H. T. Wright; K. A. Reynolds, *Curr. Opin. Microbiol.* **2007**, *10*, 447.
23. T. J. Sullivan; J. J. Truglio; M. E. Boyne; P. Novichenok; X. Zhang; C. F. Stratton; H.-J. Li; T. Kaur; A. Amin; F. Johnson; R. A. Slayden; C. Kisker; P. J. Tonge, *ACS Chem. Biol.* **2006**, *1*, 43.
24. Y. Sato; S. Nomura; Y. Kamio; S. Omura; T. Hata, *J. Antibiot.* **1967**, *20*, 344.
25. H. Oishi; T. Noto; H. Sasaki; K. Suzuki; T. Hayashi; H. Okazaki; K. Ando; M. Sawada, *J. Antibiot.* **1982**, *35*, 391.
26. J. Wang; S. M. Soisson; K. Young; W. Shoop; S. Kodali; A. Galgoci; R. Painter; G. Parthasarathy; Y. S. Tang; R. Cummings; S. Ha; K. Dorso; M. Motyl; H. Jayasuriya; J. Ondeyka; K. Herath; C. Zhang; L. Hernandez; J. Allocco; A. Basilio; J. R. Tormo; O. Genilloud; F. Vicente; F. Pelaez; L. Colwell; S. H. Lee; B. Michael; T. Felcetto; C. Gill; L. L. Silver; J. D. Hermes; K. Bartizal; J. Barrett; D. Schmatz; J. W. Becker; D. Cully; S. B. Singh, *Nature* **2006**, *441*, 358.
27. M. M. Alhamadsheh; F. Musayev; A. A. Komissarov; S. Sachdeva; H. T. Wright; N. Scarsdale; G. Florova; K. A. Reynolds, *Chem. Biol.* **2007**, *14*, 513.
28. R. A. Daines; I. Pendrak; K. Sham; G. S. Van Aller; A. K. Konstantinidis; J. T. Lonsdale; C. A. Janson; X. Qiu; M. Brandt; S. S. Khandekar; C. Silverman; M. S. Head, *J. Med. Chem.* **2003**, *46*, 5.
29. J. Wang; S. Kodali; S. H. Lee; A. Galgoci; R. Painter; K. Dorso; F. Racine; M. Motyl; L. Hernandez; E. Tinney; S. L. Colletti; K. Herath; R. Cummings; O. Salazar; I. Gonzalez; A. Basilio; F. Vicente; O. Genilloud; F. Pelaez; H. Jayasuriya; K. Young; D. F. Cully; S. B. Singh, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 7612.
30. A. C. Price; K. H. Choi; R. J. Heath; Z. Li; S. W. White; C. O. Rock, *J. Biol. Chem.* **2001**, *276*, 6551.
31. M. Moche; G. Schneider; P. Edwards; K. Dehesh; Y. Lindqvist, *J. Biol. Chem.* **1999**, *274*, 6031.
32. M. B. Austin; J. P. Noel, *Nat. Prod. Rep.* **2003**, *20*, 79.
33. K. H. Choi; L. Kremer; G. S. Besra; C. O. Rock, *J. Biol. Chem.* **2000**, *275*, 28201.
34. S. Sachdeva; F. N. Musayev; M. M. Alhamadsheh; J. N. Scarsdale; H. T. Wright; K. A. Reynolds, *Chem. Biol.* **2008**, *15*, 402.
35. M. M. Alhamadsheh; N. C. Waters; D. P. Huddler; M. Kreishman-Deitrick; G. Florova; K. A. Reynolds, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 879.
36. X. He; A. M. Reeve; U. R. Desai; G. E. Kellogg; K. A. Reynolds, *Antimicrob. Agents Chemother.* **2004**, *48*, 3093.
37. W. L. Delano, http://www.pymol.org, **2002**.
38. S. R. Luckner; C. A. Machutta; P. J. Tonge; C. Kisker, *Structure* **2009**, *17*, 1.
39. M. B. Austin; M. Izumikawa; M. E. Bowman; D. W. Udwary; J. L. Ferrer; B. S. Moore; J. P. Noel, *J. Biol. Chem.* **2004**, *279*, 45162.
40. J. M. Jez; J. P. Noel, *J. Biol. Chem.* **2000**, *275*, 39640.
41. J. M. Jez; M. B. Austin; J. Ferrer; M. E. Bowman; J. Schroder; J. P. Noel, *Chem. Biol.* **2000**, *7*, 919.
42. J. M. Jez; J. L. Ferrer; M. E. Bowman; R. A. Dixon; J. P. Noel, *Biochemistry* **2000**, *39*, 890.
43. J. L. Ferrer; J. M. Jez; M. E. Bowman; R. A. Dixon; J. P. Noel, *Nat. Struct. Biol.* **1999**, *6*, 775.
44. M. J. Theisen; I. Misra; D. Saadat; N. Campobasso; H. M. Miziorko; D. H. Harrison, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16442.
45. B. J. Bahnson, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 16399.
46. W.-H. Kunau; V. Dommes; H. Schulz, *Prog. Lipid Res.* **1995**, *34*, 267.
47. Y. Modis; R. K. Wierenga, *J. Mol. Biol.* **2000**, *297*, 1171.
48. S. Thompson; F. Mayerl; O. P. Peoples; S. Masamune; A. J. Sinskey; C. T. Walsh, *Biochemistry* **1989**, *28*, 5735.
49. H. F. Gilbert; B. J. Lennox; C. D. Mossman; W. C. Carle, *J. Biol. Chem.* **1981**, *256*, 7371.
50. S. Masamune; M. A. J. Palmer; R. Gamboni; S. Thompson; J. T. Davis; S. F. Williams; O. P. Peoples; A. J. Sinskey; C. T. Walsh, *J. Am. Chem. Soc.* **1989**, *111*, 1879.
51. M. A. Palmer; E. Differding; R. Gamboni; S. F. Williams; O. P. Peoples; C. T. Walsh; A. J. Sinskey; S. Masamune, *J. Biol. Chem.* **1991**, *266*, 8369.
52. S. F. Williams; M. A. Palmer; O. P. Peoples; C. T. Walsh; A. J. Sinskey; S. Masamune, *J. Biol. Chem.* **1992**, *267*, 16041.
53. H. F. Gilbert, *Biochemistry* **1981**, *20*, 5643.

54. D. Hyndman; D. R. Bauman; V. V. Heredia; T. M. Penning, *Chem. Biol. Interact.* **2003**, *143–144*, 621.

55. J. M. Jez; M. J. Bennett; B. P. Schlegel; M. Lewis; T. M. Penning, *Biochem. J.* **1997**, *326* (Pt 3), 625.

56. G. Sanli; J. I. Dudley; M. Blaber, *Cell Biochem. Biophys.* **2003**, *38*, 79.

57. T. M. Penning, *J. Steroid Biochem. Mol. Biol.* **1999**, *69*, 211.

58. F. Hoffmann; E. Maser, *Drug Metab. Rev.* **2007**, *39*, 87.

59. U. Oppermann; C. Filling; M. Hult; N. Shafqat; X. Wu; M. Lindh; J. Shafqat; E. Nordling; Y. Kallberg; B. Persson; H. Jornvall, *Chem. Biol. Interact.* **2003**, *143–144*, 247.

60. Y. Kallberg; U. Oppermann; H. Jornvall; B. Persson, *Protein Sci.* **2002**, *11*, 636.

61. H. Jörnvall; B. Persson; M. Krook; S. Atrian; R. Gonzàlez-Duarte; J. Jeffery; D. Ghosh, *Biochemistry* **1995**, *34*, 6003.

62. B. Persson; M. Krook; H. Jornvall, *Eur. J. Biochem.* **1991**, *200*, 537.

63. O. Danielsson; S. Atrian; T. Luque; L. Hjelmqvist; R. Gonzalez-Duarte; H. Jornvall, *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 4980.

64. N. Tanaka; T. Nonaka; T. Tanabe; T. Yoshimoto; D. Tsuru; Y. Mitsui, *Biochemistry* **1996**, *35*, 7715.

65. Y. J. Liu; J. B. Thoden; J. Kim; E. Berger; A. M. Gulick; F. J. Ruzicka; H. M. Holden; P. A. Frey, *Biochemistry* **1997**, *36*, 10675.

66. J. B. Thoden; A. D. Hegeman; G. Wesenberg; M. C. Chapeau; P. A. Frey; H. M. Holden, *Biochemistry* **1997**, *36*, 6294.

67. T. Tanabe; N. Tanaka; K. Uchikawa; T. Kabashima; K. Ito; T. Nonaka; Y. Mitsui; M. Tsuru; T. Yoshimoto, *J. Biochem.* **1998**, *124*, 634.

68. M. P. Patel; W. S. Liu; J. West; D. Tew; T. D. Meek; S. H. Thrall, *Biochemistry* **2005**, *44*, 16753.

69. M. Fisher; J. T. Kroon; W. Martindale; A. R. Stuitje; A. R. Slabas; J. B. Rafferty, *Structure* **2000**, *8*, 339.

70. A. C. Price; Y. M. Zhang; C. O. Rock; S. W. White, *Biochemistry* **2001**, *40*, 12772.

71. M. Cohen-Gonsaud; S. Ducasse; F. Hoh; D. Zerbib; G. Labesse; A. Quemard, *J. Mol. Biol.* **2002**, *320*, 249.

72. S. Pillai; C. Rajagopal; M. Kapoor; G. Kumar; A. Gupta; N. Surolia, *Biochem. Biophys. Res. Commun.* **2003**, *303*, 387.

73. S. R. Wickramasinghe; K. A. Inglis; J. E. Urch; S. Muller; D. M. van Aalten; A. H. Fairlamb, *Biochem. J.* **2006**, *393*, 447.

74. A. C. Price; Y. M. Zhang; C. O. Rock; S. W. White, *Structure* **2004**, *12*, 417.

75. R. G. Silva; L. P. de Carvalho; J. S. Blanchard; D. S. Santos; L. A. Basso, *Biochemistry* **2006**, *45*, 13064.

76. R. G. Silva; L. A. Rosado; D. S. Santos; L. A. Basso, *Arch. Biochem. Biophys.* **2008**, *471*, 1.

77. H. Marrakchi; S. Ducasse; G. Labesse; H. Montrozier; E. Margeat; L. Emorine; X. Charpentier; M. Daffe; A. Quemard, *Microbiology* **2002**, *148*, 951.

78. D. Tasdemir; G. Lack; R. Brun; P. Ruedi; L. Scapozza; R. Perozzo, *J. Med. Chem.* **2006**, *49*, 3345.

79. S. Ducasse-Cabanot; M. Cohen-Gonsaud; H. Marrakchi; M. Nguyen; D. Zerbib; J. Bernadou; M. Daffe; G. Labesse; A. Quemard, *Antimicrob. Agents Chemother.* **2004**, *48*, 242.

80. H. Bergler; P. Wallner; A. Ebeling; B. Leitinger; S. Fuchsbichler; H. Aschauer; G. Kollenz; G. Hogenauer; F. Turnowsky, *J. Biol. Chem.* **1994**, *269*, 5493.

81. Y. M. Zhang; H. Marrakchi; S. W. White; C. O. Rock, *J. Lipid Res.* **2003**, *44*, 1.

82. R. P. Massengo-Tiasse; J. E. Cronan, *J. Biol. Chem.* **2007**, *283*, 1308.

83. R. J. Heath; C. O. Rock, *Nature* **2000**, *406*, 145.

84. R. J. Heath; N. Su; C. K. Murphy; C. O. Rock, *J. Biol. Chem.* **2000**, *275*, 40128.

85. J. Saito; M. Yamada; T. Watanabe; M. Iida; H. Kitagawa; S. Takahata; T. Ozawa; Y. Takeuchi; F. Ohsawa, *Protein Sci.* **2008**, *17*, 691.

86. F. Turnowsky; K. Fuchs; C. Jeschek; G. Hogenauer, *J. Bacteriol.* **1989**, *171*, 6555.

87. H. Bergler; G. Hogenauer; F. Turnowsky, *J. Gen. Microbiol.* **1992**, *138*, 2093.

88. R. J. Heath; C. O. Rock, *J. Biol. Chem.* **1995**, *270*, 26538.

89. H. Bergler; S. Fuchsbichler; G. Hogenauer; F. Turnowsky, *Eur. J. Biochem.* **1996**, *242*, 689.

90. Y. Ji; D. Yin; B. Fox; D. J. Holmes; D. Payne; M. Rosenberg, *FEMS Microbiol. Lett.* **2004**, *231*, 177.

91. A. Quemard; J. C. Sacchettini; A. Dessen; C. Vilcheze; R. Bittman; W. R. Jacobs, Jr.; J. S. Blanchard, *Biochemistry* **1995**, *34*, 8235.

92. K. L. Fillgrove; V. E. Anderson, *Biochemistry* **2000**, *39*, 7001.

93. C. Baldock; J. B. Rafferty; S. E. Sedelnikova; P. J. Baker; A. R. Stuitje; A. R. Slabas; T. R. Hawkes; D. W. Rice, *Science* **1996**, *274*, 2107.

94. A. Dessen; A. Quemard; J. S. Blanchard; W. R. Jacobs, Jr.; J. C. Sacchettini, *Science* **1995**, *267*, 1638.

95. R. Perozzo; M. Kuo; A. S. Sidhu; J. T. Valiyaveettil; R. Bittman; W. R. Jacobs, Jr.; D. A. Fidock; J. C. Sacchettini, *J. Biol. Chem.* **2002**, *277*, 13106.

96. H. H. Lee; J. Moon; S. W. Suh, *Proteins* **2007**, *69*, 691.

97. M. V. Dias; I. B. Vasconcelos; A. M. Prado; V. Fadel; L. A. Basso; W. F. de Azevedo, Jr.; D. S. Santos, *J. Struct. Biol.* **2007**, *159*, 369.

98. D. A. Rozwarski; C. Vilcheze; M. Sugantino; R. Bittman; J. C. Sacchettini, *J. Biol. Chem.* **1999**, *274*, 15582.

99. M. J. Stewart; S. Parikh; G. Xiao; P. J. Tonge; C. Kisker, *J. Mol. Biol.* **1999**, *290*, 859.

100. J. B. Rafferty; J. W. Simon; C. Baldock; P. J. Artymiuk; P. J. Baker; A. R. Stuitje; A. R. Slabas; D. W. Rice, *Structure* **1995**, *3*, 927.

101. S. Parikh; D. P. Moynihan; G. Xiao; P. J. Tonge, *Biochemistry* **1999**, *38*, 13623.

102. K. L. Fillgrove; V. E. Anderson, *Biochemistry* **2001**, *40*, 12412.

103. A. F. Bell; C. F. Stratton; X. Zhang; P. Novichenok; A. A. Jaye; P. A. Nair; S. Parikh; R. Rawat; P. J. Tonge, *J. Am. Chem. Soc.* **2007**, *129*, 6425.

104. J. Marcinkeviciene; W. Jiang; L. M. Kopcho; G. Locke; Y. Luo; R. A. Copeland, *Arch. Biochem. Biophys.* **2001**, *390*, 101.

105. H. Xu; T. J. Sullivan; J. Sekiguchi; T. Kirikae; I. Ojima; W. Mao; F. L. Rock; M. R. K. Alley; F. Johnson; S. G. Walker; P. J. Tonge, *Biochemistry* **2008**, *47*, 4228.

106. A. R. Slabas; C. Sidebottom; R. Kessell; A. Hellyer; M. P. Tombs, *Biochem. Soc. Trans.* **1986**, *14*, 581.

107. Y. M. Zhang; M. S. Rao; R. J. Heath; A. C. Price; A. J. Olson; C. O. Rock; S. W. White, *J. Biol. Chem.* **2001**, *276*, 8231.

108. Y. M. Zhang; B. Wu; J. Zheng; C. O. Rock, *J. Biol. Chem.* **2003**, *278*, 52935.

109. S. Rafi; P. Novichenok; S. Kolappan; X. Zhang; C. F. Stratton; R. Rawat; C. Kisker; C. Simmerling; P. J. Tonge, *J. Biol. Chem.* **2006**, *281*, 39285.

110. M. A. Seefeld; W. H. Miller; K. A. Newlander; W. J. Burgess; W. E. DeWolf; Jr.; P. A. Elkins; M. S. Head; D. R. Jakas; C. A. Janson; P. M. Keller; P. J. Manley; T. D. Moore; D. J. Payne; S. Pearson; B. J. Polizzi; X. Qiu; S. F. Rittenhouse; I. N. Uzinskas; N. G. Wallis; W. F. Huffman, *J. Med. Chem.* **2003**, *46*, 1627.

111. W. H. Miller; M. A. Seefeld; K. A. Newlander; I. N. Uzinskas; W. J. Burgess; D. A. Heerding; C. C. Yuan; M. S. Head; D. J. Payne; S. F. Rittenhouse; T. D. Moore; S. C. Pearson; V. Berry; W. E. DeWolf, Jr.; P. M. Keller; B. J. Polizzi; X. Qiu; C. A. Janson; W. F. Huffman, *J. Med. Chem.* **2002**, *45*, 3246.
112. J. H. Yum; C. K. Kim; D. Yong; K. Lee; Y. Chong; C. M. Kim; J. M. Kim; S. Ro; J. M. Cho, *Antimicrob. Agents Chemother.* **2007**, *51*, 2591.
113. A. Roujeinikova; S. Sedelnikova; G. J. de Boer; A. R. Stuitje; A. R. Slabas; J. B. Rafferty; D. W. Rice, *J. Biol. Chem.* **1999**, *274*, 30811.
114. D. A. Rozwarski; G. A. Grant; D. H. R. Barton; W. R. Jacobs, Jr.; J. C. Sacchettini, *Science* **1998**, *279*, 98.
115. A. Banerjee; E. Dubnau; A. Quemard; V. Balasubramanian; K. S. Um; T. Wilson; D. Collins; G. de Lisle; W. R. Jacobs, Jr., *Science* **1994**, *263*, 227.
116. K. Johnsson; D. S. King; P. G. Schultz, *J. Am. Chem. Soc.* **1995**, *117*, 5009.
117. J. A. Marcinkeviciene; R. S. Magliozzo; J. S. Blanchard, *J. Biol. Chem.* **1995**, *270*, 22290.
118. R. Rawat; A. Whitty; P. J. Tonge, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13881.
119. W. H. Ward; G. A. Holdgate; S. Rowsell; E. G. McLean; R. A. Pauptit; E. Clayton; W. W. Nichols; J. G. Colls; C. A. Minshull; D. A. Jude; A. Mistry; D. Timms; R. Camble; N. J. Hales; C. J. Britton; I. W. Taylor, *Biochemistry* **1999**, *38*, 12514.
120. S. Sivaraman; T. J. Sullivan; F. Johnson; P. Novichenok; G. Cui; C. Simmerling; P. J. Tonge, *J. Med. Chem.* **2004**, *47*, 509.
121. S. Sivaraman; J. Zwahlen; A. F. Bell; L. Hedstrom; P. J. Tonge, *Biochemistry* **2003**, *42*, 4406.
122. S. L. Parikh; G. Xiao; P. J. Tonge, *Biochemistry* **2000**, *39*, 7645.
123. R. A. Copeland; D. L. Pompliano; T. D. Meek, *Nat. Rev. Drug Discovery* **2006**, *5*, 730.
124. H. Lu; K. England; C. W. am Ende; J. J. Truglio; S. Luckner; B. G. Reddy; N. L. Marlenee; S. E. Knudson; D. L. Knudson; R. A. Bowen; C. Kisker; R. A. Slayden; P. J. Tonge, *ACS Chem. Biol.* **2009**, *4*, 221.
125. M. E. Boyne; T. J. Sullivan; C. W. amEnde; H. Lu; V. Gruppo; D. Heaslip; A. G. Amin; D. Chatterjee; A. Lenaerts; P. J. Tonge; R. A. Slayden, *Antimicrob. Agents Chemother.* **2007**, *51*, 3562.
126. P. J. Tonge; C. Kisker; R. A. Slayden, *Curr. Top. Med. Chem.* **2007**, *7*, 489.
127. C. W. am Ende; S. E. Knudson; N. Liu; J. Childs; T. J. Sullivan; M. Boyne; H. Xu; Y. Gegina; D. L. Knudson; F. Johnson; C. A. Peloquin; R. A. Slayden; P. J. Tonge, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 3029.
128. J. J. P. Kim; R. Miura, *Eur. J. Biochem.* **2004**, *271*, 483.
129. S. Ghisla; C. Thorpe, *Eur. J. Biochem.* **2004**, *271*, 494.
130. N. Gregersen; P. Bross; B. S. Andresen, *Eur. J. Biochem.* **2004**, *271*, 470.
131. C. Thorpe; J. J. Kim, *FASEB J.* **1995**, *9*, 718.
132. J. J. P. Kim; M. Wang; R. Paschke, *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 7523.
133. S. Ghisla; C. Thorpe; V. Massey, *Biochemistry* **1984**, *23*, 3154.
134. B. Pohl; T. Raichle; S. Ghisla, *Eur. J. Biochem.* **1986**, *160*, 109.
135. L. M. Schopfer; V. Massey; S. Ghisla; C. Thorpe, *Biochemistry* **1988**, *27*, 6599.
136. T. D. Poulsen; M. Garcia-Viloca; J. L. Gao; D. G. Truhlar, *J. Phys. Chem. B.* **2003**, *107*, 9567.
137. J. Wu; A. F. Bell; L. Luo; A. W. Stephens; M. T. Stankovich; P. J. Tonge, *Biochemistry* **2003**, *42*, 11846.
138. Y. Ikeda; D. G. Hine; K. Okamura-Ikeda; K. Tanaka, *J. Biol. Chem.* **1985**, *260*, 1326.
139. S. M. Lau; R. K. Brantley; C. Thorpe, *Biochemistry* **1988**, *27*, 5089.
140. P. Vock; S. Engst; M. Eder; S. Ghisla, *Biochemistry* **1998**, *37*, 1848.
141. J. K. Johnson; Z. X. Wang; D. K. Srivastava, *Biochemistry* **1992**, *31*, 10564.
142. J. Wu; A. F. Bell; A. A. Jaye; P. J. Tonge, *J. Am. Chem. Soc.* **2005**, *127*, 8424.
143. J. D. Pellett; K. M. Sabaj; A. W. Stephens; A. F. Bell; J. Wu; P. J. Tonge; M. T. Stankovich, *Biochemistry* **2000**, *39*, 13982.
144. S. M. Lau; P. Powell; H. Buettner; S. Ghisla; C. Thorpe, *Biochemistry* **1986**, *25*, 4184.
145. I. Rudik; C. Thorpe, *Arch. Biochem. Biophys.* **2001**, *392*, 341.
146. S. Engst; P. Vock; M. Wang; J. J. P. Kim; S. Ghisla, *Biochemistry* **1999**, *38*, 257.
147. N. D. Lenn; M. T. Stankovich; H. W. Liu, *Biochemistry* **1990**, *29*, 3709.
148. B. D. Johnson; G. J. Mancini-Samuelson; M. T. Stankovich, *Biochemistry* **1995**, *34*, 7047.
149. P. C. Engel; V. Massey, *Biochem. J.* **1971**, *125*, 879.
150. Y. Nishina; K. Sato; I. Hazekawa; K. Shiga, *J. Biochem.* **1995**, *117*, 800.
151. Y. Nishina; K. Sato; K. Shiga; S. Fujii; K. Kuroda; R. Miura, *J. Biochem.* **1992**, *111*, 699.
152. H. Tamaoki; Y. Nishina; K. Shiga; R. Miura, *J. Biochem.* **1999**, *125*, 285.
153. J. J. Barycki; L. K. O'Brien; A. W. Strauss; L. J. Banaszak, *J. Biol. Chem.* **2000**, *275*, 27186.
154. S. Y. Yang; X. Y. He; H. Schulz, *Febs J.* **2005**, *272*, 4874.
155. J. J. Barycki; L. K. O'Brien; J. M. Bratt; R. G. Zhang; R. Sanishvili; A. W. Strauss; L. J. Banaszak, *Biochemistry* **1999**, *38*, 5786.
156. J. J. Barycki; L. K. O'Brien; A. W. Strauss; L. J. Banaszak, *J. Biol. Chem.* **2001**, *276*, 36718.
157. J. P. Taskinen; T. R. Kiema; J. K. Hiltunen; R. K. Wierenga, *J. Mol. Biol.* **2006**, *355*, 734.
158. M. S. Ylianttila; N. V. Pursiainen; A. M. Haapalainen; A. H. Juffer; Y. Poirier; J. K. Hiltunen; T. Glumoff, *J. Mol. Biol.* **2006**, *358*, 1286.
159. D. J. Brock; L. R. Kass; K. Bloch, *J. Biol. Chem.* **1967**, *242*, 4432.
160. L. R. Kass; D. J. Brock; K. Bloch, *J. Biol. Chem.* **1967**, *242*, 4418.
161. J. E. Cronan, Jr.; W. B. Li; R. Coleman; M. Narasimhan; D. de Mendoza; J. M. Schwab, *J. Biol. Chem.* **1988**, *263*, 4641.
162. D. P. Clark; D. DeMendoza; M. L. Polacco; J. E. Cronan, Jr., *Biochemistry* **1983**, *22*, 5897.
163. S. Mohan; T. M. Kelly; S. S. Eveland; C. R. Raetz; M. S. Anderson, *J. Biol. Chem.* **1994**, *269*, 32896.
164. R. J. Heath; C. O. Rock, *J. Biol. Chem.* **1996**, *271*, 27795.
165. M. Leesong; B. S. Henderson; J. R. Gillig; J. M. Schwab; J. L. Smith, *Structure* **1996**, *4*, 253.
166. M. S. Kimber; F. Martin; Y. Lu; S. Houston; M. Vedadi; A. Dharamsi; K. M. Fiebig; M. Schmid; C. O. Rock, *J. Biol. Chem.* **2004**, *279*, 52593.
167. P. L. Swarnamukhi; S. K. Sharma; P. Bajaj; N. Surolia; A. Surolia; K. Suguna, *FEBS Lett.* **2006**, *580*, 2653.
168. L. Zhang; W. Liu; T. Hu; L. Du; C. Luo; K. Chen; X. Shen; H. Jiang, *J. Biol. Chem.* **2008**, *283*, 5370.

169. V. Bhowruth; A. K. Brown; G. S. Besra, *Microbiology* **2008**, *154*, 1866.
170. S. K. Sharma; M. Kapoor; T. N. Ramya; S. Kumar; G. Kumar; R. Modak; S. Sharma; N. Surolia; A. Surolia, *J. Biol. Chem.* **2003**, *278*, 45661.
171. R. R. Rando; K. Bloch, *J. Biol. Chem.* **1968**, *243*, 5627.
172. J. P. Richard; T. L. Amyes, *Curr. Opin. Chem. Biol.* **2001**, *5*, 626.
173. J. R. Mohrig; K. A. Moerke; D. L. Cloutier; B. D. Lane; E. C. Person; T. B. Onasch, *Science* **1995**, *269*, 527.
174. P. Willadsen; H. Eggerer, *Eur. J. Biochem.* **1975**, *54*, 247.
175. J. M. Schwab; A. Habib; J. B. Klassen, *J. Am. Chem. Soc.* **1986**, *108*, 5304.
176. V. E. Anderson; G. G. Hammes, *Biochemistry* **1984**, *23*, 2088.
177. O. Gawron; A. J. Glaid; A. Lomonte; S. Gary, *J. Am. Chem. Soc.* **1958**, *80*, 5856.
178. O. Gawron; T. P. Fondy, *J. Am. Chem. Soc.* **1959**, *81*, 6333.
179. M. Cohn; J. E. Pearson; E. L. Oconnell; I. A. Rose, *J. Am. Chem. Soc.* **1970**, *92*, 4095.
180. K. A. Reynolds; K. A. Holland, *Chem. Soc. Rev.* **1997**, *26*, 337.
181. W. H. Kunau; V. Dommes; H. Schulz, *Prog. Lipid Res.* **1995**, *34*, 267.
182. Y. Poirier; V. D. Antonenkov; T. Glumoff; J. K. Hiltunen, *Biochim. Biophys. Acta* **2006**, *1763*, 1413.
183. P. Bhaumik; M. K. Koski; T. Glumoff; J. K. Hiltunen; R. K. Wierenga, *Curr. Opin. Struct. Biol.* **2005**, *15*, 621.
184. H. A. Hofstein; Y. Feng; V. E. Anderson; P. J. Tonge, *Biochemistry* **1999**, *38*, 9508.
185. R. M. Waterson; R. L. Hill, *J. Biol. Chem.* **1972**, *247*, 5258.
186. A. Thibblin; W. P. Jencks, *J. Am. Chem. Soc.* **1979**, *101*, 4963.
187. J. A. Gerlt; J. W. Kozarich; G. L. Kenyon; P. G. Gassman, *J. Am. Chem. Soc.* **1991**, *113*, 9667.
188. J. A. Gerlt; P. G. Gassman, *J. Am. Chem. Soc.* **1992**, *114*, 5928.
189. B. J. Bahnson; V. E. Anderson, *Biochemistry* **1989**, *28*, 4173.
190. B. J. Bahnson; V. E. Anderson, *Biochemistry* **1991**, *30*, 5894.
191. B. J. Bahnson; V. E. Anderson; G. A. Petsko, *Biochemistry* **2002**, *41*, 2621.
192. R. L. Dordine; B. J. Bahnson; P. J. Tonge; V. E. Anderson, *Biochemistry* **1994**, *33*, 14733.
193. M. Dai; Y. Feng; P. J. Tonge, *J. Am. Chem. Soc.* **2001**, *123*, 506.
194. C. K. Engel; M. Mathieu; J. P. Zeelen; J. K. Hiltunen; R. K. Wierenga, *EMBO J.* **1996**, *15*, 5135.
195. C. K. Engel; T. R. Kiema; J. K. Hiltunen; R. K. Wierenga, *J. Mol. Biol.* **1998**, *275*, 847.
196. A. F. Bell; Y. Feng; H. A. Hofstein; S. Parikh; J. Wu; M. J. Rudolph; C. Kisker; A. Whitty; P. J. Tonge, *Chem. Biol.* **2002**, *9*, 1247.
197. G. Muller-Newen; W. Stoffel, *Biochemistry* **1993**, *32*, 11405.
198. G. Muller-Newen; U. Janssen; W. Stoffel, *Eur. J. Biochem.* **1995**, *228*, 68.
199. M. M. Benning; K. L. Taylor; R.-Q. Liu; G. Yang; H. Xiang; G. Wesenberg; D. Dunaway-Mariano; H. M. Holden, *Biochemistry* **1996**, *35*, 8103.
200. W. J. Wu; V. E. Anderson; D. P. Raleigh; P. J. Tonge, *Biochemistry* **1997**, *36*, 2211.
201. W. J. Wu; Y. Feng; X. He; H. S. Hofstein; D. P. Raleigh; P. J. Tonge, *J. Am. Chem. Soc.* **2000**, *122*, 3987.
202. Y. Feng; H. A. Hofstein; J. Zwahlen; P. J. Tonge, *Biochemistry* **2002**, *42*, 12883.
203. R. L. Dordine; P. J. Tonge; P. R. Carey; V. E. Anderson, *Biochemistry* **1994**, *33*, 12635.
204. R. L. D'Ordine; J. Pawlak; B. J. Bahnson; V. E. Anderson, *Biochemistry* **2002**, *41*, 2630.
205. P. J. Tonge; V. E. Anderson; R. Fausto; M. Kim; M. PusztaiCarey; P. R. Carey, *Biospectroscopy* **1995**, *1*, 387.
206. A. F. Bell; J. Wu; Y. Feng; P. J. Tonge, *Biochemistry* **2001**, *40*, 1725.
207. W. P. Jencks, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1975**, *43*, 219.
208. J. A. Gerlt; P. C. Babbitt, *Annu. Rev. Biochem.* **2001**, *70*, 209.
209. R. B. Hamed; E. T. Batchelar; I. J. Clifton; C. J. Schofield, *Cell. Mol. Life Sci.* **2008**.
210. B. J. Wong; J. A. Gerlt, *Biochemistry* **2004**, *43*, 4646.
211. B. J. Wong; J. A. Gerlt, *J. Am. Chem. Soc.* **2003**, *125*, 12076.
212. T. Haller; T. Buckel; J. Retey; J. A. Gerlt, *Biochemistry* **2000**, *39*, 4622.
213. R. Meganathan; R. Bentley; H. Taber, *J. Bacteriol.* **1981**, *145*, 328.
214. J. L. Whittingham; J. P. Turkenburg; C. S. Verma; M. A. Walsh; G. Grogan, *J. Biol. Chem.* **2003**, *278*, 1744.
215. E. D. Eberhard; J. A. Gerlt, *J. Am. Chem. Soc.* **2004**, *126*, 7188.
216. M. C. Sleeman; C. J. Schofield, *J. Biol. Chem.* **2004**, *279*, 6730.
217. B. Gerratana; S. O. Arnett; A. Stapon; C. A. Townsend, *Biochemistry* **2004**, *43*, 15936.
218. H. M. Holden; M. M. Benning; T. Haller; J. A. Gerlt, *Acc. Chem. Res.* **2001**, *34*, 145.
219. J. Dong; P. R. Carey; Y. S. Wei; L. S. Luo; X. F. Lu; R. Q. Liu; D. Dunaway-Mariano, *Biochemistry* **2002**, *41*, 7453.
220. J. Clarkson; P. J. Tonge; K. L. Taylor; D. Dunaway-Mariano; P. R. Carey, *Biochemistry* **1997**, *36*, 10192.
221. K. L. Taylor; R. Q. Liu; P. H. Liang; J. Price; D. Dunaway-Mariano; P. J. Tonge; J. Clarkson; P. R. Carey, *Biochemistry* **1995**, *34*, 13881.
222. J. Dong; X. F. Lu; Y. S. Wei; L. S. Luo; D. Dunaway-Mariano; P. R. Carey, *Biochemistry* **2003**, *42*, 9482.
223. J. J. Truglio; K. Theis; Y. Feng; R. Gajda; C. Machutta; P. J. Tonge; C. Kisker, *J. Biol. Chem.* **2003**, *278*, 42352.
224. U. Igbavboa; E. Leistner, *Eur. J. Biochem.* **1990**, *192*, 441.
225. D. A. Pelletier; C. S. Harwood, *J. Bacteriol.* **1998**, *180*, 2330.
226. T. Maier; M. Leibundgut; N. Ban, *Science* **2008**, *321*, 1315.
227. C. Khosla; Y. Tang; A. Y. Chen; N. A. Schnarr; D. E. Cane, *Annu. Rev. Biochem.* **2007**, *76*, 195.
228. R. Veyron-Churlet; S. Bigot; O. Guerrini; S. Verdoux; W. Malaga; M. Daffe; D. Zerbib, *J. Mol. Biol.* **2005**, *353*, 847.
229. N. A. Kruh; R. Rawat; B. P. Ruzsicska; P. J. Tonge, *Protein Sci.* **2007**, *16*, 1617.
230. N. A. Kruh; J. G. Borgaro; B. P. Ruzsicska; H. Xu; P. J. Tonge, *J. Biol. Chem.* **2008**, *283*, 31719.
231. J. A. Karlowsky; N. M. Laing; T. Baudry; N. Kaplan; D. Vaughan; D. J. Hoban, *Antimicrob. Agents Chemother.* **2007**, *51*, 1580.
232. F. P. Kuhajda; E. S. Pizer; J. N. Li; N. S. Mani; G. L. Frehywot; C. A. Townsend, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3450.

Biographical Sketches

Huaning Zhang was born in Beijing, China, on 26 September 1978. She graduated from Peking University and obtained her B.Sc. in medicinal chemistry in July 2001. Zhang joined Tyco Healthcare Pte. Ltd., Beijing Representative Office (China) as a drug registration specialist in July 2001, and was affiliated with Beijing Huizhong Tianyuan Technology Co. Ltd. (China) from January 2002 until July 2003. After spending 6 months in studying the biosynthesis of prodiginine in *Streptomyces coelicolor* with Professor Gregory L. Challis at the University of Warwick (UK), she joined the chemistry Ph.D. program at Stony Brook University, in 2004, where she is now a member of Professor Peter J. Tonge's research group. Her current research focuses on the mechanism and inhibition of enzymes from the menaquinone biosynthesis pathway in *Mycobacterium tuberculosis* with particular emphasis on the 1,4-dihydroxynaphthoyl-CoA synthase, MenB.

Carl A. Machutta was born on 12 June 1979 in Levittown, NY. He obtained his B.S. in biochemistry and Ph.D. in chemistry at Stony Brook University under the mentorship of Professor Peter J. Tonge. He is currently a postdoctoral associate in this laboratory and his research has focused on the biophysical, kinetic, and structural characterization of protein drug targets as well as their inhibition mechanism. Currently, interligand NOE NMR, a fragment-based approach, is being utilized to rationally modify a lead compound for the development of novel chemotherapeutics for *Mycobacterium tuberculosis* infection.

Peter J. Tonge was born in Northampton, England, in 1961. He obtained his B.Sc. and Ph.D. degrees in biochemistry at the University of Birmingham and then moved to the National Research Council of Canada (NRC) in 1986 as a NATO-SERC postdoctoral fellow. After spells at NRC as a research associate and a research officer, followed by an appointment as a staff investigator at The Picower Institute for Medical Research, he joined as a faculty at Stony Brook University in 1996, where he is currently a full professor. His research focuses on using precise information on enzyme mechanisms to develop enzyme inhibitors with a specific focus on antibacterial drug discovery. He also uses steady-state and ultrafast vibrational spectroscopy to probe enzyme mechanisms and to understand the photochemistry of fluorescent as well as light-activated proteins.

8.08 Diels–Alderases

Hideaki Oikawa, Hokkaido University, Sapporo, Japan

8.08.1 Introduction

The Diels–Alder reaction is synthetically very useful and is one of the most widely used reactions in organic synthesis because it forms a six-membered ring from a 1,3-diene and a dienophile with high regio- and stereoselectivity under mild conditions.[1] In addition, for creating four chiral centers or quaternary stereogenic centers in organic synthesis, the Diels–Alder reaction is a powerful tool and has been applied to the synthesis of complex pharmaceutical and biologically active compounds.[2–4]

Natural products presumably biosynthesized through a [4 + 2] cycloaddition frequently occur in the literature. Several reviews on natural Diels–Alder-type cycloadducts[5–10] covered more than 300 cycloadducts, including polyketides, terpenoids, phenylpropanoids, alkaloids, and natural products formed through mixed biosynthetic pathways. Representative examples of natural [4 + 2] adducts are shown in **Figure 1**. These include intramolecular adducts pinnatoxin B[11,12] (**5**) and nargenicin[13,14] (**6**), a simple intermolecular adduct plagiospiroside A[15] (**2**), tetramer quatromicin A_3[16] (**4**), and hetero-Diels–Alder dimeric adduct xuxuarine $E\alpha$[17] (**3**). In addition, there are unique metabolites: intra- and intermolecular [4 + 2] adduct longithorone A[18] (**1**). Although considerable efforts have been made to identify the enzymatic Diels–Alder reaction, there was no report on the enzyme catalyzing the Diels–Alder reaction until our first report in 1995.[19,20]
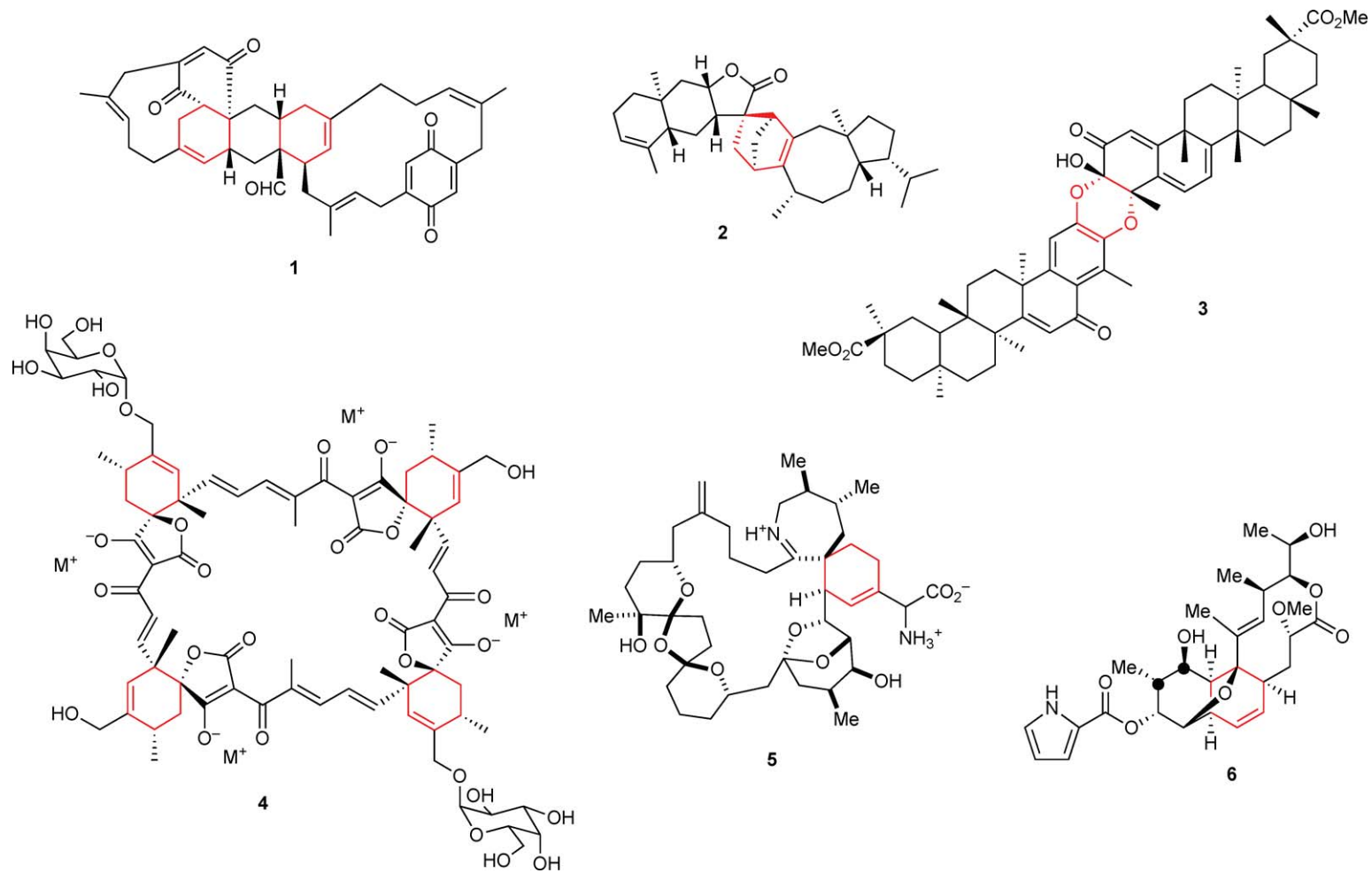
**Figure 1** Representative examples of natural [4 + 2] adducts.
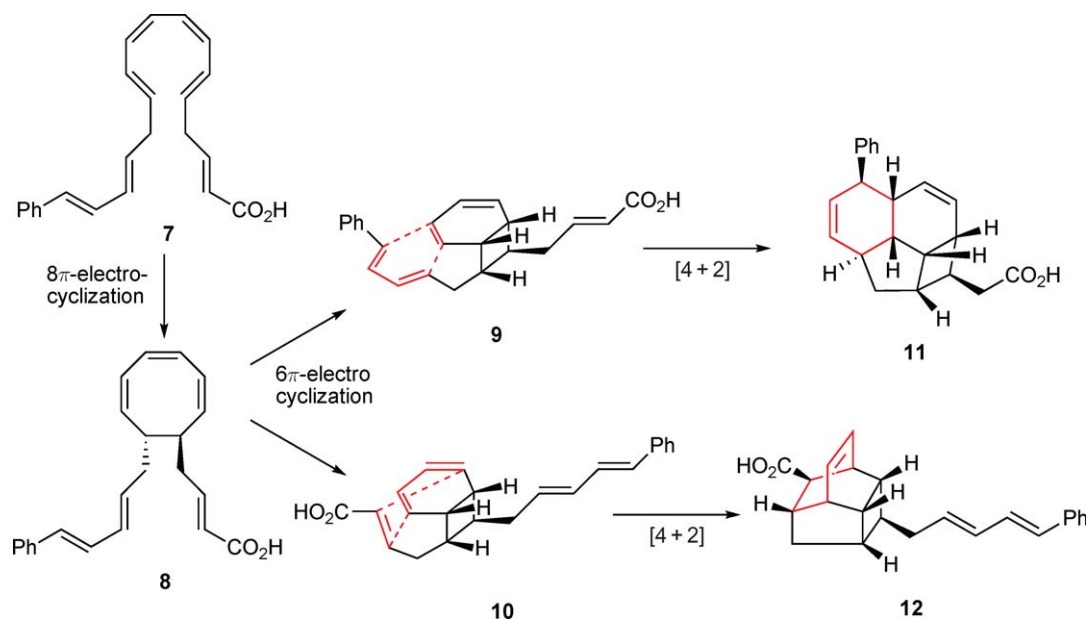
Scheme 1

The indication that natural products may be biosynthesized through the biological Diels–Alder reaction can be realized from several observations: (1) the isolation of an adduct with the corresponding precursor; (2) co-occurrence of adducts and their regio- and diastereoisomers; (3) nonenzymatic feasibility of a plausible cycloaddition; and (4) chirality of the adducts. Careful examination of the structures of the natural products sometimes provides useful information on biosynthesis. Biosynthesis of the endiandric acids is a good example of a natural [4 + 2] adduct. The endiandric acids B (11) and C (12) were isolated from the plant *Endiandra introsa* as a mixture of structurally related compounds.[21] Based on a logical retrosynthetic disconnection of 11 and 12, an elegant biogenesis as shown in **Scheme 1** was proposed by Black and coworkers.[22] Starting from a common precursor 7, the first electrocyclization affords cyclooctatriene (8). Conformational isomers of 8 provided 6π-electrocyclization products 9 and 10 that are finally converted into 11 and 12, respectively, by Diels–Alder cycloaddition. Nicolaou *et al.*[23] proved that this route is chemically feasible. Since all endiandric acids were isolated as racemates, their formation would proceed in a nonenzymatic manner. In the biosynthesis of endiandric acids, the reactive polyene precursor is released from the active site of the corresponding polyene formation enzyme (possibly polyketide synthase or dehydrogenase), and is spontaneously cyclized without the assistance of an enzyme to give a series of structurally related cycloadducts.

Marine alkaloids sceptrin (14) and ageriferin (15) from the Caribbean sponge *Agelas conifera* are regarded as [2 + 2]- and [4 + 2] adducts of monomer hymenidin (13), respectively.[24–26] Since the monomer is achiral, the isolation of these alkaloids as optically active forms indicates their formation through an enzymatic Diels–Alder cycloaddition. Based on the abundance of 14 more than 15 in the sponge extracts, an alternative pathway through stepwise ionic rearrangement was proposed as shown in **Scheme 2**. Along this line, a recent biomimetic total synthesis[27] showed that conversion of 14 into 15 proceeds when heated at 195 °C for 1 min in H_2O (microwave). Considering the rapid exchange of deuterium at C2 of imidazole (in CD_3OD, 80 °C, 5 min), they concluded that this rearrangement proceeded by an ionic tandem shift. Thus, co-occurrence of monomer and formal [4 + 2]-adduct is not a conclusive evidence of the involvement of Diels-Alderase.

Bisanthraquinone (−)-flavoskyrin[28] (18a) is originally proposed to be biosynthesized through an enzymatic hetero-Diels–Alder reaction of monomer 16a (**Scheme 3**). In the total synthesis[29] inspired by the biogenesis, oxidation of 16b with MnO_2 gave dimeric adduct 18b as a single isomer, suggesting that regio- and stereoselectivity are controlled by substrate and reaction conditions. When dihydroquinone (16c)
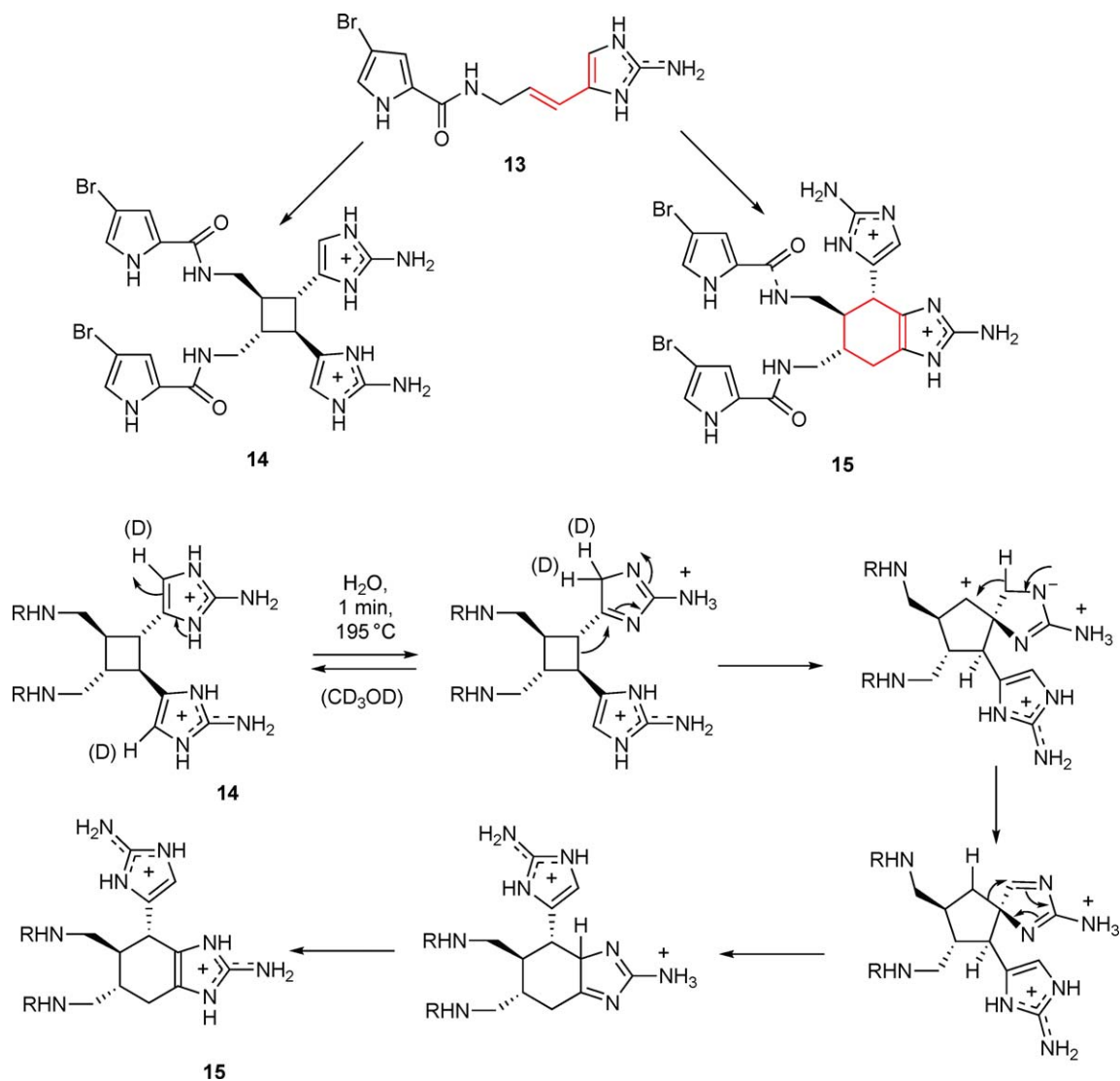
Scheme 2

with bulky protecting group was treated under the same conditions, the reaction slowed down to give unstable intermediate **17c**, which has only a single bond between two quinones and was quantitatively converted into **18c** in CDCl₃. The flavoskyrin analog **18b** was further oxidatively converted into rugulosin analog **19b** possibly via a similar mechanism. A careful examination of these transformations strongly suggested that biosynthetic dimerization is nothing but stepwise Michael reactions.

In addition to the examples shown above, there are alternative pathways (**Scheme 4**) to construct cyclohexene rings by a cationic (mechanism A), a nucleophilic (mechanism B), and a radical mechanism (mechanism C). Therefore, it is difficult to distinguish the Diels–Alder reaction and alternative reactions in cyclohexene formation, and to determine enzymatic and nonenzymatic [4 + 2] adducts. Through our studies on Diels-Alderases,[8,20] we recognized that 'Diels-Alderase' catalyzes not only the formation of reactive species but also cycloaddition at the same active site. This indicates that Diels-Alderase is, at least part of them, a producer of a reactive substrate for cycloadditions. If this is general, a Diels-Alderase could be any type of enzyme such as an oxidase, a dehydrogenase, a decarboxylase, or
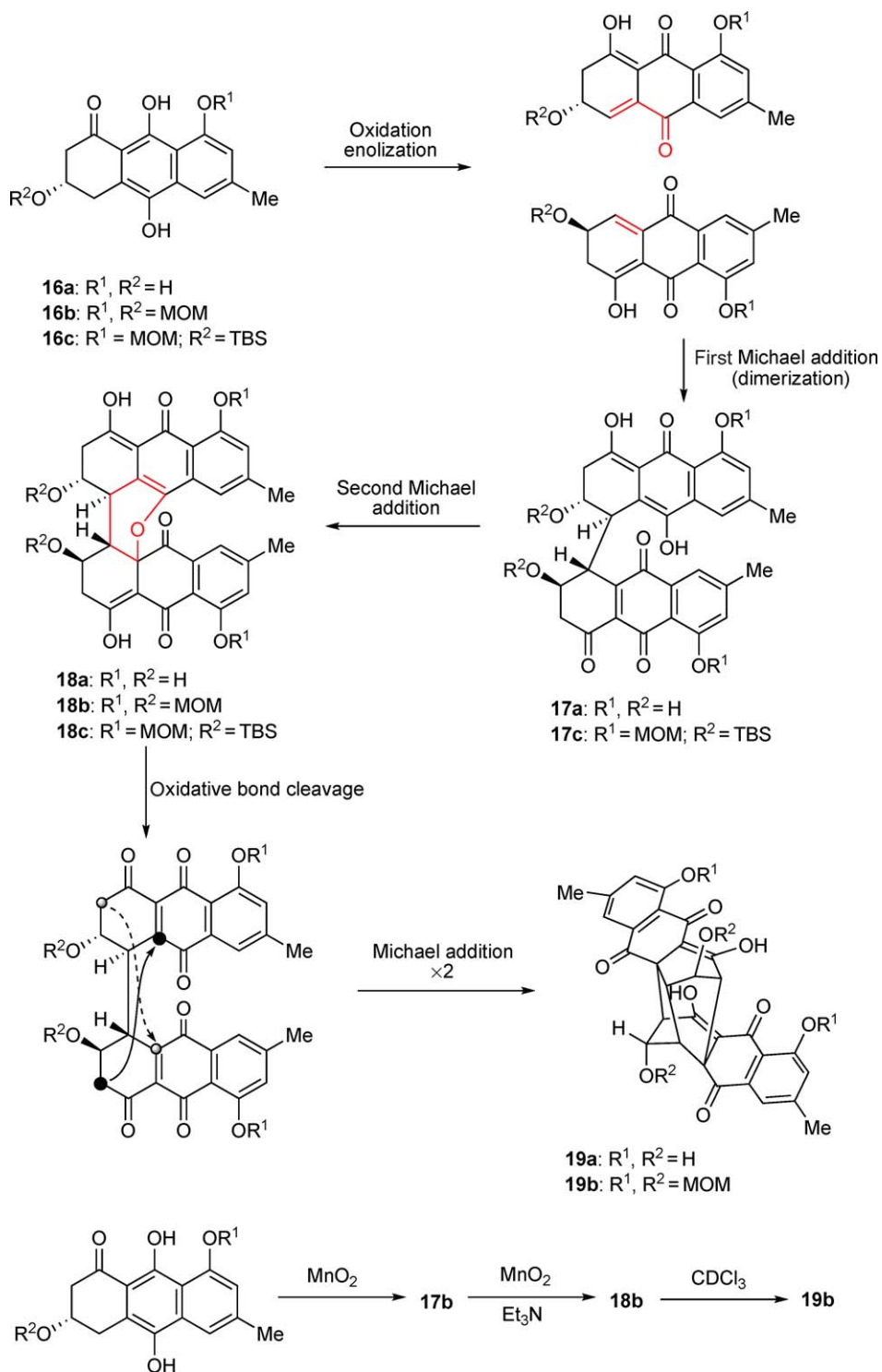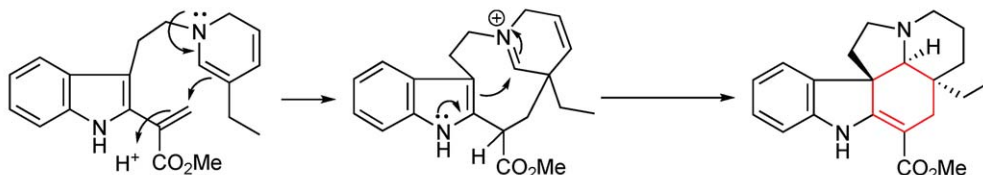
Scheme 3

a polyketide synthase (PKS), which have an active site for the Diels–Alder reaction. The function and catalytic mechanism of natural Diels-Alderases are of great interest due to the diversity of molecular skeletons in natural Diels–Alder adducts. When cycloaddition proceeds after releasing the reactive substrate from the
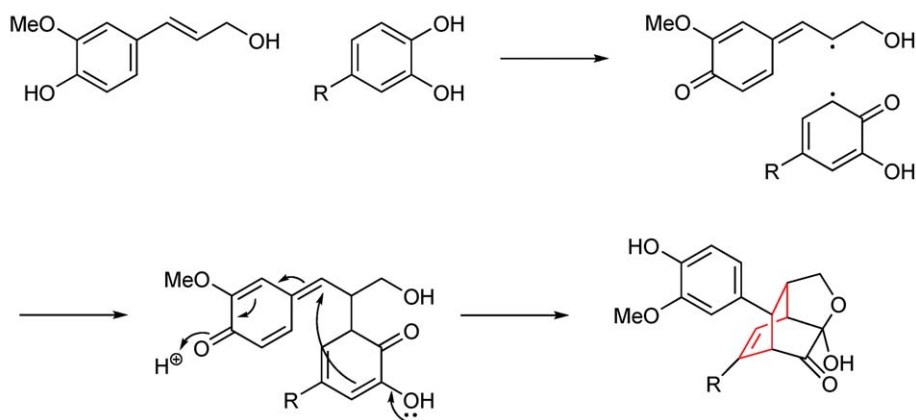
Mechanism A
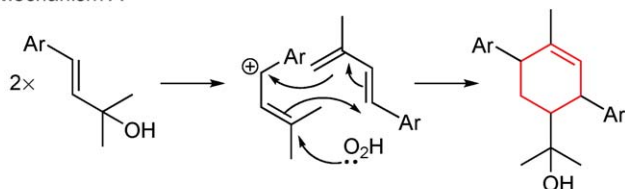
Mechanism B

Mechanism C

Scheme 4

active site, $[4+2]$ adducts in nonenzymatic reaction could be obtained as either chiral or racemic forms. Obviously, these enzymes simply producing reactive substrate are not Diels-Alderases. In view of the formation of natural $[4+2]$ adducts, however, both type of enzymes are important. The enzymes simply producing reactive diene and dienophile and their mechanisms on cycloadditions are also dealt with in this chapter.

8.08.2   Enzymatic Formation of Diels–Alder Adducts

In many cases, it is difficult to predict that a target molecule is actually biosynthesized by a Diels-Alderase based only on its structure. Demonstration of the skeletal construction of natural products with a biomimetic Diels–Alder reaction is an effective way to confirm the involvement of Diels–Alder reactions in their biosynthesis. This section describes putative $[4+2]$ adducts enzymatically biosynthesized in addition to nonenzymatic $[4+2]$ adducts. The biomimetic synthesis not only provides information on the feasibility of the Diels–Alder reaction but also provides diastereomer ratio and detailed reaction conditions to afford adducts.

8.08.2.1   Involvement of Oxidation Enzyme in the Formation of Natural [4 + 2] Adducts

There are numerous examples in which enzymatic oxidation might produce a reactive species in Diels–Alder reactions. Oxidations include dehydrogenation giving 1,3-diene, and phenol oxidation affording cyclic dienone, which is acted as a diene or a dienophile.

8.08.2.1.1   Plausible nonenzymatic [4 + 2] cycloaddition after oxidative transformation of the substrate

8.08.2.1.1(i)   Asatone and related metabolites   Dimeric [4 + 2] adducts are frequently found in natural products. Retrosynthetic analysis of such a compound provides information on the dimerization reaction. The dimer of neolignan, asatone (**22**), was isolated from the plant *Asarum teitonense* (**Scheme 5**).[30] Later, two closely related novel trimers, heterotropatrione (**23a**) and isoheterotropatrione (**23b**), were also isolated.[31] Based on the oligomeric structure of neolignans, biosynthetic pathways of these metabolites were proposed as shown in **Scheme 5**. Oxidation of a phenol **20**, and the subsequent addition of methanol produced a dienone **21** that dimerizes to give **22** through a Diels–Alder reaction. Further cycloadditions of **22** with **21** yield **23a** and **23b**, respectively. Having no optical activity, all these lignans would be formed by spontaneous cycloadditions of the dienone **21** in the absence of an enzyme after enzymatic oxidation of **20**. This proposal was supported by the result that the anodic oxidation of phenol **20** produced reactive quinone methide that underwent cycloaddition to give **22** quantitatively.[32] Later, an example is introduced where an oxidase catalyzes not only an oxidation reaction but also a Diels–Alder reaction in the same active site of solanapyrone synthase (SPS) (Section 8.08.4.1).

8.08.2.1.1(ii)   Diarylheptanoid anigorufone   Bazan *et al.*[33] proposed the involvement of a Diels–Alder reaction in the biosynthesis of phenylphenalenone lachanthocarpone (**27**) based on the conversion of diarylheptanoid (**24**) into **27**, with NaIO$_4$, through the orthoquinone **26** (**Scheme 6**). To examine the intermediacy of the diarylheptanoid in the biosynthesis of anigorufone (**28**), Steiner and coworkers fed the ^{13}C-labeled diarylhepta-noid (**25**) to the cultured root of *Anigozanthos preissii*.[34] The anigorufone (**28**) isolated showed significant incorporation establishing the involvement of the oxidation of **25** into **26**, followed by a Diels–Alder
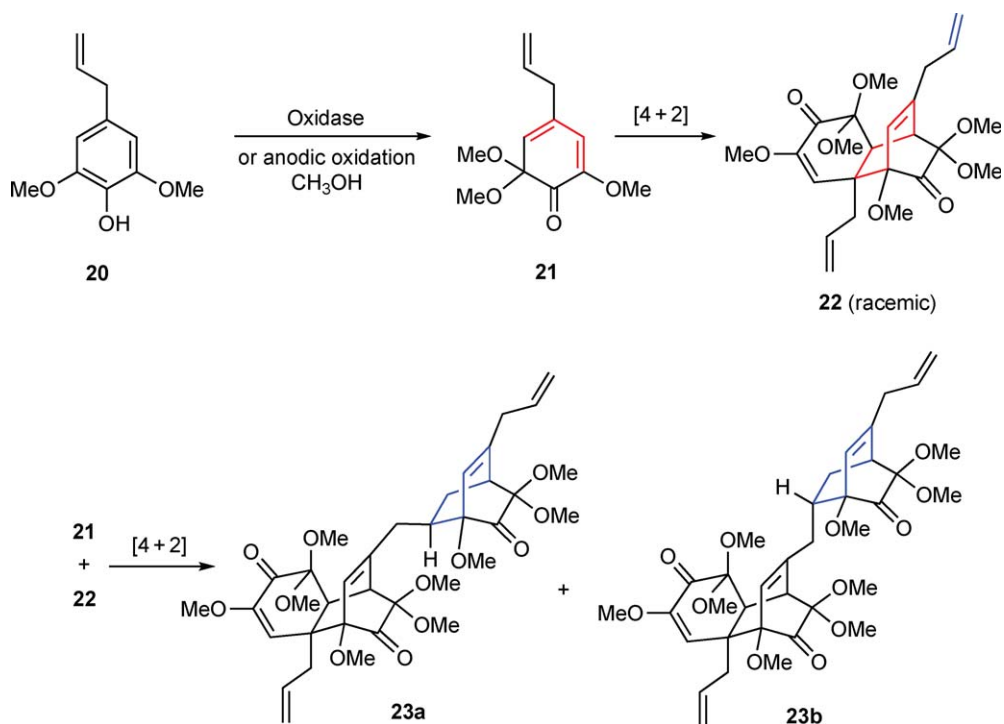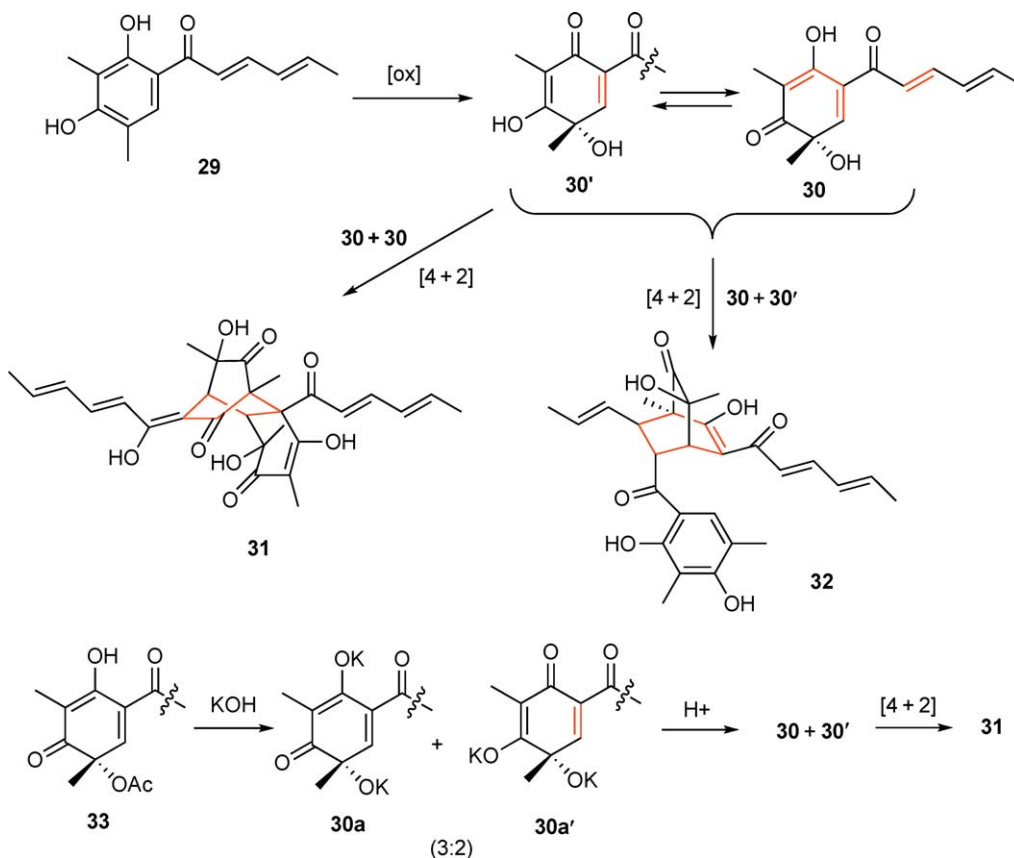

Scheme 5
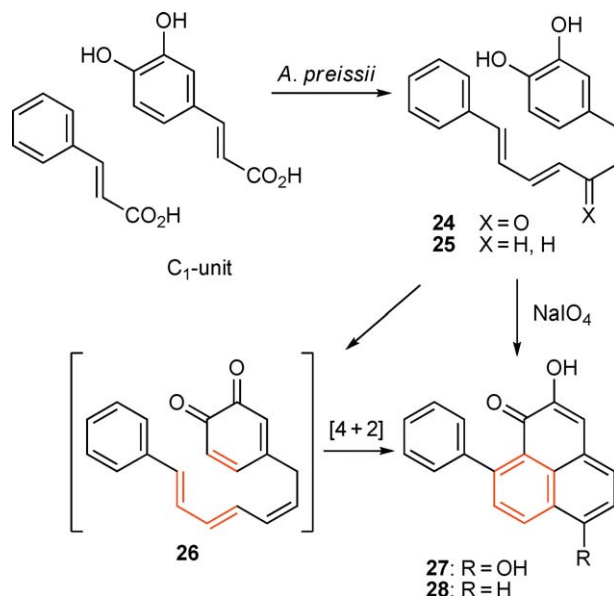
Scheme 6

reaction as proposed by Bazan *et al.* As in the case of asatone (**22**), an oxidase provided the reactive precursor **26** for a cycloaddition. Then, the cycloaddition proceeded after releasing it from the active site of the oxidase to give racemic products. Again, involvement of a Diels-Alderase is not essential in this case.

8.08.2.1.1(iii) Bisorbicillinol and aquaticol   Next, examples in which a chirality of a product is not conclusive evidence for distinguishing nonenzymatic and enzymatic reactions are introduced. The bisorbicillinoids are a family of structurally diverse fungal metabolites represented by bisorbicillinol (**31**), which shows DPPH radical scavenging activity.[35,36] Based on extensive studies of the structure elucidation and biosynthesis of the bisorbicillinoids, Abe *et al.*[37] proposed that the stable monomer sorbicillin (**29**) enantioselectively oxidized to the reactive sorbicillinols (**30**, **30′**), which dimerizes through two different [4 + 2] cycloadditions to provide bisorbicillinol (**31**) and sorbiquinol (**32**) as shown in **Scheme 7**. During the purification of sorbicillinol (**30**), it was found that the concentration of a solution of **30** caused a [4 + 2] cycloaddition to give **32**, indicating that **30** is highly reactive and that the conversion is nonenzymatic and occurs under mild conditions with complete regio- and stereoselectivity.[37,38] In synthetic studies of the bisorbicillinoids,[36] basic hydrolysis of acetate **33** gave two discrete quinolates (bis-deprotonated forms of **30a** and **30a′**) that underwent cycloaddition after subsequent acidification. Involvement of a Diels-Alderase is not necessary in this case because a nonenzymatic reaction provided a single product and because enantioselective oxidation of **29** introduced the chirality in **30**, thus determining the stereochemistry in **31**.

The detection of a significant amount of monomer **29** in the bisorbicillinol-producing fungus indicated that an oxidase provided the chiral reactive substrate **30** and that the Diels–Alder reaction of **30** was promoted in the aqueous medium without a Diels-Alderase.

Similar enzyme-catalyzed phenolic oxidation could produce a [4 + 2] adduct in the biosynthesis of monoterpene aquaticol[39] (**36**). Recent total synthesis of aquaticol (**36**)[40] showed that a mixture of orthoquinols (**35a**) and **35b** prepared by the oxidation of chiral monomer **34** with SIBX provided only two diastereomers **36** and **37** out of four expected products as shown in **Scheme 8**. This indicated that chiral orthoquinol (**35a**) gave a single diastereomer **36**. On the basis of the computational results, the energy difference between the transition states A (**36**-like) and B (**37**-like) is more than 9.9 kcal mol^{-1}, which was explained by the hyperconjugative effects. In this case, again, reactivity and remarkable diastereoselectivity originates from the substrate. Examples of bisorbicillinoids and aquaticol suggest that the [4 + 2] adduct was provided from a reactive hydroxydienone, which was derived oxidatively from an aromatic ring is most likely the nonenzymatic reaction products.
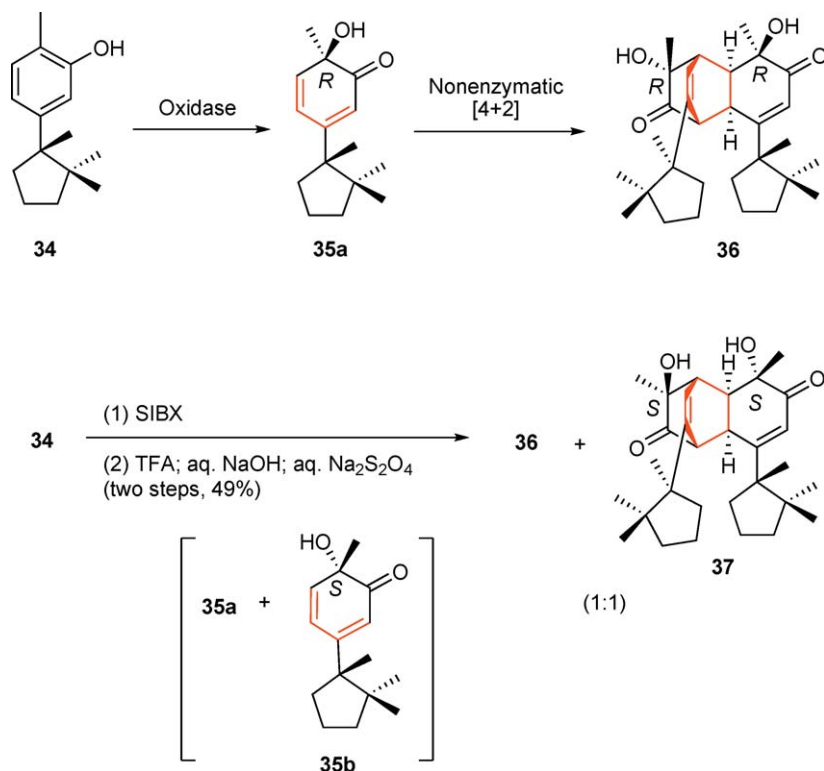
Scheme 7

8.08.2.1.1(iv) Torreyanic acid

This is one of the excellent examples where biomimetic synthesis explains an unusual event in biosynthesis. Porco *et al.*[41] extensively studied the synthesis of the fungal metabolite torreyanic acid (**38**).[42] Isolation of the corresponding monomer (+)-ambuic acid (**40**)[43] suggested the involvement of a Diels–Alder reaction in the biosynthesis of **38**. In the biomimetic synthesis[41] of **38**, oxidation of alcohol **40** with Dess–Martin periodinane provided aldehyde **41**, which rapidly converted through oxaelectrocyclization into *syn*- and *anti*pyrans **42a** and **42b**, respectively (**Scheme 9**). Although aldehyde **41** and 2*H*-pyrans **42a** and **42b** existed as an equilibrium mixture, both steric and substituent effects shifted the equilibrium to the formation of the 2*H*-pyrans over aldehyde **41**. Spontaneous Diels–Alder dimerization of pyrans **42a** and **42b** proceeded with complete regio- and diastereos-electivity to give *endo*-adduct **38a**. In the retro-Diels–Alder reaction of **42a** at 60 °C, signals originating from pyrans **42a** and **42b** were detected in the ^{1}H-NMR spectrum, but no aldehyde signal was observed. Theoretical calculations indicate that dienal formation is a disfavored process (11–12 kcal mol^{-1}) in the electrocyclization and that a remarkable difference exists between the energies of the transition states in the Diels–Alder reaction: the most favored one, **39**, is 9.4 kcal mol^{-1} more stable than the alternative. These calculations and the easily accomplished reaction to reverse the reaction explain the exclusive formation of a single diastereomer of **38a**. The high-reactivity values of the substrates **42a** and **42b**, which may be produced by a corresponding dehydrogenase, and the excellent diastereoselectivity in the cycloaddition indicate that the corresponding Diels–Alder reaction proceeds in a nonenzymatic manner.

8.08.2.1.1(v) Grandione

Grandione[44] (**44**) from *Torreya grandis* is a dimer of an *endo*-selective hetero-Diels–Alder adduct between the *o*-quinone moieties present in the modified abietane diterpene monomer **43**. Biomimetic oxidation[45] of demethylsalvicanol (**43**) in solid state at room temperature afforded a single [4 + 2]
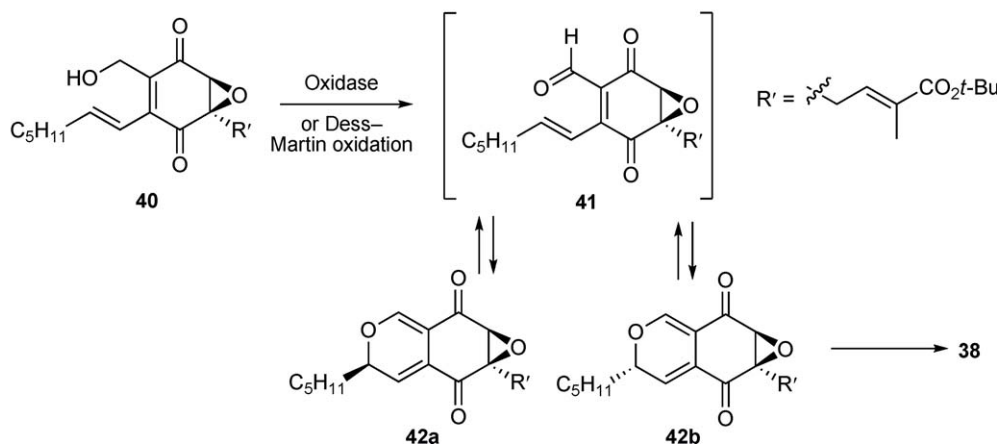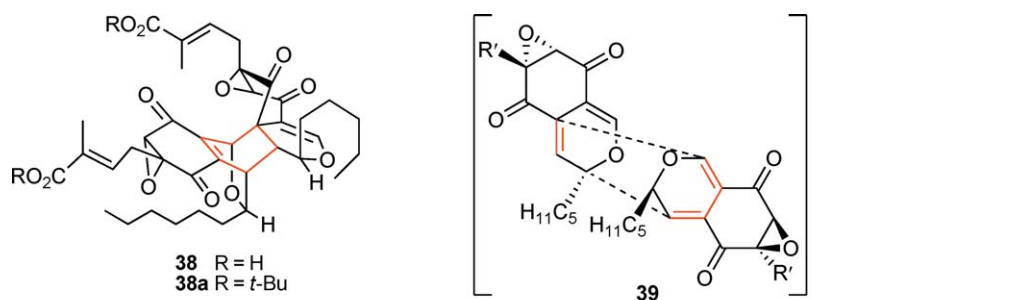
Scheme 8

adduct that was identical to **44** (**Scheme 10**). On the basis of X-ray crystallographic analysis of the synthetic sample, it is necessary to revise the original structure of grandione (**44′**) to **44**. In the biosynthesis of **44**, enzymatic oxidation converts catechol moiety of **44** enantioselectively to reactive *o*-quinone, which could afford the [4 + 2] adduct nonenzymatically.

Examples shown in this section indicate that a significant number of natural [4 + 2] adducts are produced in a nonenzymatic manner and that chirality of the adducts is not a conclusive evidence for the involvement of Diels-Alderases.
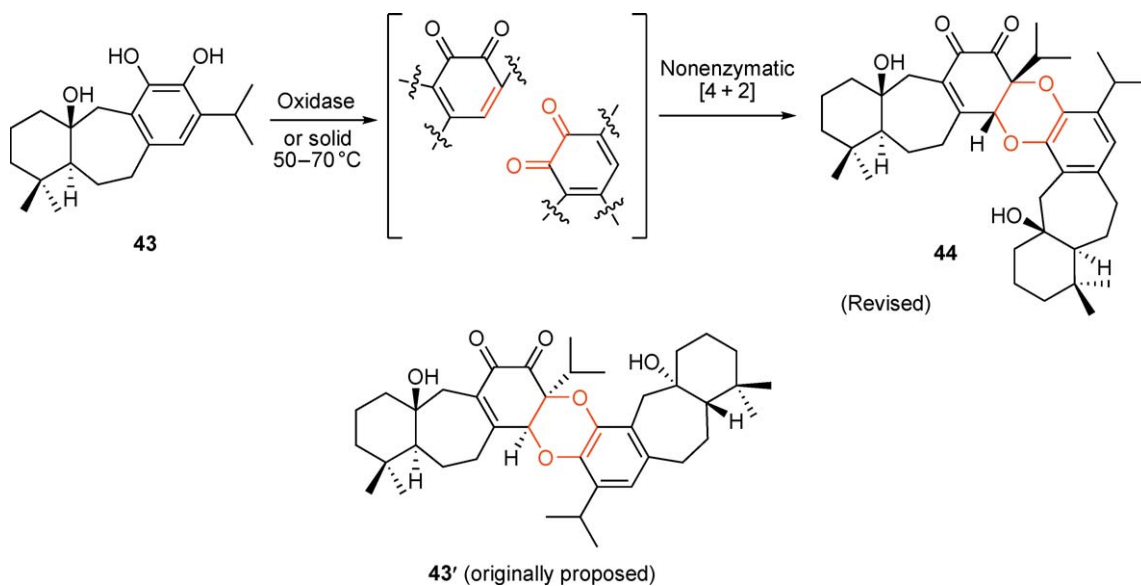
8.08.2.1.2 Plausible enzymatic [4 + 2] cycloaddition after oxidative transformation of the substrate

8.08.2.1.2(i) Pinnatal Biosynthesis of antimalarial naphthoquinone pinnatal derivatives[46,47] involves series of pericyclic reactions.[48] Starting from geranylated naphthoquinone **46**, oxidation and the subsequent dehydration gave reactive achiral orthoquinone **47** that would be converted into chiral adduct pyranokunthone A (**49**) as shown in **Scheme 11**. On the other hand, 6π-electrocyclization of the alternative isomeric orthoquinone **48** afforded pyranokunthone B (**50**). In these oxidations from **47** and **48** to **49** and **50**, respectively, achiral precursors gave optically active products, indicating the involvement of Diels-Alderases, which probably catalyze both oxidation and cycloaddition. Actually, biomimetic syntheses[48] of **49** and **50** have been achieved starting from commercially available naphthoquinone **45** and geranial. In the pinnatal biosynthesis, another oxidase that converts **50** into aldehyde **51** could catalyze the hetero-Diels–Alder reaction to give pinnatal (**52**).
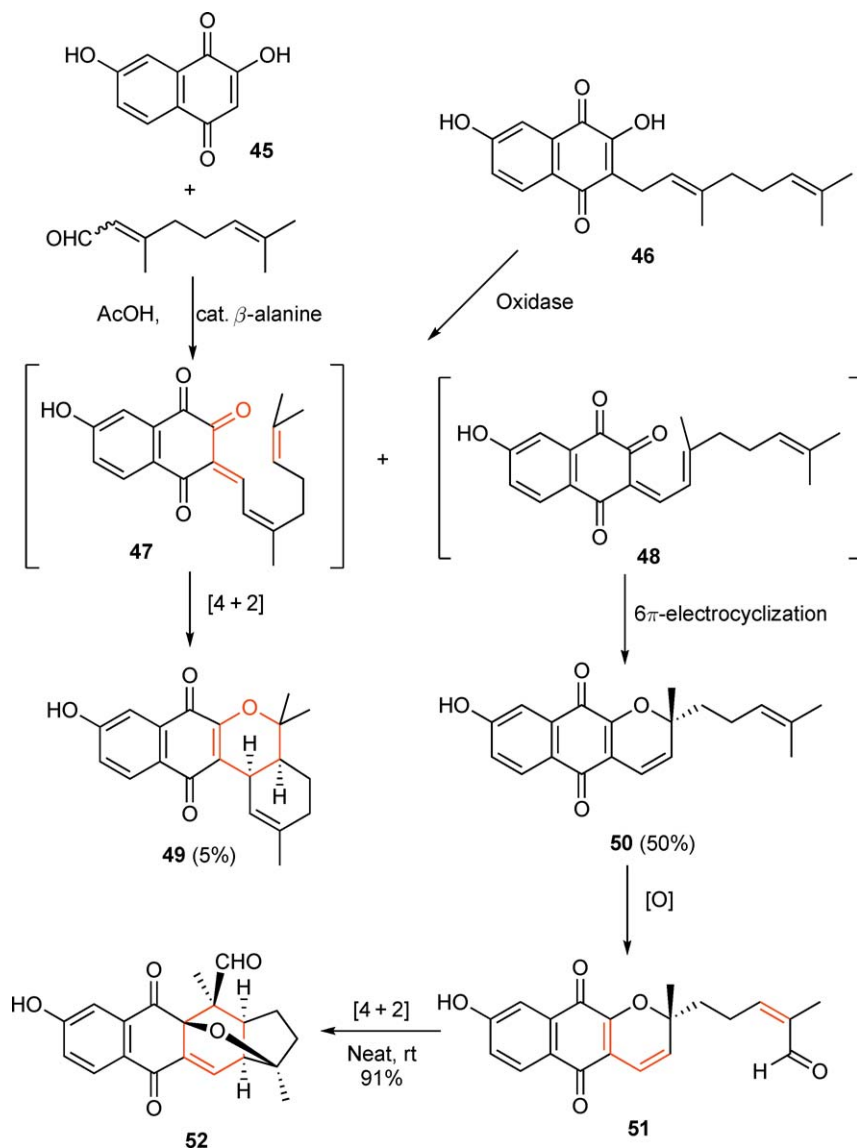
8.08.2.1.2(ii) Perovskone A Diels–Alder reaction between different types of terpene units is plausible in several cases. Perovskone[49] (**55**) from the heartwood of *Chamaecyparis obtuse* is regarded as an adduct between a chiral quinone **53** derived from an abietane-type diterpene (similar to that of grandione) and the linear monoterpene *trans*-β-ocimene as shown in **Scheme 12**. In the total synthesis[50] of perovskone, the Diels–Alder reaction of **53** and *trans*-α-ocimene with Lewis acid Eu(fod)₃ proceeded in good regio- and

38 R = H
38a R = *t*-Bu

39

40 → Oxidase or Dess–Martin oxidation → **41**

R′ = CO₂*t*-Bu

42a **42b** → **38**

Scheme 9

43 → Oxidase or solid 50–70 °C → Nonenzymatic [4 + 2] → **44** (Revised)

43′ (originally proposed)

Scheme 10

Scheme 11

diastereoselectivity to afford a desired diastereomer **56** as a major product. Since this biomimetic [4 + 2] cycloaddition accompanied the formation of unnatural *endo*-adduct, Diels-Alderase, which could catalyze phenol oxidation to quinone **53**, is responsible for the construction of the perovskone skeleton.

8.08.2.1.2(iii)   Kuwanons   Kuwanons **57a, 57b, 58a,** and **58b**, which are constituents of plants belonging to the family Moraceae, are phytoalexins that consist of pairs of diastereomers that are derived from chalcones and stilbenes.[51] Based on their structures, kuwanons I (**57a**) and J (**57b**) are regarded as dimeric adducts of a precursor chalcone with a diene derived from the prenyl side chain. Isolation of these diastereomers as optically active forms strongly indicated that the achiral precursors afford *endo*- and *exo*-adducts through an enzymatic Diels–Alder reaction.[51] Chemical feasibility of the corresponding [4 + 2] cycloaddition[52] was confirmed by the fact that reaction of two different monomers **59** and **60** gave *exo*- and *endo*-adducts **61a** and **61b**, respectively (**Scheme 13**). Nomura and Hano[51] reported a series of
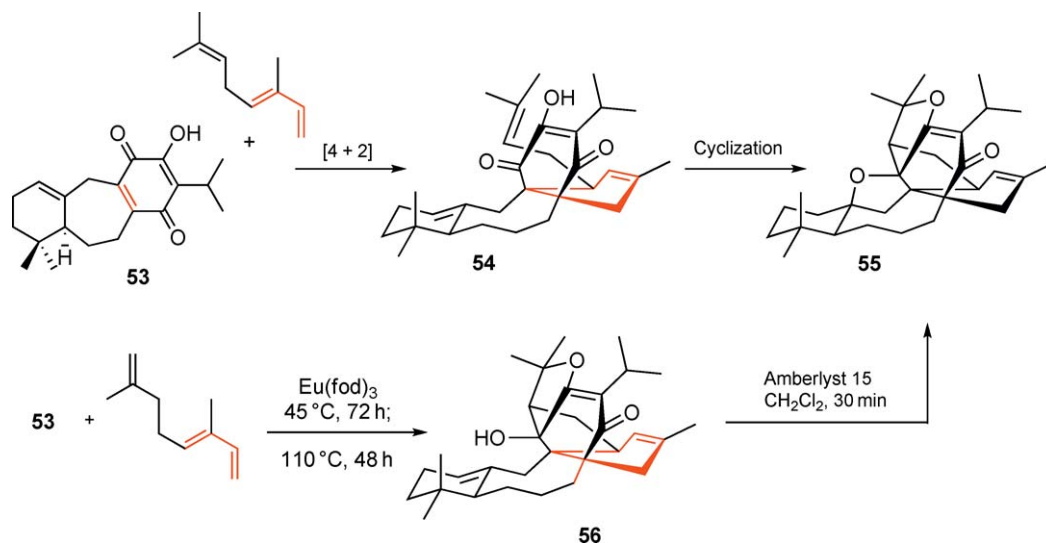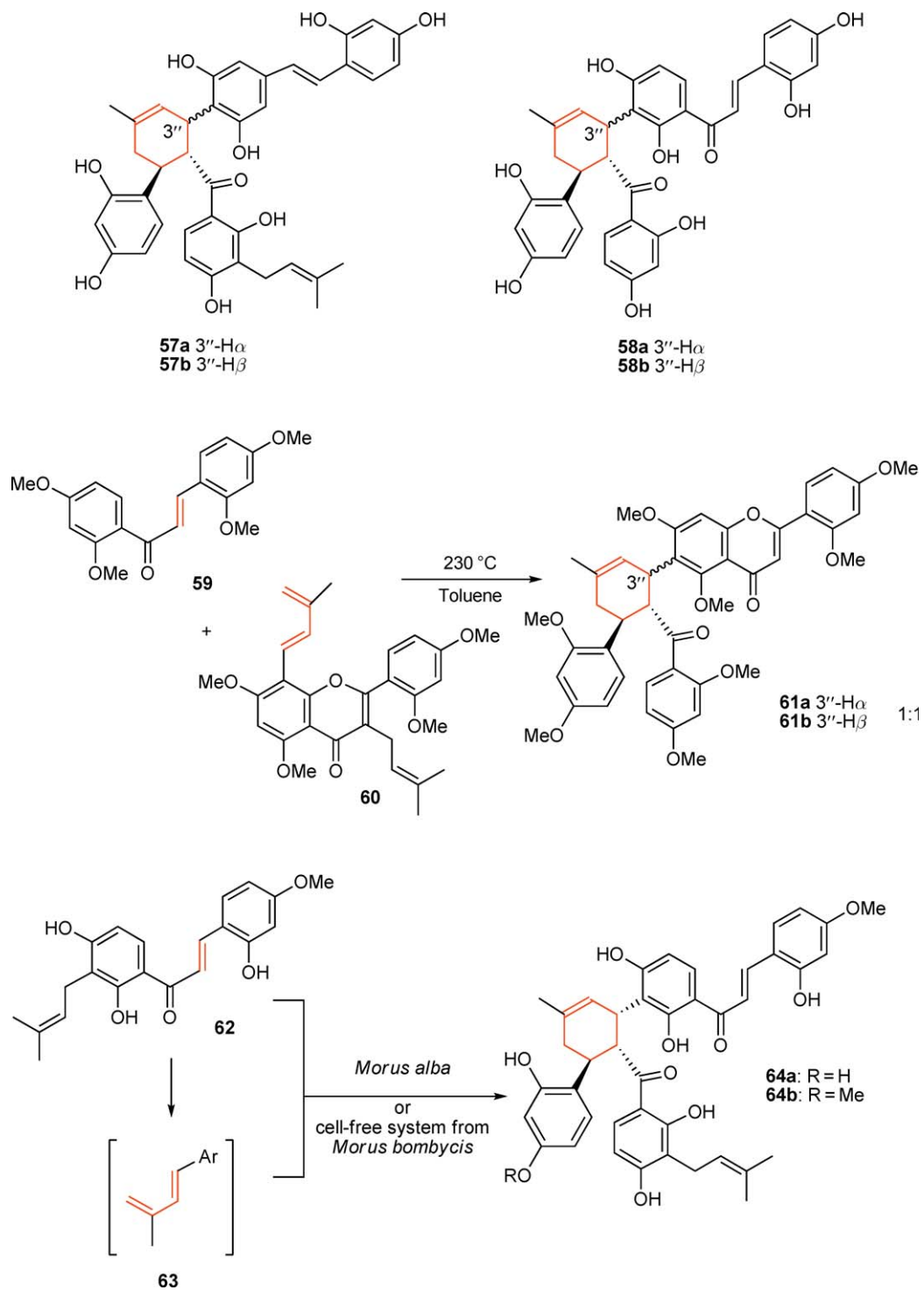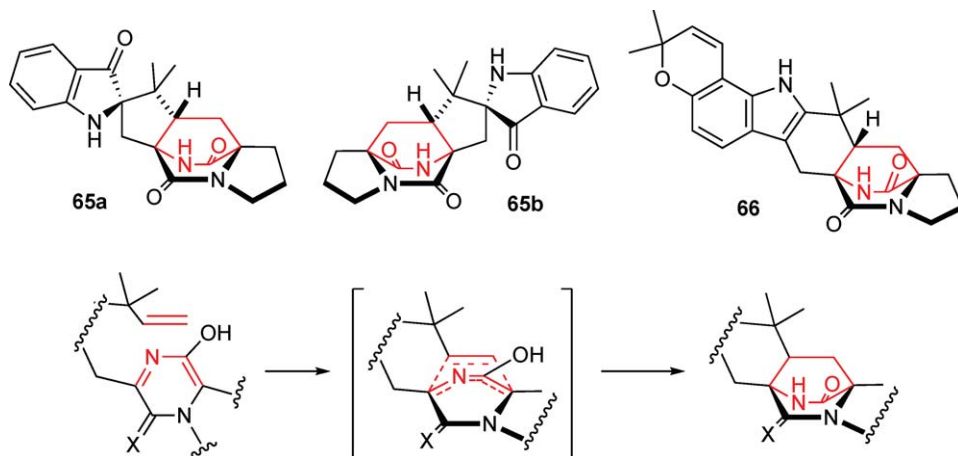
Scheme 12

incorporation experiments to examine this hypothesis. In a feeding experiment with the nonnatural methoxychalcone **62** into the callus tissue of *Morus alba*, the dimeric adducts **64a** and **64b** were obtained, indicating that dehydrogenation of the prenyl group followed by [4 + 2] cycloaddition between **62** and **63** yielded the nonnatural adduct **64b** (**Scheme 13**).[53] In the case of metabolites described in Section 8.08.2.1.1, an oxidative enzyme may convert a precursor into a reactive diene or dienophile and release it from the active site. The subsequent Diels–Alder reaction gave racemic products. On the other hand, a plausible dehydrogenase involving kuwanon biosynthesis may convert **62** into the reactive diene **63**, but, in this case, the same enzyme may provide an active site for the intermolecular Diels–Alder reaction between **62** and **63** to afford chiral **64b**. Although the cell-free system from *Morus bombycis* showed the catalytic activity of similar transformation observed in cell culture,[54] unfortunately, no characterization of the corresponding enzyme has been reported.

8.08.2.1.2(iv) Brevianamides Brevianamides A (**65a**) and B (**65b**) and stephacidin A (**66**) constitute a structurally unique family of fungal metabolites containing the bicyclo[2.2.2]diazaoctane skeleton, which is proposed to be constructed through an intramolecular Diels–Alder reaction between an azadiene and an isoprene unit (**Figure 2**). To date, more than 30 compounds belonging to this family have been reported in the literature.[55] Extensive synthetic and biosynthetic studies of this family have been carried out by Williams[54] and Williams and Cox,[56] and have recently been reviewed.

It was found that isotopically labeled deoxybrevianamide E (**70**) was efficiently incorporated into **65a** and **65b** when fed to the culture of brevianamide-producing fungus (**Scheme 14**). This established that **70** is an actual intermediate of brevianamides, and that prenylation takes place after the formation of the bicyclo[2.2.2]diazaoctane skeleton.[57] Improved biomimetic total synthesis of **65a** and **65b** was achieved through a hetero-Diels–Alder reaction of **67**[58] (**Scheme 14**). These mild cycloaddition conditions require conversion of enamide into the azadiene obligate combination of PBu_3 and DEAD. This biomimetic synthesis provides a strong support for the assertion that the plausible Diels–Alder reaction in **Scheme 14** is chemically feasible. The oxidative enzyme as shown in Section 8.08.2.1.2 (kuwanons) could be responsible for the formation of the azadiene and the subsequent Diels–Alder reaction. Significant progress has been made in the biosynthesis of the brevianamides and their structural analogs, but identification of the real substrates for the Diels–Alder reaction and elucidation of the mechanism forming the reactive azadiene system remain to be solved.

57a 3″-Hα
57b 3″-Hβ

58a 3″-Hα
58b 3″-Hβ

61a 3″-Hα
61b 3″-Hβ   1:1

230 °C
Toluene

59

+

60

62

63

Morus alba
or
cell-free system from
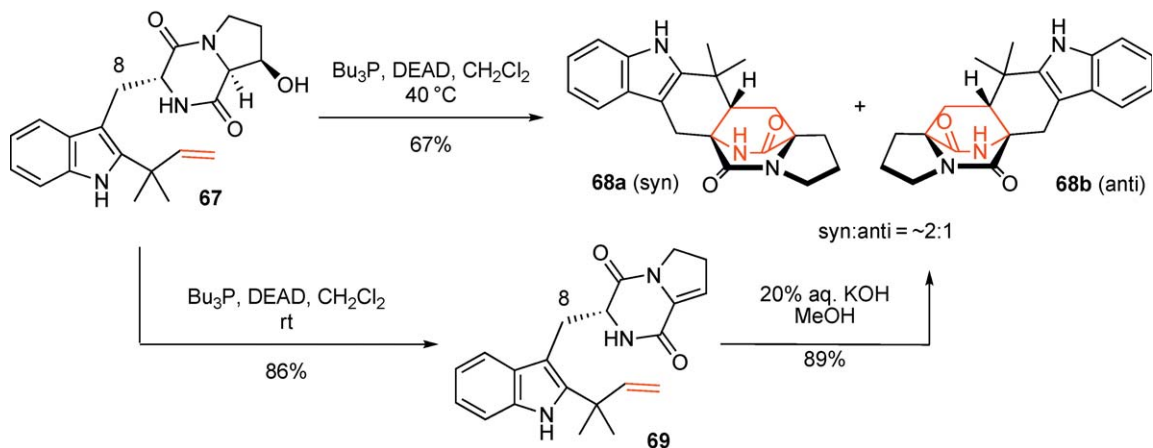*Morus bombycis*

64a: R = H
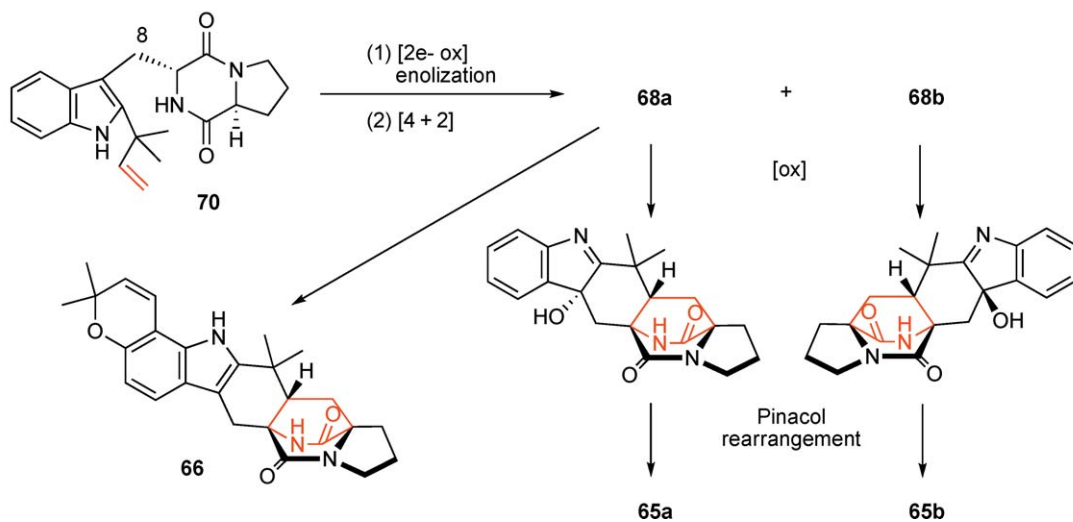**64b**: R = Me

Scheme 13

Figure 2 Structures of breviamide family metabolites.

(i) Biomimetic Diels–Alder reaction

(ii) Biosynthetic pathways of brevianamides

Scheme 14

8.08.2.2   Involvement of Polyketide Synthase in the Formation of Natural [4 + 2] Adducts

8.08.2.2.1   Plausible enzymatic [4 + 2] cycloaddition of putative intermediate produced by polyketide synthase

Polyketide synthase type-I[59] (type-I PKS) can provide linear products such as **71** with *E,E*-conjugated polyene system and *E*-enone system, which are reactive substrates for [4 + 2] cycloadditions (**Figure 3**). PKS–nonribosomal peptide synthethases[59,60] (NRPS) produce similar polyene such as **73** with acyltetramate moiety at the terminal of the polyketide backbone. Intramolecular Diels–Alder reactions of polyene polyketide intermediate provide various carbocyclic systems such as **72 and 74**. Although the length between diene and dienophile on the polyketide chain may be variable, a common carbocycle produced by the biosynthetic Diels–Alder reaction is decalin in the metabolites produced by fungi and actinomycetes.[5–8] Differences in the number of branched methyl groups, oxidation level, and amino acids increase the number of members of this family. The biosynthetic Diels–Alder reaction usually affords a single adduct out of the possible diastereomers, except plant derived adducts. The stereochemistry of the decalin system in the adducts is usually *trans* from *endo*-transition state, while cis-junction from the *exo*-transition state is also found in several cases: Representative examples belonging to this class are microbial metabolites; phomopsidin[61] (**75**, an inhibitor of microtubule assembly); decumbenones[62] (**76**, an inhibitor of fungal melanization); tubelactomicin[63] (**77**, an antimicrobial agent); UCS1025A[64] (**78**, an antitumor antibiotic); phomasetin[65] (**79**); TAN-1813[66] (**80**, a novel *ras*-farnesyltransferase inhibitor); and integramycin[67] (**81**, an inhibitor of HIV-1 integrase) as shown in **Figure 3**. Among them, *endo*-adducts are predominant, but there are several *exo*-adducts such as **75** and **81**. Several plant origin metabolites were also reported as described in the next section.

A number of total syntheses of decalin polyketides have been reported,[3] supporting the feasibility of biosynthetic cycloadditions. Actual involvement of PKS (lovastatin PKS and chlorothricin PKS) and PKS–NRPS (cheaoglobosin PKS–NRPS and equisetin PKS–NRPS) as Diels-Alderases is discussed in Section 8.08.4. Intramolecular Diels–Alder reaction could occur either during or after chain elongation of polyketide chain.

8.08.2.2.2   Plausible nonenzymatic [4 + 2] cycloaddition of putative intermediate produced by polyketide synthase

Piperacea alkaloids piperstachine (**82**), cyclostachines A (**83a**), and B (**83b**) were isolated from *Piper trichostachyon*.[68] Biogenetically, it is regarded that an intramolecular Diels–Alder reaction of linear precursor **82** would be expected to give *exo*-adduct **83a** and *endo*-adduct **83b** as shown in **Scheme 15**.[69] The molecular skeleton of the linear precursor can be regarded as a polyketide. Although the corresponding PKS affording a reduced polyketide chain has not been reported, PKS similar to type-III PKS chalcone synthase, which utilizes cinnamoyl moiety (C6–C3 unit) as a starter unit, produces **82**. These alkaloids were obtained as racemates, indicating the nonenzymatic formation of **83a and 83b**. In this case, PKS can be regarded as a producer of the reactive substrate. In the nonenzymatic cycloaddition of the substrate of Diels-Alderase SPS (Section 8.08.4.1), remarkably high reactivity in aqueous medium was observed. Thus, nonenzymatic cyclization of piperstachine might proceed to give adducts during prolonged reaction time *in vivo*.

Structurally related [4 + 2] adducts, brombyins II (**85a**) and III (**85b**), were obtained as racemic forms from the bark of *Brombya platynema* F. Muell (Rutaceae).[70] Co-occurrence of a 1-piperonyldodecane intermediate analog suggests that the putative triene precursor **84** similar to piperstachine (**82**) could subsequently undergo two possible Diels–Alder cyclizations to yield two racemic products **85a and 85b** in approximately equal quantities (**Scheme 15**).[71]

8.08.2.2.3   Plausible enzymatic imine formation preceding [4 + 2] cycloaddition

In the field of marine natural products, there are a number of adducts that may be produced by an intramolecular [4 + 2] cycloaddition of appropriate dienes and α,β-unsaturated imines. Recently, the biomimetic synthesis of an intramolecular Diels–Alder reaction through iminium ion has been reported. The remarkable reactivity of this intermediate was realized. In the case of [4 + 2] cycloaddition involving iminium ion, formation of imine should
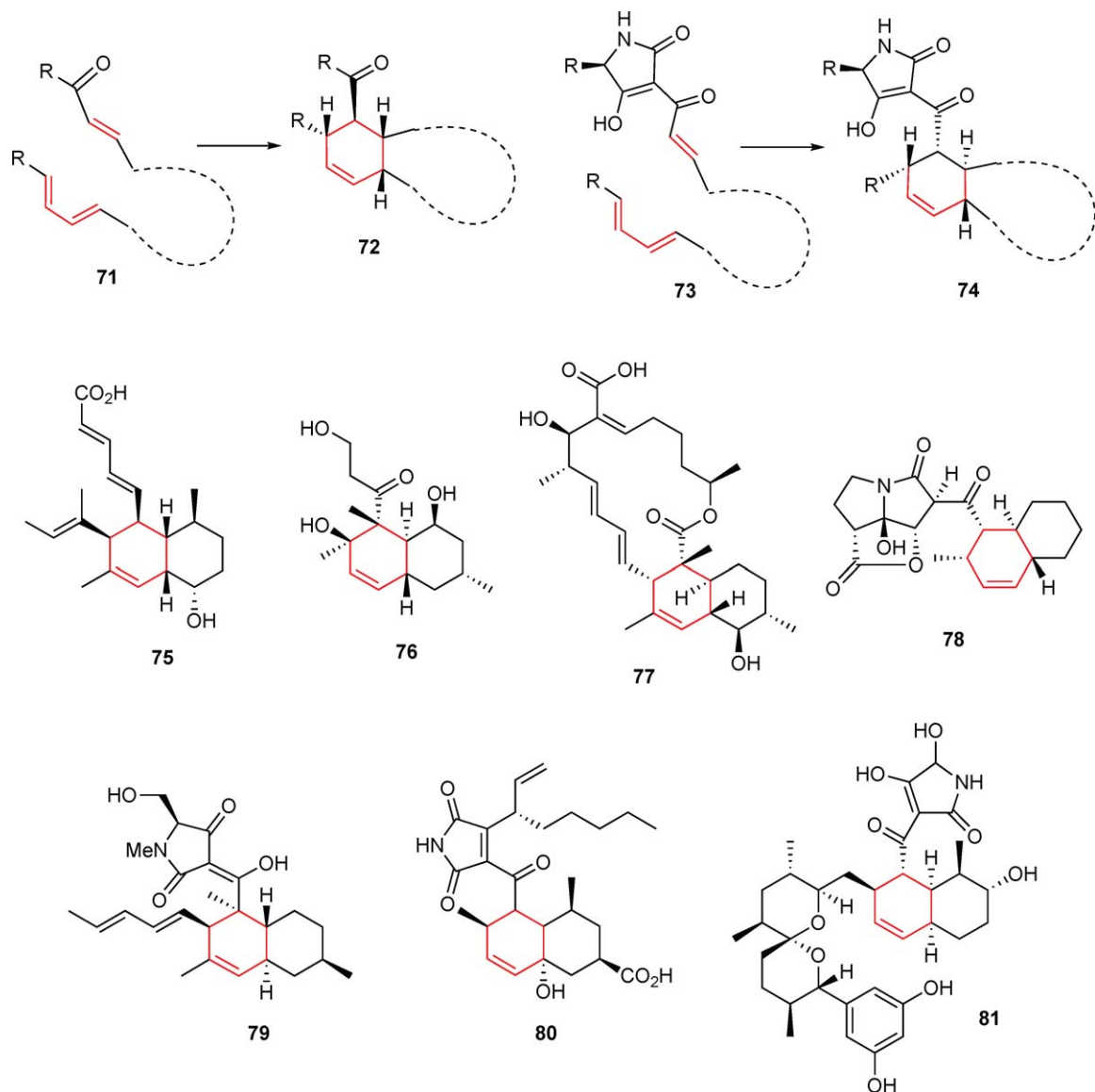
Figure 3 Representative examples of intramolecular [4 + 2] adducts possible catalyzed by PKS or PKS–NRPS.

occur prior to [4 + 2] cycloaddition. If one applies the empirical rule that Diels-Alderases possess a catalytic role, and imine-forming enzyme such as transaminase could be a Diels-Alderase. In this section, several examples are introduced along this line.

8.08.2.2.3(i) Himbacine The new drug leads of Alzheimer's disease, himbacine[72] (**91b**) and related alkaloids were isolated from the bark of *Galbulimima belgraveana*. A unique biogenetic proposal of **91b** has been reported by Baldwin and coworkers.[73] Tetraene butenolide **86** possibly constructed by PKS and reductase is converted into a chiral iminium ion **87a** or **87b** by enzyme-catalyzed reductive amination. The resultant highly reactive iminium ion **87a** or **87b** undergoes spontaneous intramolecular Diels–Alder reaction to afford *endo*-adduct **88a**, which is further reduced to imine **91b**. This elegant proposal was examined by a biomimetic synthesis as shown in **Scheme 16**. Starting from a protected amine **92**, one-pot transformation including *N*-deprotection with TFA, concomitant imine formation and cycloaddition followed by reductive quenching gave a mixture of
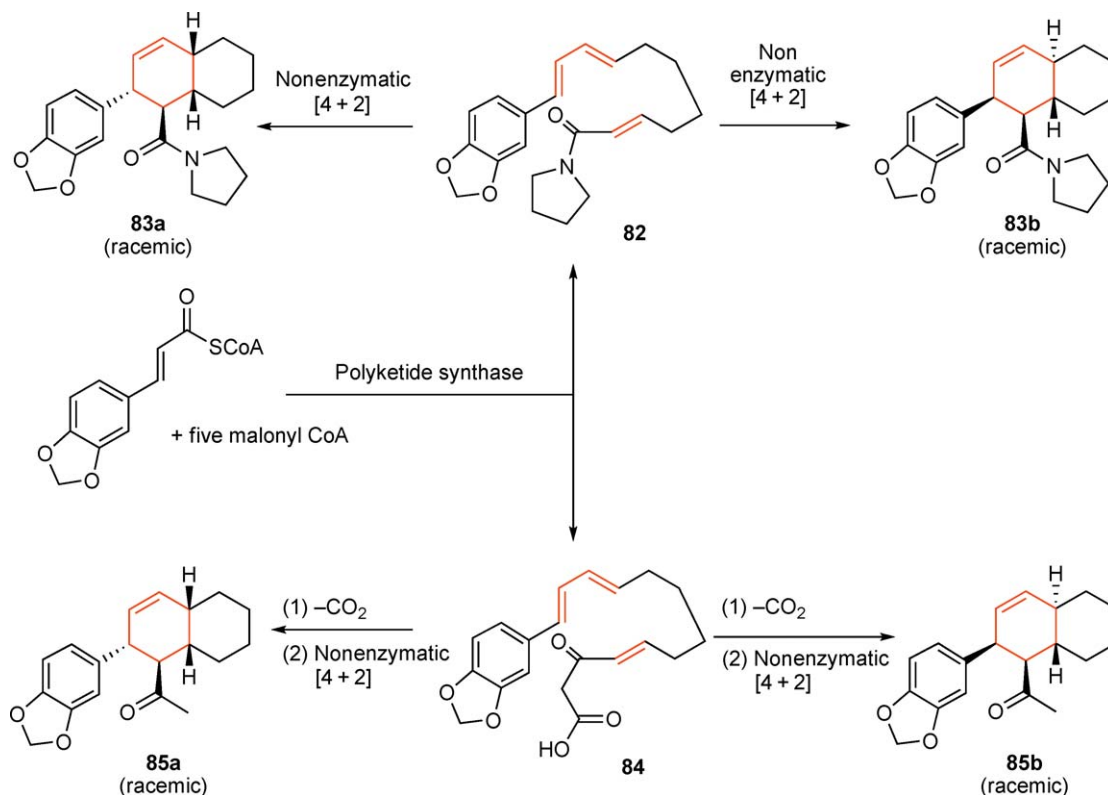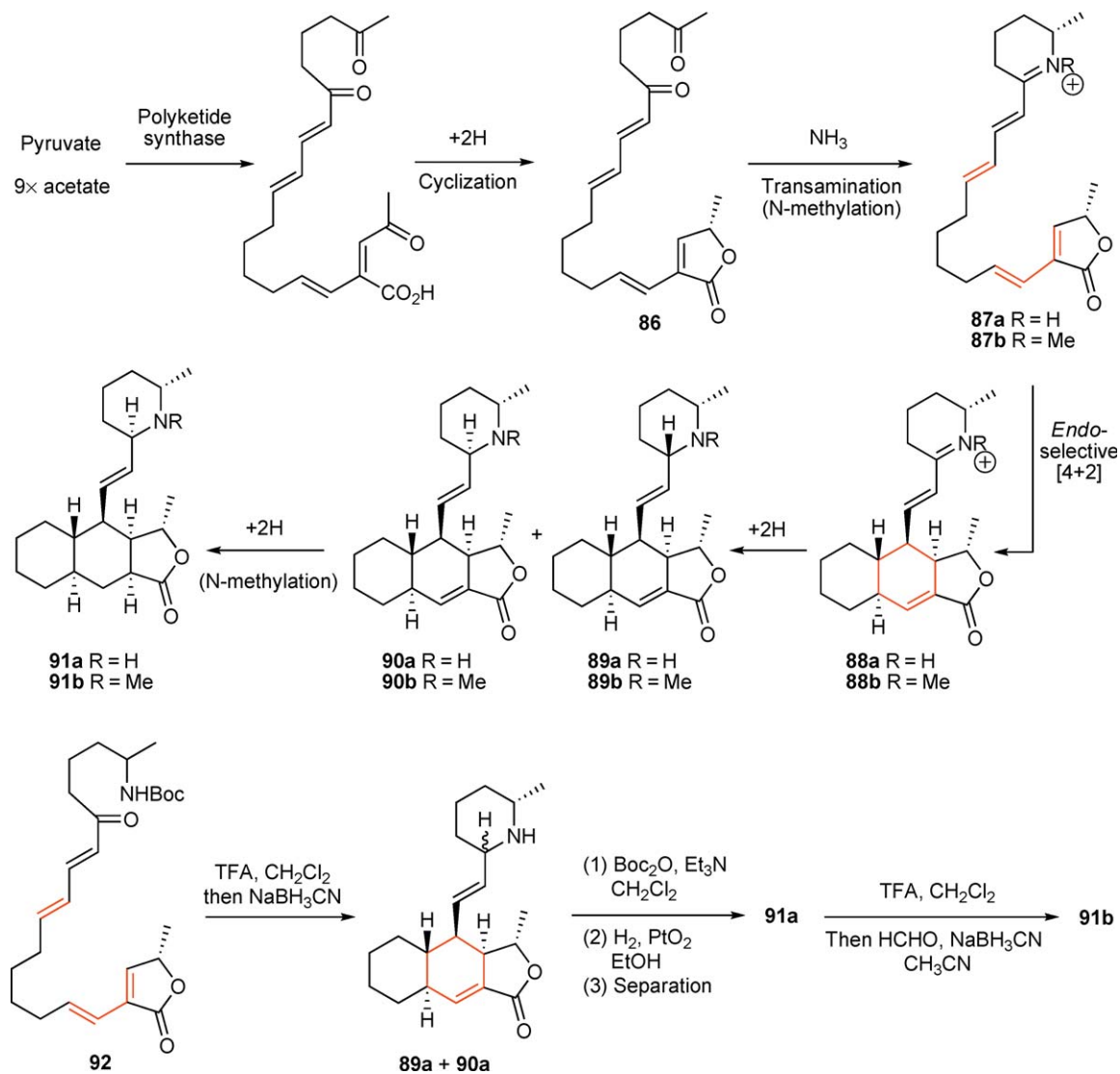
Scheme 15

two diastereomeric himbacine intermediates **89a** and **90a** concerning the piperidine ring. The efficient synthesis of himbacine (**91b**) provided a strong support for the biogenetic proposal. The remarkable enhancement of cycloaddition reactivity with iminium salt indicated that the corresponding transaminase could be a Diels-Alderase. Since a significant number of natural [4 + 2] adducts have an imine or iminium moiety akin to cyclohexene ring, transaminases could be Diels-Alderase.

8.08.2.2.3(ii)   Symbioimine   Iminium ion intermediate is proposed in the biosynthesis of osteoclastogenesis inhibitor, symbioimine[74] (**97**) from *Symbiodinium* sp., a culture from dinoflagellate that has a symbiotic relationship with the acoel flatworm *Amphysicolops* sp. Original biogenetic analysis involves *exo*-selective intramolecular Diels–Alder reaction and subsequent iminium ion formation through **98** as shown in **Scheme 17**. An alternative proposal involving *endo*-selective cycloaddition of dihydropyridinium ion **94** derived from **93** and the subsequent epimerization of **95** was examined by biomimetic synthesis.[75] Observation that the [4 + 2] cycloaddition of dihydropyridinium ion derived from **99** gave a single diastereomer **96b** strongly supports the involvement of Diels-Alderase, very likely a transaminase, which provides reactive dihydropyridinium ion **94**.

8.08.2.2.3(iii)   Manzamines   The manzamines belong to a family of unique heterocyclic alkaloids isolated from marine sponges. Manzamines A (**100**) and B (**101**) were isolated from *Haliclona* sp.,[76,77] and later, many manzamine-related alkaloids were isolated from other sponges. In 1992, plausible biogenetic precursors ircinals A (**102**) and B (**103**) from *Ircinia* sp. were reported.[78] At the same time, Baldwin and Whitehead proposed an elegant biogenetic pathway involving a Diels–Alder reaction.[79]

In Baldwin and Whitehead's pathway (**Scheme 18**), the intramolecular Diels–Alder cyclization of the bis-dihydropyridine intermediate **105a** via isomerization to **105b** gives iminium salt **106a**. Redox exchange and hydrolysis of the resultant **106b** affords aldehyde **108**. Isolations of **103** and keramaphidin B (**107**) from
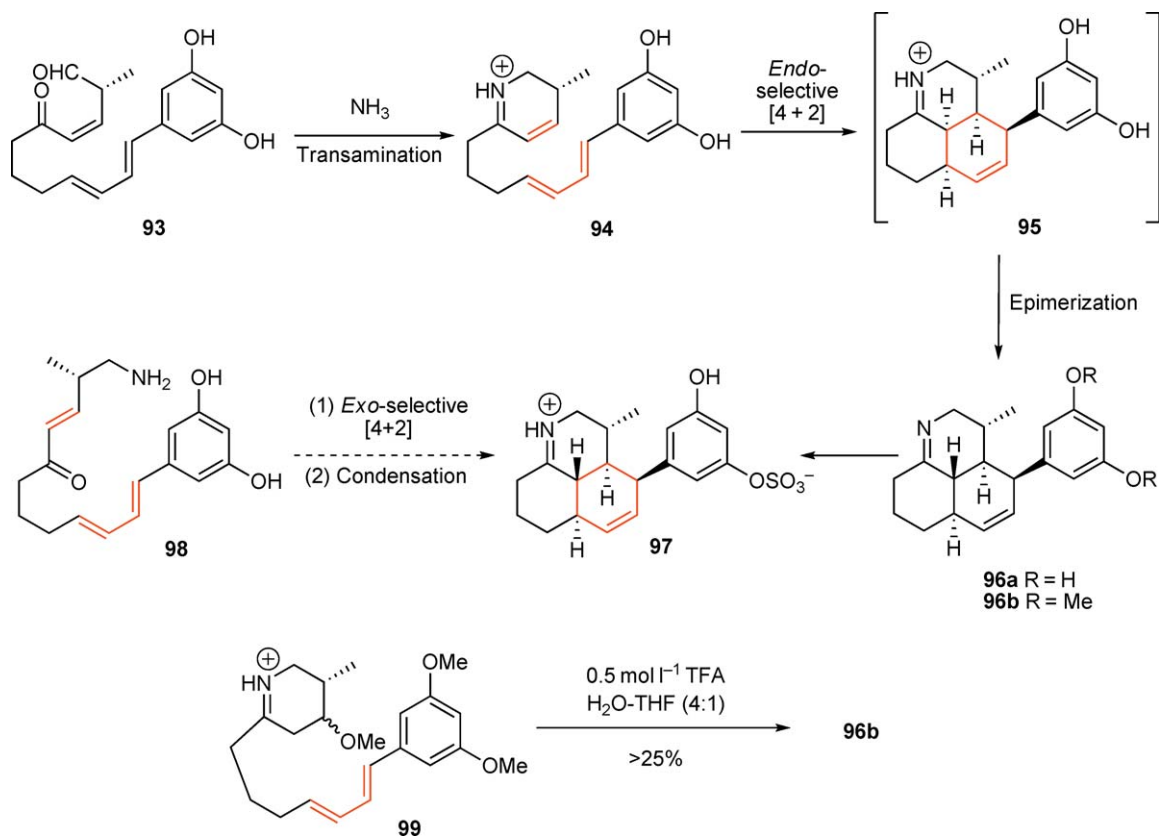
Scheme 16

Amphimedon sp.,[80] corresponding to reduced forms of plausible intermediates **108** and **106a**, strongly support this biogenetic pathway.

Baldwin *et al.*[81] have achieved the keramaphidin synthesis via an intramolecular Diels–Alder reaction of the proposed precursor, macrocyclic dihydropyridinium salt **105a** (**Scheme 18**). Cycloaddition of **105a** in a MeOH buffer followed by NaBH₄ reduction produced keramaphidin B (**107**) in 0.2–0.3% yield. This result is the first chemical evidence for the Baldwin and Whitehead's hypothesis.

8.08.2.2.4   Miscellaneous

In the preceding sections, possible involvement of biosynthetic Diels–Alder reaction with oxidative enzymes such as oxidase and dehydrogenase, PKS- and imine-forming enzyme are described. A series of cytotoxic prenylxanthones, forbesione[82] **109a**, isoforbesione **109b**, and gaudichaudic acid F[83] **110**, have been isolated from the *Garcinia* species. On the basis of biomimetic synthesis,[84,85] it is proposed that their molecular skeleton is constructed through Claisen rearrangement of prenylated phenol **111** and the subsequent intramolecular Diels–Alder reaction to afford adduct forbesione **109a** as shown in **Scheme 19(a)**. Quantum mechanical

Scheme 17

calculations[86] demonstrate that the Claisen rearrangement is reversible, and the irreversible Diels–Alder reaction determines the regioselectivity on Claisen step (four possible pathways concerning C5 and C6 positions). Since the Claisen rearrangement produces a chiral intermediate 112 from achiral precursor 111, this step would be catalyzed by a novel type of enzyme.
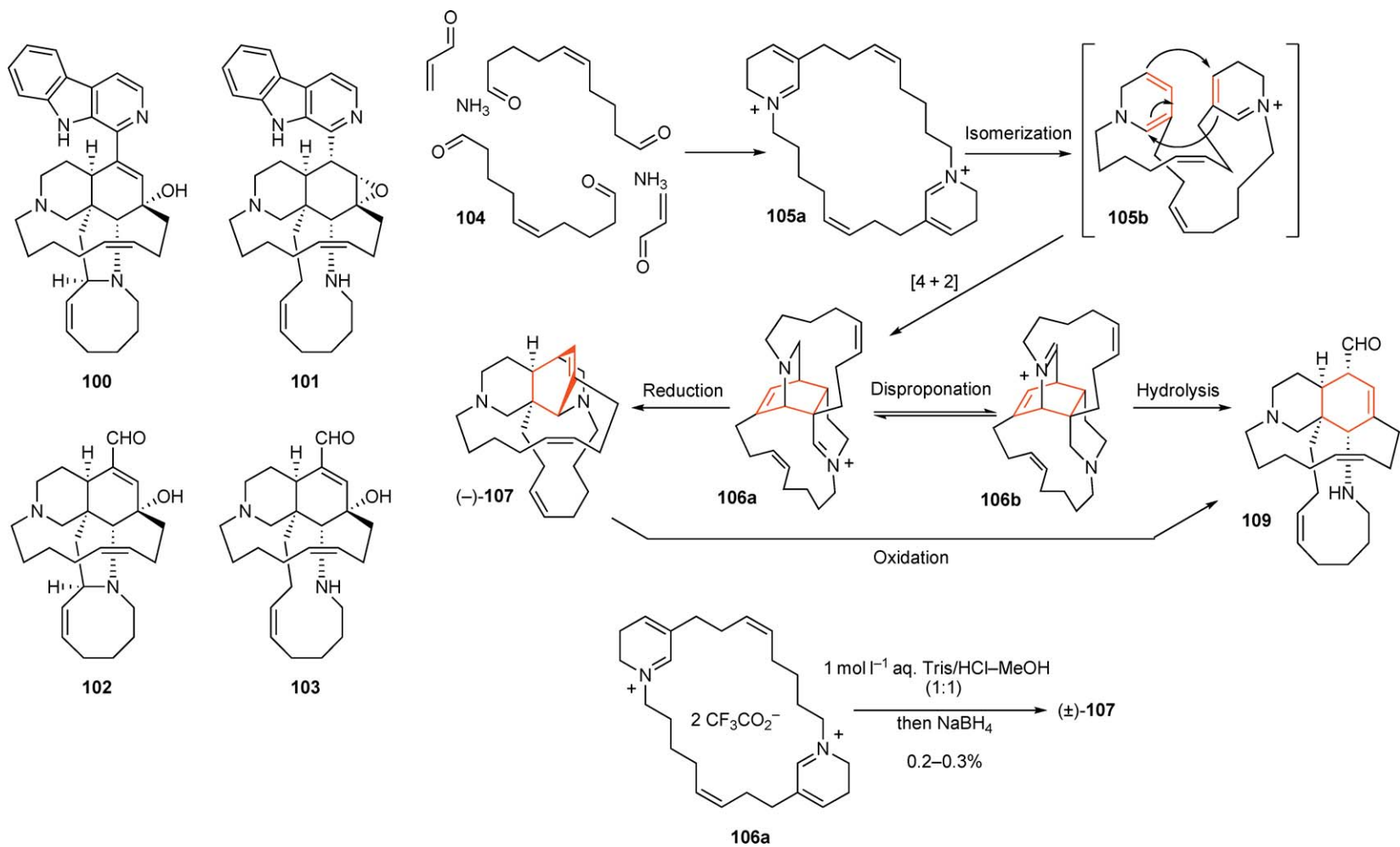
Andibenin B (117) is a highly oxidized meroterpenoid produced by the fungus *Aspergillus variecolor*. On the basis of the biosynthetic study,[87] it was proposed that a plausible intermediate 115 from Claisen rearrangement of 114 affords the adduct 116 via the intramolecular inverse-electron demand [4 + 2] cycloaddition as shown in Scheme 19(b). In this case, an enzyme catalyzes the cyclization of the terpene part on intermediate 114 derived from farnesyl diphosphate, and benzoates might provide the reactive dienophile 116.

8.08.3   Diels–Alder Reaction Catalyzed by the Biological System

There are several examples of Diels–Alder reactions catalyzed by catalytic antibodies and RNA. Their mechanisms will be discussed in this section.

8.08.3.1   Diels–Alder Ribozymes

After ribozyme (100-mer) catalyzing Diels–Alder reaction was discovered in 1997 using the unique selection system,[88] small Diels–Alder ribozyme (38-mer) with remarkable catalytic efficiency (20 000 relative to the uncatalyzed reaction) was developed.[89] Recently, the novel ribozyme that catalyzed [4 + 2] cycloaddition between tethered diene and biotinylated maleimide (Scheme 20) and the crystal structure of Diels–Alder ribozyme in the unbound form and in complex with a reaction product has been solved.[90]

Scheme 18

109a

109b

110

(a)

111

Claisen rearrangement

112

[4 + 2]

113

109a

(b)

114

115

[4 + 2]

116

117
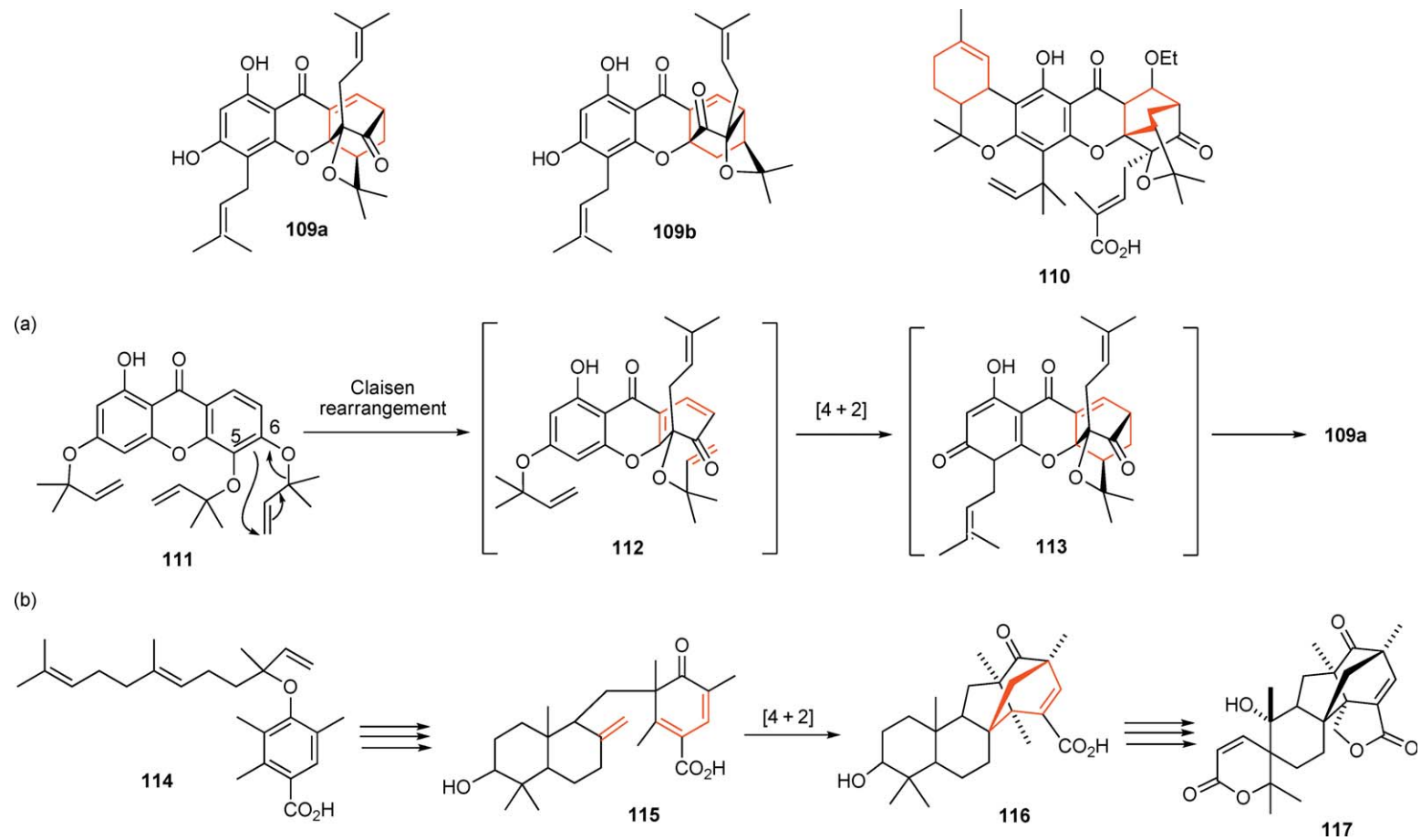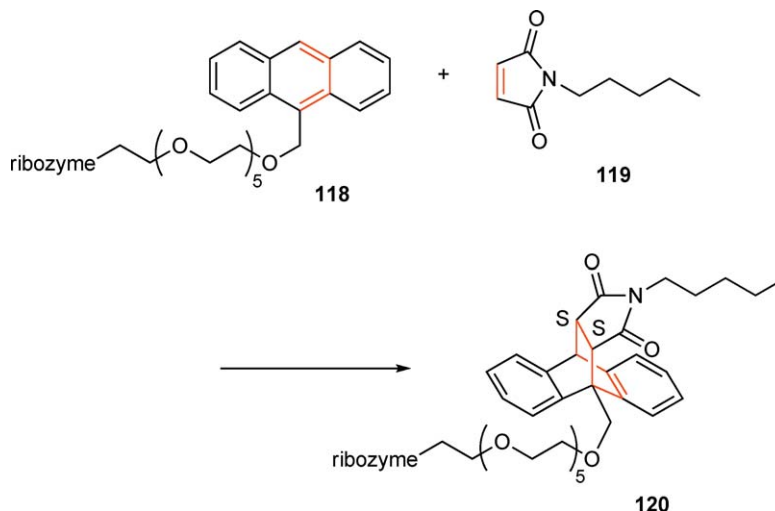
Scheme 19

Scheme 20

Based on the structure of the ribozyme complexed with the adduct, it was proposed that in the active site of the ribozyme, orientation of substrate anthracene was defined by stacking with two purine bases and also the carbonyl group of the maleimide portion was hydrogen bonded with exocyclic amine on base. Thus, these interactions not only stabilized the transition state of $[4+2]$ cycloaddition, but also activated diene and dienophile substrates. The same strategy was observed in the antibody counterpart.

8.08.3.2 Diels-Alderase Antibodies

The Diels–Alder reaction usually does not require a catalyst,[91] and therefore, rate acceleration may be achieved by stabilizing the transition state. In fact, antibodies elicited by transition state analog are able to catalyze Diels–Alder reactions. The Diels-Alderase antibodies provide a pocket acting as an entropy trap. Using this strategy several Diels-Alderase antibodies have been created to date.[92–96] Product inhibition is an inherent problem in the use of Diels-Alderase antibodies because of the resemblance between the product and the transition state.

To escape product inhibition, unstable boat-like haptens are used to prepare Diels-Alderase antibodies. Braisted and Schultz[93] raised antibody 39-A11 with hapten **124** fixed in a boat form. The antibody 39-A11 catalyzed the Diels–Alder reaction between **121** and **122** to yield the chiral adduct **123** (**Scheme 21**).

exo-Adducts are usually minor products of Diels–Alder reactions due to the absence of secondary orbital interactions between the diene and the dienophile. To catalyze this disfavored process, Lerner and coworkers developed 22C18 using a rigid bicyclic transition state analog **128a**.[95] Out of the eight possible isomers, the reaction between **125** and **126** using 22C18 gave a single product **127a** with high regio-, diastereo-, and enantioselectivity (**Scheme 22**). Using a similar strategy, *endo*-specific antibody 7D4 was prepared with hapten **128b**. Later, Janda and coworkers prepared Diels–Alder antibodies 4D5 and 13G5 with the ferrocene derivative **129** as hapten, as a loose mimic of the early transition state[96] (**Scheme 22**). Since the ferrocene ring of **129** can rotate freely, regioselectivity of the antibody obtained might be reduced. Interestingly, 13G5 catalyzed the cycloaddition between **125 and 126** to afford only ortho adduct (*exo*) **127b** with high regio-, diastereo-, and enantioselectivity. The antibodies 4D5 and 13G5 showed similar catalytic efficiency to that of the antibodies 22C18 and 7D4 raised with conformationally rigid haptens.

With hapten **134**, a close analog of the transition state, Hilvert *et al.*[92] developed the Diels-Alderase antibody 1E9 in which elimination of the product is programmed to avoid product inhibition. In their study, the cycloaddition of reactive thiophene dioxide **130** and maleimide **131** affords the unstable adduct **132** that spontaneously eliminates SO_2 followed by air oxidation to yield a stable phthalimide **133** (**Scheme 23**). The structure of the final product **133** is significantly different from that of the Diels–Alder adduct **132**, and product
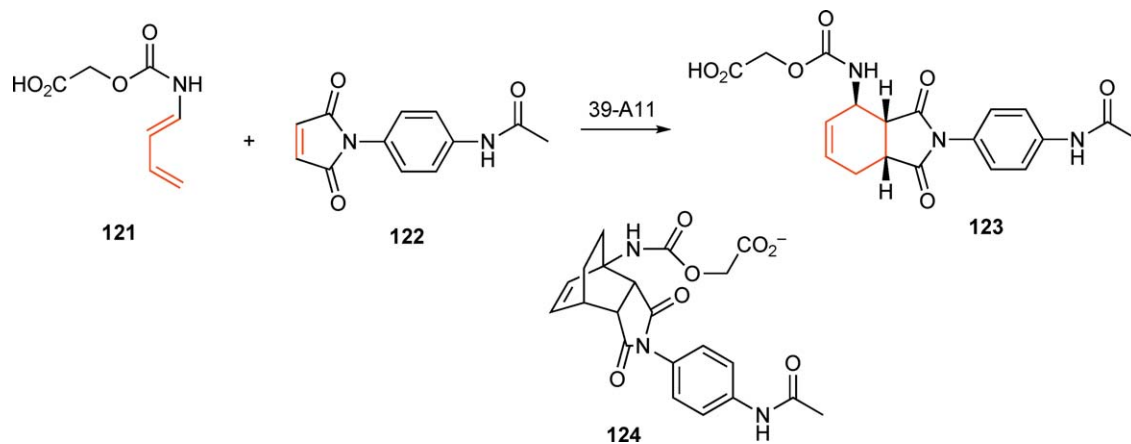
Scheme 21
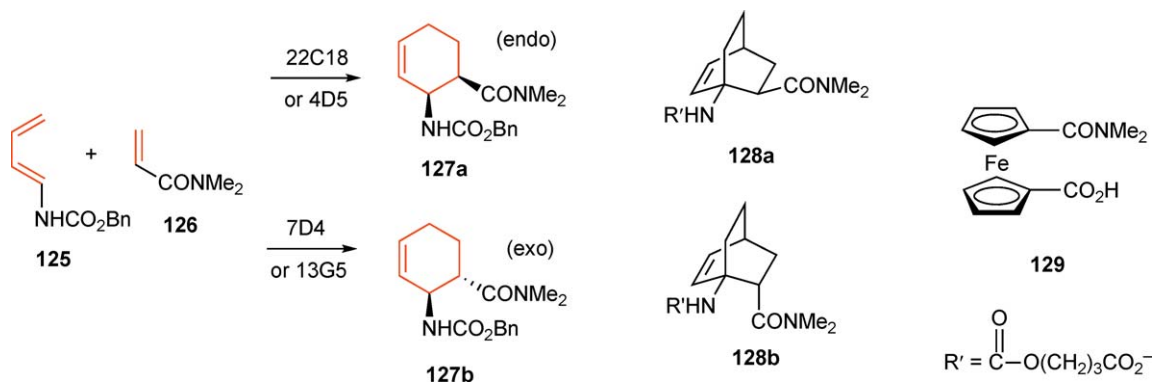
Scheme 22
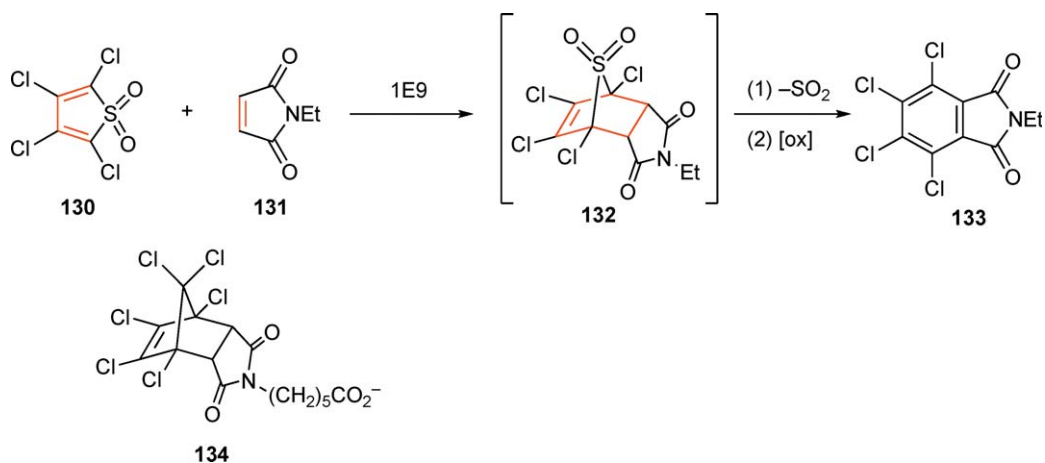
Scheme 23

inhibition is thus effectively minimized making 1E9 the most efficient catalytic antibody reported (k_{cat}/k_{uncat} 100 mol l^{-1}).

Crystal structures of three Diels-Alderase antibodies complexed with haptens (39-A11 and **124**,[97] 13G5 and **128b**,[98] and 1E9 and **134**[99]) are now available. In all cases, hydrogen bonding and significant numbers of van der Waals interactions between haptens and amino acid residues in the active sites are observed. The hydrogen bonding helps to fix the orientation and conformation of the substrate but also activates the dienophile. These data provide important information on the catalytic mechanism of Diels-Alderase antibodies.

8.08.4   Enzyme-Catalyzed Diels–Alder Reactions

Since we reported the enzymatic activity of SPS in 1995 as the first Diels-Alderase,[19] two additional Diels-Alderases, lovastatin nonaketide synthase (LNKS)[100] and macrophomate synthase (MPS),[101] have been purified and characterized. Two of these catalyze intramolecular Diels–Alder reactions while the third catalyzes an intermolecular Diels–Alder reaction. We have recently reported the detailed reaction pathway[102,103] of MPS and its catalytic mechanism based on the crystal structure.[104] In this section, we describe three natural Diels-Alderases and discuss the mechanism of their catalysis.

8.08.4.1   Solanapyrone Synthase

Solanapyrones were isolated as phytotoxic substances from phytopathogenic fungi *Alternaria solani*[105–108] and *Ascochyta rabiei*.[109] and an unidentified marine fungus.[110] The solanapyrone family consists of diastereomers A (**139a**) and D (**139b**), and their reduced forms B (**140a**) and E (**140b**). Isolation of these substances as optically active forms strongly indicates that solanapyrones are biosynthesized from the achiral linear triene precursor prosolanapyrone III (**138**) through an enzyme-catalyzed Diels–Alder reaction. The biomimetic synthesis of **139a and 139b** through a [4 + 2] cycloaddition proved the feasibility of the biosynthetic Diels–Alder reaction.[111–112] Incorporation of isotopically labeled biosynthetic precursors, prosolanapyrones I (**136**) and II (**137**), into (−)-solanapyrones unambiguously confirmed the biosynthetic pathway of solanapyrones as shown in **Scheme 24**.[18,113]

To establish the involvement of Diels-Alderases in this reaction, the enzymatic conversion of **137** and **138** was examined next. In cell-free extracts of *A. solani*, we found enzymatic activity catalyzing the Diels–Alder reaction[19] from **137** to (−)-**139a** with excellent enantioselectivity (99% ee) and relatively high *exo*-selectivity (6:1). Subsequently, we reported the partial purification and properties of the enzyme, SPS,[114] which is the first example of a Diels-Alderase. In addition, we showed that in the presence of molecular oxygen the crude enzyme converted **137** into **139a and 139b** with accompanying formation of hydrogen peroxide. Based on the chromatographic behavior of the enzyme,[113] we proposed that the single enzyme catalyzes the oxidation from the alcohol **137** to the reactive aldehyde **138** that is further converted into the adducts **139a and 139b** by the Diels–Alder reaction. Although purification of SPS was hampered by its instability, SPS has been purified as a single band on SDS-PAGE.[115] This provided information of N-terminal amino acid sequence that is identical to that from SPS synthase gene described later.

To assess the diastereoselectivity and the intrinsic reactivity of prosolanapyrones, Diels–Alder reactions were examined under various conditions.[112] In less polar solvents, heating was required for the effective cycloaddition of **136–138**. Increase in the oxidation levels of the three substituents in the prosolanapyrones enhances rate acceleration. This can be rationalized in terms of the LUMO energy of the dienophile moiety in the pyrone precursors. *Endo-/exo*-selectivities with **136–138** were essentially the same in various organic solvents, while the *endo*-selectivity was increased with increasing solvent polarity. The slight preference for *endo*-selectivity in less polar solvents suggests that there is little steric congestion in both *endo*- and *exo*-transition states as reported in the reactions of simple decatriene systems.[116]

When the nonenzymatic Diels–Alder reaction of prosolanapyrone III (**138**) was carried out in organic solvents under standard conditions, no reaction occurred. In aqueous medium, however, the reaction was accelerated and gave *endo*-adducts with high selectivity (**139a:139b** = 3:97).[112] These effects were observed in the reaction of **138** but not in that of **137**. This observation indicated that the oxidation of prosolanapyrone II
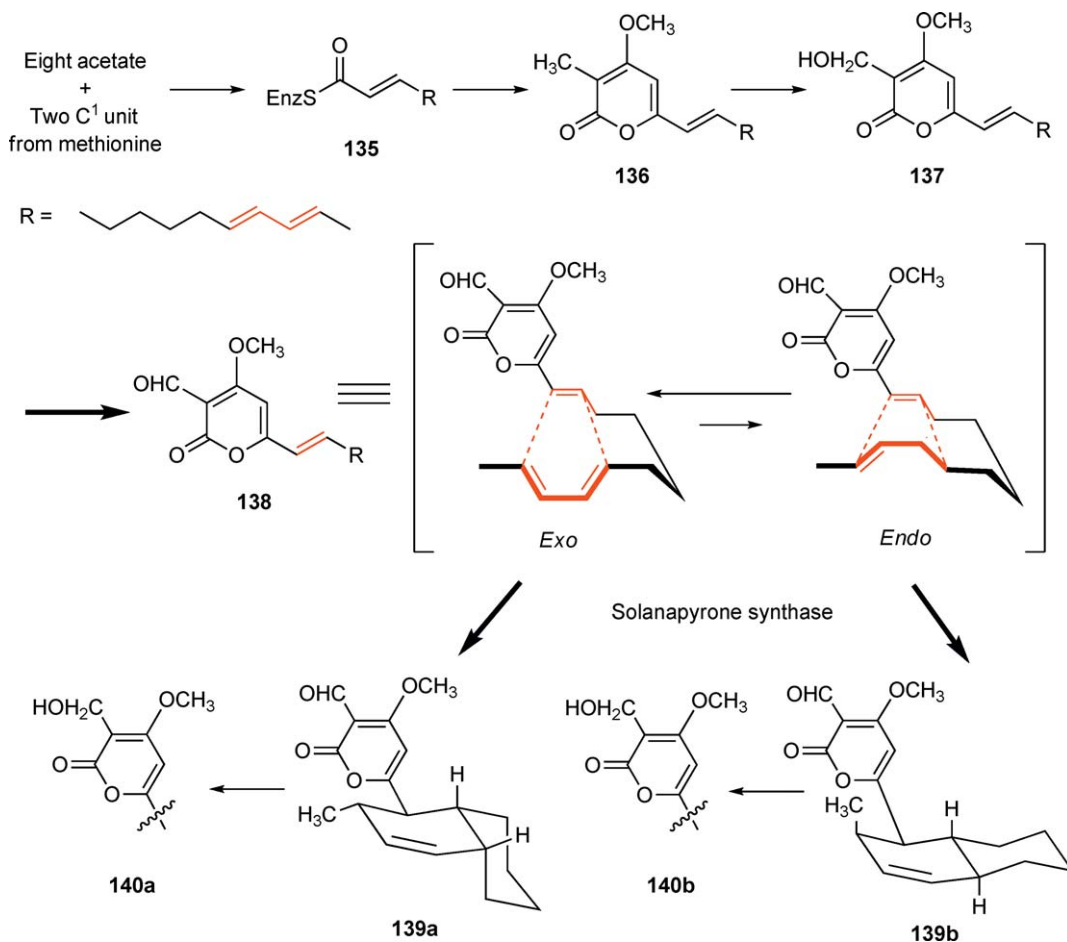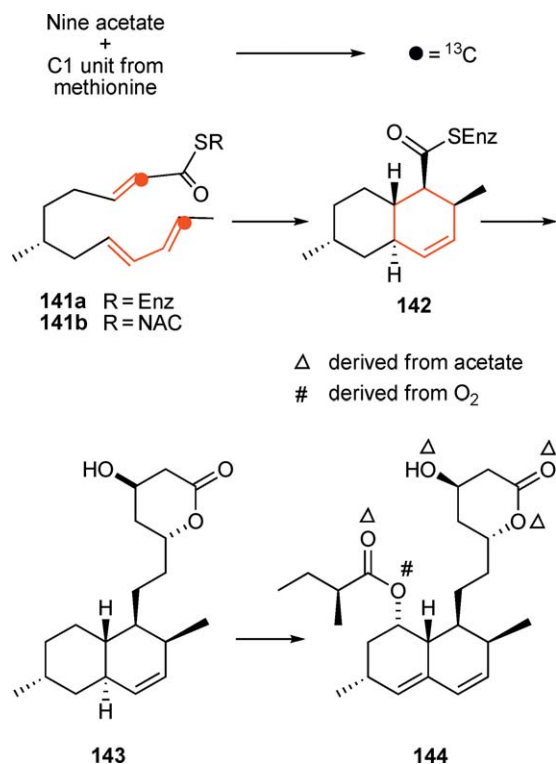
Scheme 24

(**137**) enhanced the reactivity of the substrate significantly for the Diels–Alder reaction. For Diels–Alder reactions in aqueous media, similar rate accelerations and the predominant formation of *endo*-adducts have been reported.[117,119] Breslow[117] explained this phenomenon by the hydrophobic effect: water forces the substrate to form the more compact *endo*-transition state, reducing its molecular surface exposed to the aqueous medium. On the other hand, Ruiz-López *et al.*[119] emphasized the importance of hydrogen bonding between the water and the dienophile carbonyl group to reduce the LUMO energy of the dienophile and to enhance the reactivity of the substrate. Due to the effects described above, the background reaction could not be ignored in the enzymatic reaction under standard conditions. Contrary to the nonenzymatic reaction, the enzymatic conversion of **137** provided preferentially *exo*-adduct **139a**. In general, the *exo*-selective cycloaddition cannot be achieved by simple heating or by the use of Lewis acid catalysts. These observations indicate that the major function of SPS is the oxidation of prosolanapyrone II (**137**) to the more reactive III (**138**) and the stabilization of the *exo*-transition state.[112]

Recently, the gene cluster responsible for the biosynthesis of solanapyrones has been identified by homology-based PCR and genome walking.[120] The gene cluster contains the PKS gene for backbone construction and several genes for modification enzymes, including FAD-dependent monooxygenase, most likely SPS. This monooxygenase was overexpressed in yeast and showed enzymatic activity catalyzing both the oxidation and the Diels–Alder reaction. The character of this enzyme was identical to that of crude SPS from *A. solani*. Thus, the first Diels-Alderase is ready for a detailed analysis of the reaction mechanism elucidating that the corresponding cycloaddition proceeds in a concerted manner.
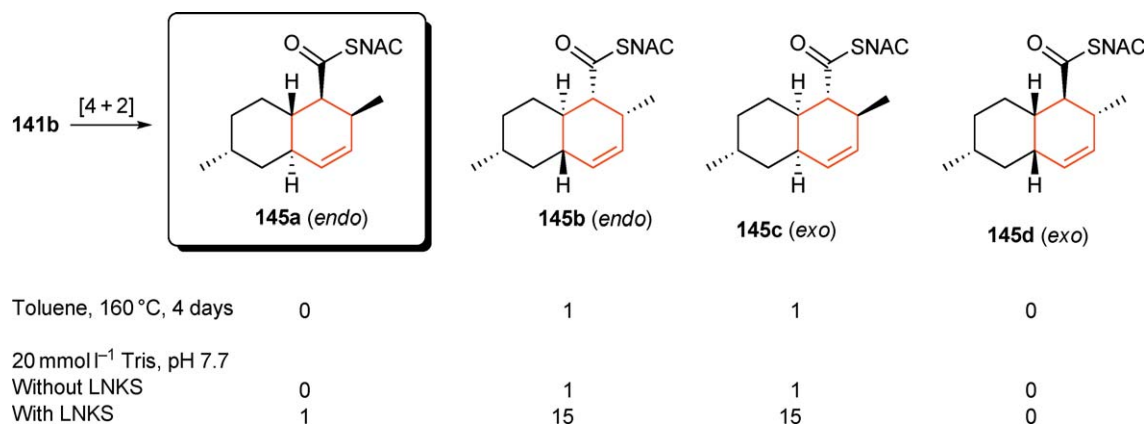
8.08.4.2   Lovastatin Nonaketide Synthase

The biosynthesis of the cholesterol-lowering drug lovastatin (**144**) isolated from *Aspergillus terreus* has been extensively investigated by Vederas and coworkers. Incorporation experiments with multiple labeled acetate and ^{18}O-oxygen suggested that oxygen atoms of the side chain are derived from acetate and the oxygen atom on the decalin ring is obtained from the molecular oxygen.[121,122] Later, a blocked mutant of *A. terreus* converted 4a,5-dihydromonacolin L (**143**) into **144**, confirming that these are intermediates (**Scheme 25**).[123] Because lovastatin (**144**) does not have an electron-withdrawing group in the dienophile moiety, it was proposed that the requisite Diels–Alder reaction occurred at the hexaketide stage. The researchers synthesized ^{13}C-labeled hexaketide precursor **141b** to test this hypothesis.[124] No incorporation was observed in the feeding experiment using labeled **141b**. In addition, recovery of **141b** was less than 1% despite considerable efforts to suppress β-oxidation of **141b**. The cycloaddition of hexaketide **141b** in aqueous media proceeded at 28 °C to afford a 1:1 mixture of adducts **145b** (*endo*) and **145c** (*exo*), while the same reaction in toluene required heating at 160 °C to yield the same adducts in the same ratio (**Scheme 26**). This clearly showed a significant rate acceleration in aqueous media (half-life of **141b**: 2 days). Thus, the facile conversion and rapid degradation of **141b** with intact cells hampered detection of significant incorporation.

In 1999, the biosynthetic gene cluster of lovastatin (**144**) was cloned by Hutchinson's group.[125] They succeeded in achieving heterologous expression of whole genes in *A. nidulans* to produce dihydromonacolin L (**143**). Collaborating with the Hutchinson group, the Vederas group started enzymatic studies using LNKS, which is responsible for the construction of lovastatin backbone.[100] Since it is known that LNKS requires lovC protein for the production of **143**, a cell-free system was prepared from the recombinant strain producing **143**. Incubation with the synthetic substrate **141b** and the required cofactors and substrates did not yield **143**, indicating that decomposition, probably hydrolysis of **141b**, occurred during incubation. Using the purified LNKS, hexaketide triene precursor **141b** was incubated without cofactors and substrates to give three adducts **145a**–**145c** in a ratio (1:15:15). Minor *endo*-product **145a** was confirmed to be the one with the same stereochemistry as natural **144**. Due to the inability of the denatured enzyme to form the adduct **145a**, LNKS

Scheme 25

141b $\xrightarrow{[4+2]}$	145a (*endo*)	145b (*endo*)	145c (*exo*)	145d (*exo*)
Toluene, 160 °C, 4 days	0	1	1	0
20 mmol l^{-1} Tris, pH 7.7 Without LNKS	0	1	1	0
With LNKS	1	15	15	0

Scheme 26

catalyzed this cycloaddition with a significant rate acceleration. In the lovastatin biosynthesis, hexaketide precursor **141b** should load on the corresponding ketosynthase domain of LNKS, then it is processed downstream to yield **144**. Since the adducts **145a–145c** were obtained as NAC thioesters; the obligatory thioester exchange did not occur in the Diels–Alder reaction. To understand the detailed mechanism of LNKS, conversion of the adduct **145a** into dihydromonacolin L (**143**) by LNKS is required. It is interesting that SPS catalyzes the Diels–Alder reaction after chain elongation, while LNKS catalyzes it during the polyketide chain construction. Since most of the modification reactions in polyketide biosynthesis take place after chain elongation, LNKS is the first enzyme capable of chain modification prior to the termination of polyketide chain extension.

8.08.4.3   Macrophomate Synthase

The phytopathogenic fungus, *Macrophoma commelinae* has the ability to transform 2-pyrone **146** into the corresponding benzoate analog macrophomate (**152**) (**Scheme 27**).[126] This complex aromatic conversion is catalyzed by a single enzyme, MPS,[101,127] with oxalacetate as a substrate for the C3-unit precursor. MPS is a Mg^{2+}-dependent enzyme with 339 amino acid residues (MW = 36 244 Da). The catalytic mechanism of the whole pathway was investigated extensively, and it was shown that it proceeds through three separate steps, including decarboxylation, two carbon–carbon bond formations, and decarboxylation with concomitant dehydration.[102,103,128] In the absence of 2-pyrone **146**, MPS simply acts as a decarboxylase with high catalytic efficiency (**Scheme 27**).

The crystal structure of the MPS complexed with pyruvate and Mg^{2+} was determined with a resolution of 1.70 Å (**Figure 4**).[104] In the crystal structure, the C-terminal 40 residues (residues 300–339) that are not important in the catalysis were deleted. The molecule is hexameric, and the protomer core region consists of eight-stranded β-barrel surrounded by 8 + 3 α-helices with a $(\beta/\alpha)_8$ barrel fold.

At the catalytic cavity, Mg^{2+} is located in an octahedral coordination site (**Figure 5(a)**). Two of the ligands of Mg^{2+} are the side-chain carboxyl oxygens of Glu185 and Asp211. Another two of the coordination sites are filled with two water molecules, which are in turn hydrogen bonded to the protein. The last two coordination sites are occupied by the C2-carbonyl and C1-carboxyl oxygen atoms of pyruvate enolate. This firmly bound structure clearly defines the precise orientation of the pyruvate enolate (**Figure 5(a)**) of which two carboxyl oxygen atoms are also hydrogen bonded to the main-chain amide protons of Gly210 and Asp211. This complex is further stabilized by interaction between the carbonyl oxygen of pyruvate and the side chain of Arg101. With these bonds, the pyruvate enolate is tightly placed in this position. The active site is at the C-terminal end of the β-barrel, which is covered by the loop from the threefold-related chain.

On the basis of this structural information, the pathway of the MPS reaction can be outlined as follows (**Scheme 27**): oxalacetate is incorporated into the active site of MPS in a similar way to that of pyruvate. Lewis acidity of the magnesium promotes decarboxylation to form the enolate anion, which is stabilized by an electron sink provided by the divalent cation.[129] Steric congestion of the peptide backbone allows the 2-pyrone
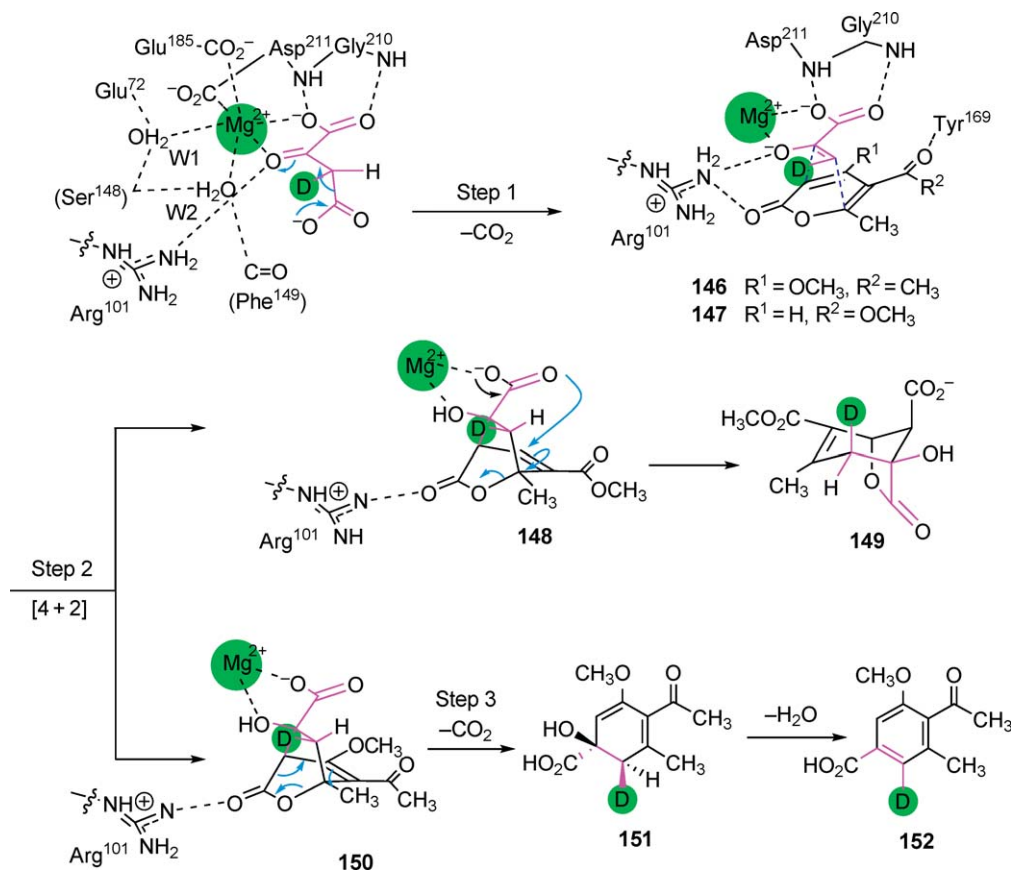
Scheme 27

146 access only from one side of the enolate plane where the catalytic pocket is open. As shown in **Figure 5(b)**, the 2-pyrone molecule is fixed in place through two hydrogen bonds between the carbonyl oxygen of 2-pyrone and Arg101 and the C5-acyl oxygen of 2-pyrone and Tyr169. The flexible loop (residues 139–170) with hydrophobic side chains (Phe149, Pro151, and Trp152) from the threefold-related protomer shields this transition state from the solvent. The stacking direction of 2-pyrone and pyruvate enolate is exactly as expected from the product. The importance of the hydrogen bonds in the present model is confirmed by the experiment with two mutants, R101S and Y169F. Both mutations dramatically disturbed MPS activity while retaining the decarboxylase activity. These hydrogen bonds act not only in substrate recognition but also enhance reactivity in the inverse electron demand Diels–Alder reaction by reducing the LUMO energy of the diene. The binding model explains the substrate specificity[130] and stereochemical course of the whole reaction pathway.[102]

   On the basis of the formation of the aberrant adduct with pyrone[103] and the observation that dehydration proceeds formally in an antisense,[102] it was proposed that the higher energy [4 + 2] adducts 148 and 150 are transformed to either the benzoate analog 152 or the rearranged product 149 as shown in **Scheme 27**. The binding structure indicates that the free carboxylate oxygen is located close to the C4 position in the electron-deficient olefin. Thus, the carboxylate in the higher-energy bicyclo[2,2,2]octane 148 attacks the β-position of the α,β-unsaturated ester moiety followed by the rearrangement and the ring opening to afford stable bicyclo[3,2,1]octane 149. In the reaction with the normal 2-pyrone 146, substitution of the electron-donating methoxy group reduces the reactivity of the α,β-unsaturated ester, resulting in the fact that the nucleophile, C1-carboxylate, cannot attack this moiety. This alters the reaction path to form the aromatic compound 152.

   The stereochemical course of conversion from the adduct 150 into macrophomate 152[102] is shown in **Scheme 27**. Since inspection of the active site does not identify any basic residue proximal to the pro-*R* proton, the C1-carboxylate may abstract the proton by Arg101 accepting the developing charge on the lactone oxygen.
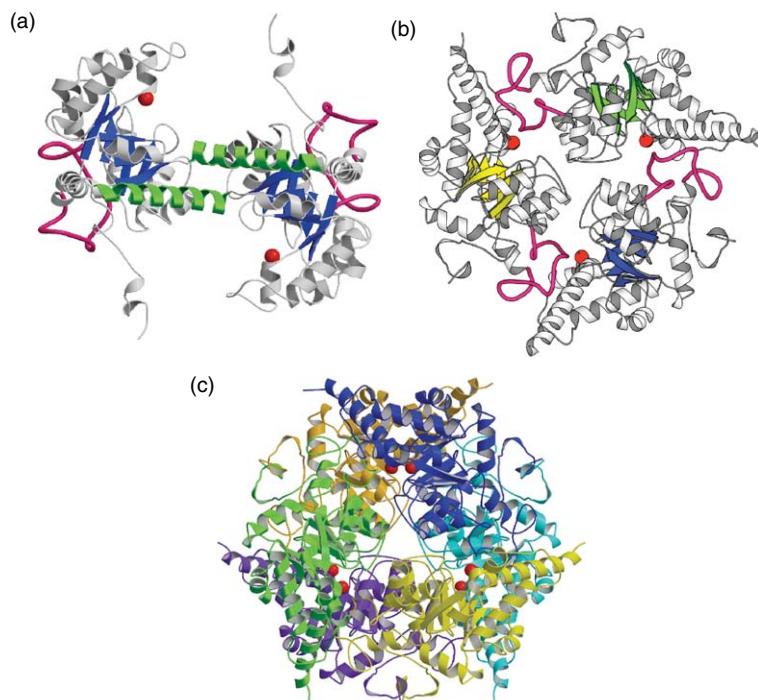
Figure 4   Crystal structures of MPS, dimer (a), trimer (b), hexamer (functional unit) (c).
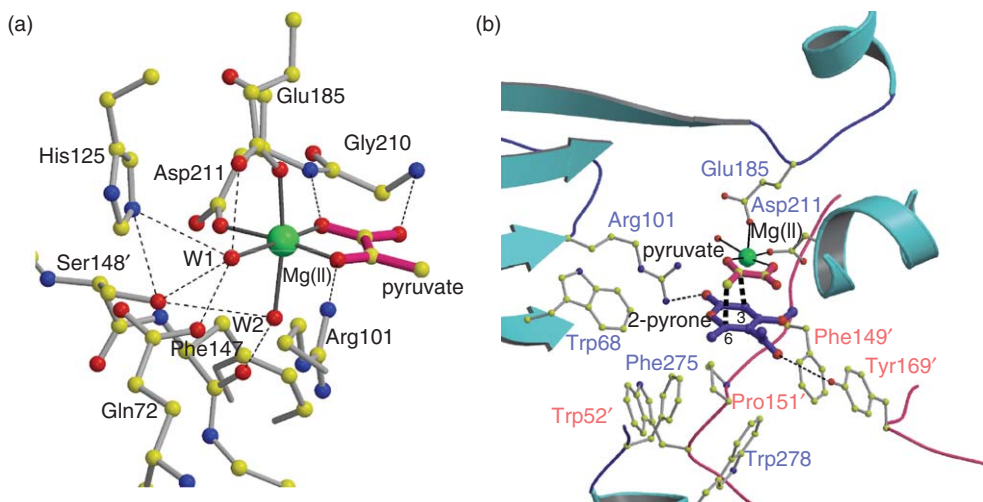
Figure 5   Structures of MPS active site, MPS and pyruvate complex (a), a model complexed with pyruvate enolate and **146** (b).

Kinetic analysis of MPS for the decarboxylation (k_{cat} 16.3 s^{-1}), the aberrant adduct formation (k_{cat} 5.9 s^{-1}), and the overall reaction (k_{cat} 0.6 s^{-1}) reveals that the last degradation step is the rate-determining step.[102]

Extensive point mutation experiments[131] on MPS identified essential amino acid residues and remarkable tolerance for mutation. Recently, it was found that the third step of the MPS reaction is decarboxylation to afford intermediate **151** and that the subsequent dehydration is not catalyzed by MPS.[132] Unstable intermediate **151** was released from the active site and dehydrated to give macrophomate (**152**) in a nonenzymatic manner.

QM/MM calculation[133] on the MPS- catalyzed reaction suggested that the transition state of the two-step Micheal-aldol route is more stable than that of the concerted Diels–Alder route, indicating that the two-step route is an energetically preferred process. To evaluate the validity of the calculation results, experimental evidence must be provided. A recent finding that the first decarboxylation, the second C—C bond formation, and the third decarboxylation to **151** are rapid processes,[132] suggested that it is difficult to distinguish between concerted or stepwise processes experimentally.

8.08.5   Natural Diels–Alder Adducts Whose Biosynthetic Gene Clusters Have Been Identified

The rapid progress in polyketide biosynthesis allowed to identify the biosynthetic gene clusters of several Diels–Alder adducts such as equisetin, chaetoglobosin, spinosyn, chlorothricin, and kijanimicin. Possible enzymes that catalyze the corresponding Diels–Alder reaction in the biosynthesis of these polyketides will be discussed.

8.08.5.1   Chaetoglobosins and Equisetin

A biosynthetic pathway of the fungal metabolite chaetoglobosin A (**155**)[134] from *Chaetomium globosum* and *Ch. subbafine* was extensively investigated by the incorporation experiments with isotopically labeled precursors showing that the backbone of **155** is constructed with a PKS–NRPS hybrid.[135–137] Detailed information on post-PKS biosynthesis is provided by the incorporation of ^{18}O-labeled acetate and oxygen gas[137] and by the inhibition studies[138] with cytochrome P-450 inhibitors, indicating that a series of introduction of the oxygen atoms with monooxygenases occurs after the construction of the cytochalasan skeleton (**Scheme 28**). The involvement of Diels-Alderase in the formation of prochaetoglobosin I (**154a**)[137] was proposed by the experimental evidence that heating of **154a** provided *exo*-adduct **154b**, indicating reversible cycloaddition between a plausible Diels–Alder precursor **153** and the *exo*- and *endo*-adducts.
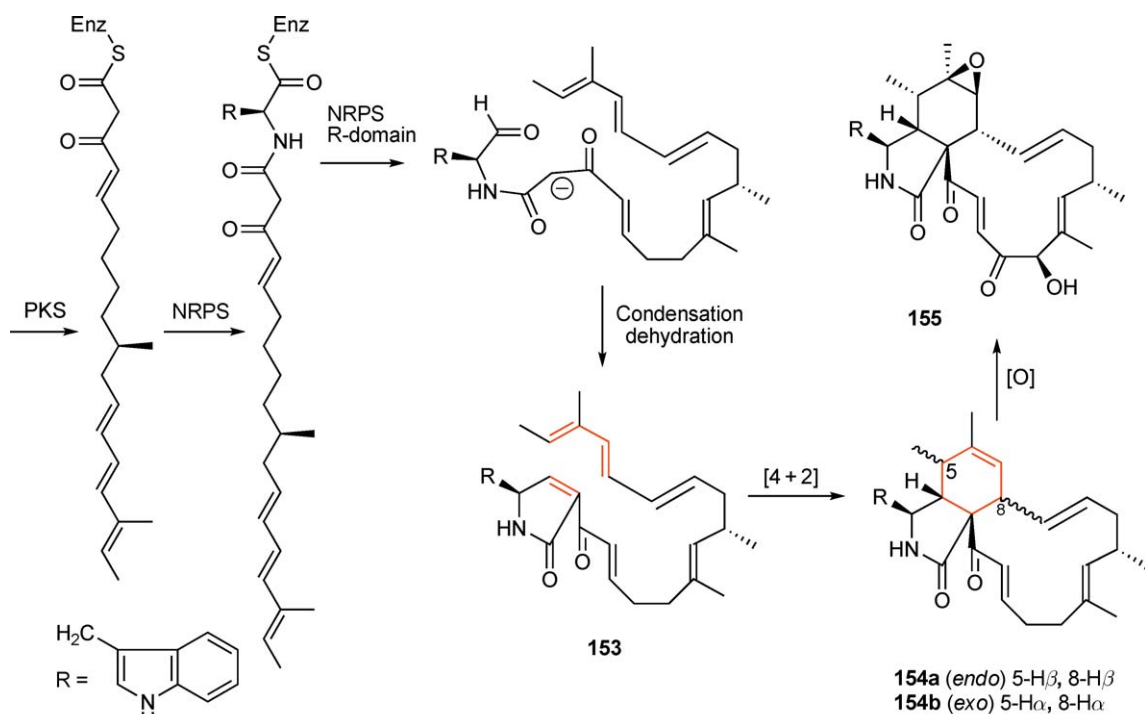
Scheme 28
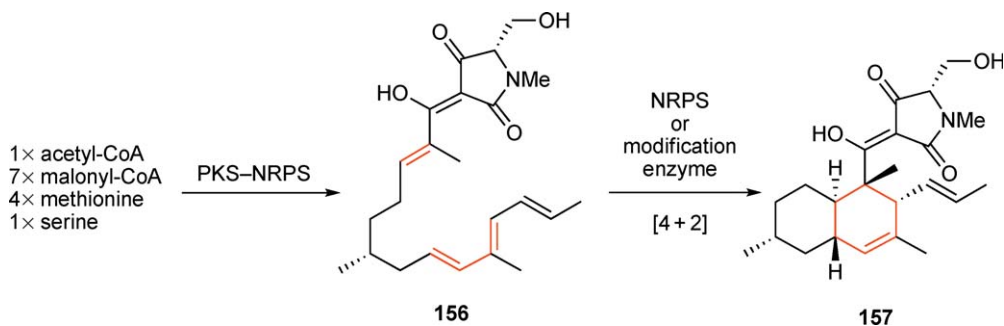
Scheme 29

Recently, the gene cluster of chaetoglobosin has been identified.[139] It consists of seven genes, a PKS–NRPS hybrid (CheA) and an enoyl reductase (CheB) for backbone construction, three monooxygenases (CheD, E, G), and two transcriptional regulators. Intriguing macrocyclic ring formation needs formation of dienophile deoxytetramic acid moiety. Since CheA has reduction domain at the C-terminal, matured polyketide backbone is cleaved off from CheA to give the aldehyde. Condensation and dehydration are necessary to form the plausible Diels–Alder precursor **153**. Currently, no information on this transformation is available from genetic analysis. It is possible to speculate that the enzyme responsible for the formation of deoxytetramic acid moiety catalyzes the intriguing Diels–Alder reaction. Since synthesis of the putative precursor analog has already been achieved, intriguing transformation can be studied at the enzyme level.[140]

Acyltetramate polyketide precursors produce not only cytochalasin-type metabolites but also alkyldecalin adducts that are frequently obtained as fungal metabolites (Section 8.08.2.3.2). The fungal metabolite, equisetin, (**157**) is a typical decalin polyketide. Incorporation experiments with stable isotope-labeled precursors established that the molecular skeleton of **157** is constructed from an octaketide and L-serine (**Scheme 29**).[141] The gene cluster of **157** is closely related with that of LNKS.[142] The intriguing Diels–Alder reaction of plausible precursor **156** might be catalyzed by a PKS–NRPS hybrid (most likely dehydration domain) during polyketide chain elongation step as in the case of LNKS. That an alternative pathway of the Diels–Alder reaction, as in the case of SPS, takes place after the completion of full-length polyketide chain elongation could not be ruled out.

8.08.5.2 Chlorothricin

Decalin polyketides constructed by the intramolecular [4+2] cycloaddition (Section 8.08.2.3.2) contain another unique class of metabolites possessing a spirotetronate substructure. This unique moiety is proposed to be formed through the second Diels–Alder reaction between the polyketide-derived polyene precursor and the γ-methylene-acyltetronate moiety. Thus, the overall skeleton of spirotetronate antibiotic represented by chlorothricin[143] (**165**) is proposed to be constructed by a tandem Diels–Alder reaction as shown in **Scheme 30**. Liu and coworkers identified a gene cluster of chlorothricin biosynthesis.[144] In the biosynthesis of **165**, the first cycloaddition from **159** to **160** might be catalyzed by PKS as in the case of LNKS. After the formation of γ-methylene-acyltetronate moiety possibly with four gene products ChlD1–ChlD4 from glycerate and a poly-ketide chain in a similar way to that for the acyltetramate polyketide chain as shown in **Schemes 28** and **29**, the second cyclization of **163** possibly with a dehydratase ChlD3 that has a motif FAD-dependent monooxygenase would occur to afford chlorothricin core **164**. Although the intermediate of the cycloaddition is rather complex, this transformation can be examined by the enzymatic reaction with the synthesizing substrate and the overexpressed enzyme. Efficient synthesis of chlorothricin core of **168** and **169** by intra- and intermolecular Diels–Alder reactions of **166** and **167** supported this hypothesis (**Scheme 31**).[145] Independently, the biosynthesis of kijanimycin, which has a similar molecular skeleton to that of chlorothricin (**165**), has been proposed on the basis of genetic information.[146]
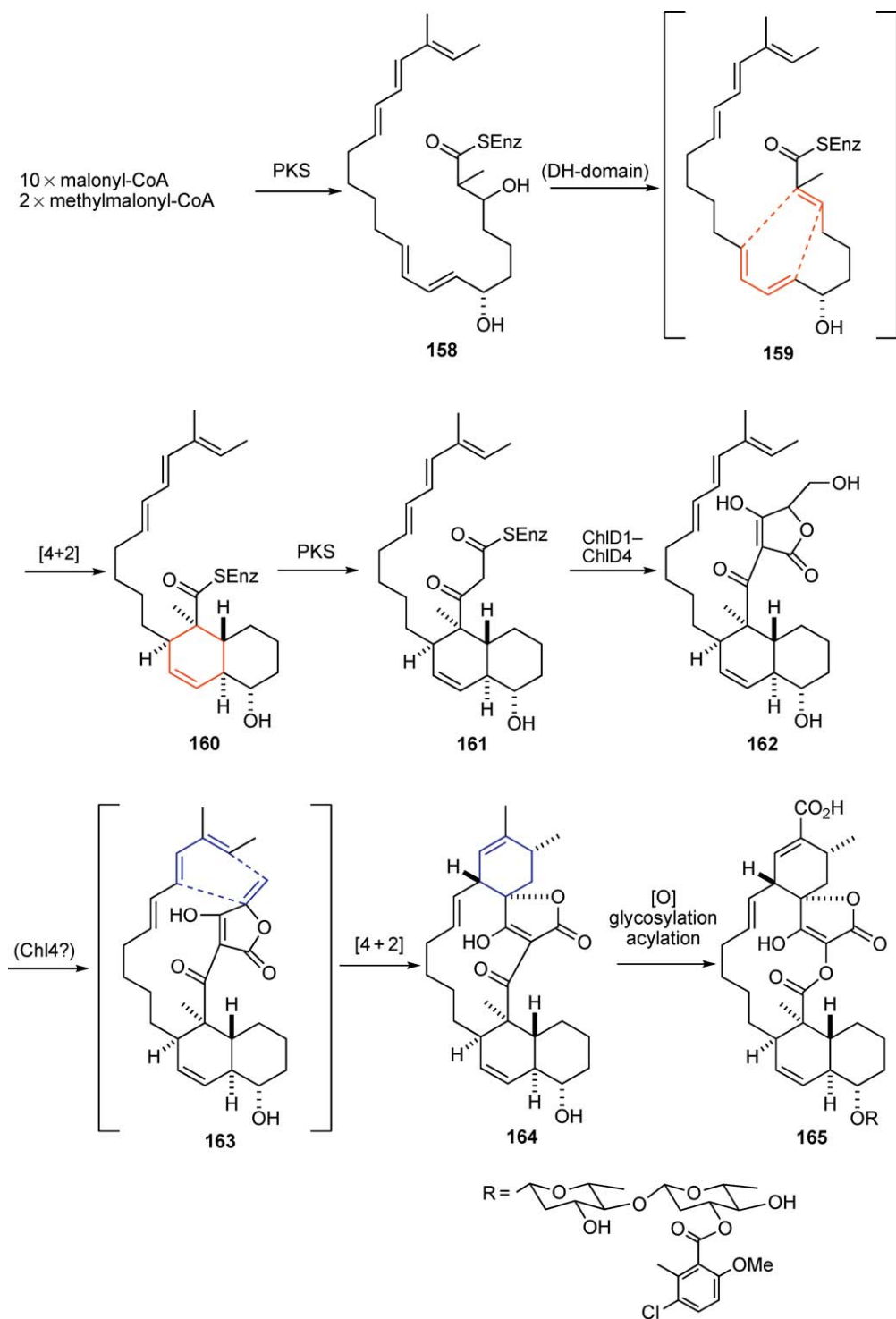
Scheme 30
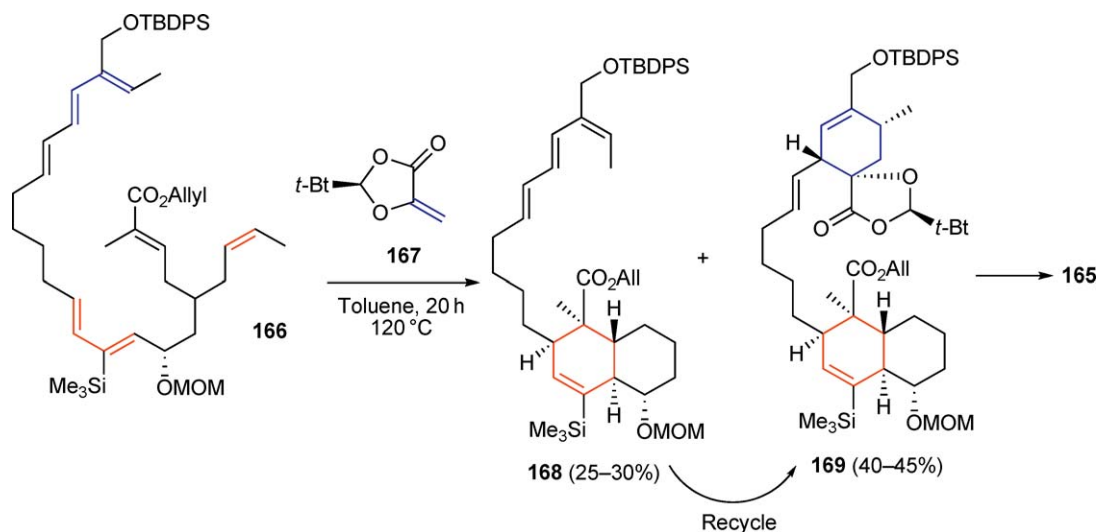
Scheme 31

8.08.5.3 Spinosyn

A different type of Diels–Alder cycloaddition is proposed in the biosynthesis of commercially available insecticide, spinosyn[147] (**174**, Naturlyte). Spinosyn (**174**) is a macrolide, that is constructed by PKS. Its biosynthetic gene cluster, consisting of genes for PKS, sugar biosynthesis, and modification enzymes, was cloned in 2001,[148] and the biosynthesis of **174** was proposed as shown in **Scheme 32**. Extensive gene disruption and bioconversion of the resultant mutants allowed to identify plausible intermediate **170** and most of the gene functions. Four genes, *spnF*, *spnJ*, *spnL*, and *spnM,* that are possibly responsible for the construction of a perhydro-*as*-indacene core **173** through a plausible Diels–Alder reaction of **172** remained as candidates of the Diels-Alderase gene. Recently, the function of *spnJ* was elucidated by the conversion of a synthesized putative precursor **170** into **171** with the overexpressed SpnJ.[149] This firmly established that the substrate of SpnJ is not an enzyme (PKS)-bound form, and narrowed down the gene responsible for [4 + 2] cycloaddition to be *spnF*, *spnL*, and *spnM*.

8.08.6 Conclusions

SPS, LNKS, and MPS, the three examples of natural Diels-Alderases, catalyze not only the Diels–Alder reaction but also oxidation, polyketide chain formation, and decarboxylation, respectively. These enzymes convert the corresponding substrates into reactive Diels–Alder substrates, which are not released from the active sites and readily undergo cycloaddition in the active sites by forcing them into reactive conformations. Thus, among the many natural plausible Diels-Alderases described in this chapter, the three examples shown above can at least be classified as producers of reactive substrates with an entropy trap for [4 + 2] cycloaddition. Compared with artificial biomolecular catalysts, these types of natural Diels-Alderases have obvious advantages to utilize highly reactive substrates that are not stable in the reaction media. The alternative candidate of Diels-Alderase is an auxiliary protein found in lignan biosynthesis that does not have a catalytic activity but has space for binding reactive species and enforcing enantioselective coupling.[150] To date, the corresponding Diels-Alderase has not yet been found.

In nature, there are a large number of bioactive secondary metabolites produced by microorganisms and plants, probably for proliferation of the producer under living conditions. Our accumulative knowledge on biosynthetic pathways of natural products indicates that the unique backbone of natural products such as polyketides, polypeptides, and terpenes is constructed by a relatively small number of biosynthetic enzyme systems. The Diels–Alder reaction provides a further option to diversify secondary metabolites since this reaction
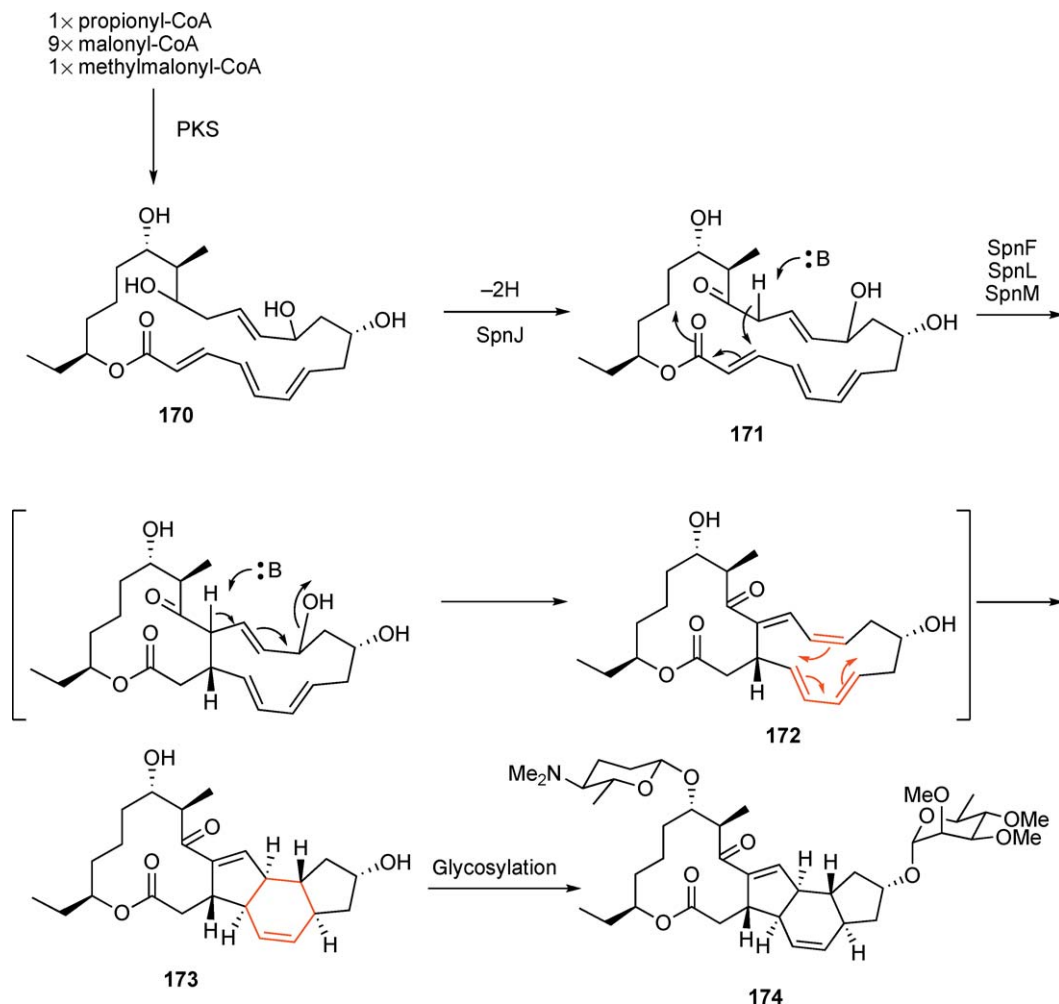
Scheme 22

requires the formation of reactive dienes and dienophiles. The cycloadditions may proceed to give optically active metabolites in the active site of the corresponding enzyme, which produces reactive diene or dienophile. On the other hand, it may proceed to afford racemates after releasing the product from the active site as in the case of the [4 + 2] adducts derived from plants. Thus, this simple system can create a variety of natural [4 + 2] adducts such as simple dimerization products, intramolecular reaction products, and multiple [4 + 2] cycloaddition products (Section 8.08.1). Further investigations are necessary to prove this hypothesis.

References

1. W. Carruthers, *Cycloaddition Reactions in Organic Symthesis*; Pergamon: Oxford, 1990.
2. K. C. Nicolaou; S. A. Snyder; T. Montagnon; G. Vassilikogiannakis, *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 1668–1698.
3. K. Takao; R. Munakata; K. Tadano, *Chem. Rev.* **2005**, *105*, 4779–4807.
4. K. C. Nicolaou; D. J. Edmonds; P. G. Bulger, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 7134–7186.
5. A. Ichihara; H. Oikawa, *Curr. Org. Chem.* **1998**, *2*, 365–394.
6. A. Ichihara; H. Oikawa, The Diels–Alder Reaction in Biosynthesis of Polyketide Phytotoxin. In *Comprehensive Natural Products Chemistry*; D. Barton, K. Nakanishi, O. Meth-Cohn, Eds.; Elsevier: Amsterdam, 1999; Vol. 1, pp 367.
7. H. Oikawa; T. Tokiwano, *Nat. Prod. Rep.* **2004**, *21*, 321–352.
8. H. Oikawa, *Bull. Chem. Soc. Jpn.* **2005**, *78*, 537–554.
9. W. L. Kelly, *Org. Biomol. Chem.* **2008**, *6* (24), 4483–4493.
10. E. M. Stocking; R. M. Williams, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 3078–3115.

11. D. Uemura; T. Chou; T. Haino; A. Nagatsu; S. Fukuzawa; S. Z. Zheng; H. S. Chen, *J. Am. Chem. Soc.* **1995**, *117*, 1155–1156.
12. N. Takada; N. Umemura; K. Suenaga; T. Chou; A. Nagatsu; T. Haino; K. Yamada; D. Uemura, *Tetrahedron Lett.* **2001**, *42*, 3491–3494.
13. W. D. Celmer; G. N. Chmurny; C. E. Moppett; R. S. Ware; P. C. Watts; E. B. Whipple, *J. Am. Chem. Soc.* **1980**, *102*, 4203–4209.
14. D. E. Cane; W. Tan; W. R. Ott, *J. Am. Chem. Soc.* **1993**, *115*, 527–535.
15. J. Sporle; H. Becker; M. P. Gupta; M. Veith; V. Huch, *Tetrahedron* **1989**, *45*, 5003–5014.
16. T. Kusumi; A. Ichikawa; H. Kakisawa; M. Tsunakawa; M. Konishi; T. Oki, *J. Am. Chem. Soc.* **1991**, *113*, 8947–8948.
17. A. G. Gonzalez; M. L. Kennedy; F. M. Rodriguez; I. L. Bazzocchi; I. A. Jimenez; A. G. Ravelo; L. Moujir, *Tetrahedron* **2001**, *57*, 1283–1287.
18. X. Fu; M. L. Ferreira; F. J. Schmitz, *J. Nat. Prod.* **1999**, *62*, 1306–1310.
19. H. Oikawa; Y. Suzuki; A. Naya; K. Katayama; A. Ichihara, *J. Am. Chem. Soc.* **1994**, *116*, 3605–3606.
20. H. Oikawa; K. Katayama; Y. Suzuki; A. Ichihara, *J. Chem. Soc. Chem. Commun.* **1995**, 1321–1322.
21. W. M. Bandaranayake; J. E. Banfield; D. S. C. Black; G. D. Fallon; B. M. Gatehouse, *Aust. J. Chem.* **1981**, *34*, 1655–1667.
22. W. M. Bandaranayake; J. E. Banfield; D. S. C. Black, *J. Chem. Soc. Chem. Commun.* **1980**, 902–903.
23. K. C. Nicolaou; N. A. Petasis; R. E. Zipkin; J. Uenishi, *J. Am. Chem. Soc.* **1982**, *104*, 5555–5557.
24. R. P. Walker; D. J. Faulkner; D. Vanengen; J. Clardy, *J. Am. Chem. Soc.* **1981**, *103*, 6772–6773.
25. J. Kobayashi; M. Tsuda; T. Murayama; H. Nakamura; Y. Ohizumi; M. Ishibashi; M. Iwamura; T. Ohta; S. Nozoe, *Tetrahedron* **1990**, *46*, 5579–5586.
26. K. L. Rinehart, *Pure Appl. Chem.* **1989**, *61*, 525–528.
27. P. S. Baran; D. P. O'Malley; A. L. Zografos, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 2674–2677.
28. S. Seo; U. Sankawa; Y. Ogihara; Y. Iitaka; S. Shibata, *Tetrahedron* **1973**, *29*, 3721–3726.
29. K. C. Nicolaou; Y. H. Lim; C. D. Papageorgiou; J. L. Piper, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 7917–7921.
30. S. Yamamura; Y. Terada; Y. Chen; M. Hong; H. Hsu; K. Sasaki; Y. Hirata, *Bull. Chem. Soc. Jpn.* **1976**, *49*, 1940.
31. M. Niwa; Y. Terada; M. Nonoyama; S. Yamalnura, *Tetrahedron Lett.* **1979**, *49*, 813–816.
32. A. Nishiyama; A. Eto; Y. Terada; M. Iguchi; S. Yamamura, *Chem. Pharm. Bull.* **1983**, *31*, 2820.
33. A. C. Bazan; J. M. Edwards; U. Weiss, *Tetrahedron* **1978**, *34*, 3005–3015.
34. D. Holscher; B. Scheider, *J. Chem. Soc. Chem. Commun.* **1995**, 525–526.
35. N. Abe; T. Murata; A. Hirota, *Biosci. Biotechnol. Biochem.* **1998**, *62*, 661–666.
36. K. C. Nicolaou; G. Vassilikogiannakis; K. B. Simonsen; P. S. Baran; Y. L. Zhong; V. P. Vidali; E. N. Pitsinos; E. A. Couladouros, *J. Am. Chem. Soc.* **2000**, *122*, 3071–3079.
37. N. Abe; O. Sugimoto; K. Tanji; A. Hirota, *J. Am. Chem. Soc.* **2000**, *122*, 12606–12607.
38. N. Abe; A. Hirota, *J. Synth. Org. Chem. Jpn.* **2004**, *62*, 584–597.
39. B. N. Su; L. Yang; K. Gao; Z. J. Jia, *Planta Med.* **2000**, *66*, 281–283.
40. J. Gagnepain; F. Castet; W. Quideau, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 1533–1535.
41. C. Li; R. P. Johnson; J. A. Porco, *J. Am. Chem. Soc.* **2003**, *125*, 5095–5106.
42. J. C. Lee; G. A. Strobel; E. Lobkovsky; J. Clardy, *J. Org. Chem.* **1996**, *61*, 3232–3233.
43. J. Y. Li; G. A. Strobel, *Phytochemistry* **2001**, *57*, 261–265.
44. B. Galli; F. Gasparrini; V. Lanzotti; D. Misiti; R. Riccio; C. Villani; G. F. He; Z. W. Ma; W. F. Yin, *Tetrahedron* **1999**, *55*, 11385–11394.
45. Y. Aoyagi; Y. Takahashi; Y. Satake; H. Fukaya; K. Takeya; R. Aiyama; T. Matsuzaki; S. Hashimoto; T. Shiina; T. Kurihara, *Tetrahedron Lett.* **2005**, *46*, 7885–7887.
46. K. C. Joshi; P. Singh; S. Taneja; P. J. Cox; R. A. Howie; R. H. Thomson, *Tetrahedron* **1982**, *38*, 2703–2708.
47. B. Onegi; C. Kraft; I. Kohler; M. Freund; K. Jenett-Siems; K. Siems; G. Beyer; M. F. Melzig; U. Bienzle; E. Eich, *Phytochemistry* **2002**, *60*, 39–44.
48. J. P. Malerich; T. J. Maimone; G. I. Elliott; D. Trauner, *J. Am. Chem. Soc.* **2005**, *127*, 6276–6283.
49. A. Parvez; M. I. Choudhary; F. Akhter; M. Noorwala; F. V. Mohammad; N. M. Hasan; T. Zamir; V. U. Ahmad, *J. Org. Chem.* **1992**, *57*, 4339–4340.
50. G. Majetich; Y. Zhang, *J. Am. Chem. Soc.* **1994**, *49*, 4979–4980.
51. T. Nomura; Y. Hano, *Nat. Prod. Rep.* **1994**, *11*, 205–218.
52. T. Nomura; T. Fukai; T. Narita; S. Terada; J. Uzawa; Y. Iitaka; M. Takasugi; S. Ishikawa; S. Nagao; T. Masamune, *Tetrahedron Lett.* **1981**, *22*, 2195–2198.
53. Y. Hano; T. Nomura; S. Ueda, *J. Chem. Soc. Chem. Commun.* **1990**, 610–613.
54. T. Nomura, *Yakugaku Zasshi* **2001**, *121*, 535–556.
55. R. M. Williams, *Chem. Pharm. Bull.* **2002**, *50*, 711–740.
56. R. M. Williams; R. J. Cox, *Acc. Chem. Res.* **2003**, *36*, 127–139.
57. R. M. Williams; J. F. Sanzcervera; F. Sancenon; J. A. Marco; K. Halligan, *J. Am. Chem. Soc.* **1998**, *120*, 1090–1091.
58. T. J. Greshock; R. M. Williams, *Org. Lett.* **2007**, *9*, 4255–4258.
59. M. A. Fischbach; C. T. Walsh, *Chem. Rev.* **2006**, *106*, 3468–3496.
60. J. Schumann; C. Hertweck, *J. Biotechnol.* **2006**, *124*, 690–703.
61. H. Kobayashi; S. Meguro; T. Yoshimoto; M. Namikoshi, *Tetrahedron* **2003**, *59*, 455–459.
62. Y. Fujii; M. Asahara; M. Ichinoe; H. Nakajima, *Phytochemistry* **2002**, *60*, 703–708.
63. M. Igarashi; H. Nakamura; H. Naganawa; T. Takeuchi, *J. Antibiot.* **2000**, *53*, 1102–1107.
64. T. Agatsuma; T. Akama; S. Nara; S. Matsumiya; R. Nakai; H. Ogawa; S. Otaki; S. Ikeda; Y. Saitoh; Y. Kanda, *Org. Lett.* **2002**, *4*, 4387–4390.
65. S. B. Singh; D. L. Zink; M. A. Goetz; A. W. Dombrowski; J. D. Polishook; D. J. Hazuda, *Tetrahedron Lett.* **1998**, *39*, 2243–2246.
66. T. Ishii; K. Hayashi; T. Hida; Y. Yamamoto; Y. Nozaki, *J. Antibiot.* **2000**, *53*, 765–778.
67. S. B. Singh; D. L. Zink; B. Heimbach; O. Genilloud; A. Teran; K. C. Silverman; R. B. Lingham; P. Felock; D. J. Hazuda, *Org. Lett.* **2002**, *4*, 1123–1126.
68. N. Viswanathan; V. Balakrishnan; B. S. Joshi; W.v. Philipsborn, *Helv. Chim. Acta* **1975**, *58*, 2026–2035.

69. B. S. Joshi; N. Viswanathan; D. H. Gawad; V. Balakrishnan; W.v. Philipsborn, *Helv. Chim. Acta* **1975**, *58*, 2295–2303.
70. I. C. Parsons; A. I. Gray; T. G. Hartley; B. W. Skelton; P. G. Waterman; A. H. White, *J. Chem. Soc. Perkin Trans. 1* **1992** 645–649.
71. I. C. Parsons; A. I. Gray; T. G. Hartley; P. G. Waterman, *Phytochemistry* **1993**, *33*, 479–482.
72. R. F. C. Brown; R. Drummond; A. C. Fogerty; G. K. Hughes; J. T. Pinhey; E. Ritchie; W. C. Taylor, *Aust. J. Chem.* **1956**, *9*, 283.
73. K. Tchabanenko; R. M. Adlington; A. R. Cowley; J. E. Baldwin, *Org. Lett.* **2005**, *7*, 585–588.
74. M. Kita; M. Kondo; T. Koyama; K. Yamada; T. Matsumoto; K. H. Lee; J. T. Woo; D. Uemura, *J. Am. Chem. Soc.* **2004**, *126*, 4794–4795.
75. J. Kim; R. J. Thomson, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 3104–3106.
76. R. Sakai; T. Higa; C. W. Jefford; G. Bernardinelli, *J. Am. Chem. Soc.* **1986**, *108*, 6404–6405.
77. R. Sakai; S. Kohmoto; T. Higa; C. W. Jefford; G. Bernardinelli, *Tetrahedron Lett.* **1987**, *28*, 5493–5496.
78. K. Kondo; H. Shigemori; Y. Kikuchi; M. Ishibashi; T. Sasaki; J. Kobayashi, *J. Org. Chem.* **1992**, *57*, 2480–2483.
79. J. E. Baldwin; R. C. Whitehead, *Tetrahedron Lett.* **1992**, *33*, 2059–2062.
80. J. Kobayashi; M. Tsuda; N. Kawasaki; K. Matsumoto; T. Adachi, *Tetrahedron Lett.* **1994**, *35*, 4383–4386.
81. J. E. Baldwin; T. D. W. Claridge; A. J. Culshaw; F. A. Heupel; V. Lee; D. R. Spring; R. C. Whitehead, *Chem. A Eur. J.* **1999**, *5*, 3154–3161.
82. Y. W. Leong; L. J. Harrison; G. J. Bennett; H. T. W. Tan, *J. Chem. Res. Synop.* **1996**, 392–393.
83. Y. J. Xu; S. C. Yip; S. Kosela; E. Fitri; M. Hana; S. H. Goh; K. Y. Sim, *Org. Lett.* **2000**, *2*, 3945–3948.
84. E. J. Tisdale; I. Slobodov; E. A. Theodorakis, *Org. Biomol. Chem.* **2003**, *1*, 4418–4422.
85. K. C. Nicolaou; J. Li, *Angew. Chem. Int. Ed. Engl.* **2001**, *40*, 4264–4268.
86. A. E. Hayden; H. Xu; K. C. Nicolaou; K. N. Houk, *Org. Lett.* **2006**, *8*, 2989–2992.
87. A. J. Bartlett; J. S. Holker; E. O'Brien, *J. Chem. Soc. Chem. Commun.* **1981**, 1198–1200.
88. T. M. Tarasow; S. L. Tarasow; B. E. Eaton, *Nature* **1997**, *389*, 54–57.
89. B. Seelig; A. Jaschke, *Chem. Biol.* **1999**, *6*, 167–176.
90. A. Serganov; S. Keiper; L. Malinina; V. Tereshko; E. Skripkin; C. Hobartner; A. Polonskaia; A. T. Phan; R. Wombacher; R. Micura; Z. Dauter; A. Jaschke; D. J. Patel, *Nat. Struct. Mol. Biol.* **2005**, *12*, 218–224.
91. J. Sauer; R. Sustmann, *Angew. Chem. Int. Ed. Engl.* **1980**, *19*, 779–807.
92. D. Hilvert; K. W. Hill; K. D. Nared; M.-T. Auditor, *J. Am. Chem. Soc.* **1989**, *111*, 9261–9262.
93. A. C. Braisted; P. G. Schultz, *J. Am. Chem. Soc.* **1991**, *112*, 7430–7431.
94. C. J. Suckling; M. C. Tedford; L. M. Bence; J. I. Irvine; W. H. Stimson, *J. Chem. Soc. Perkin Trans. 1* **1993**, 1925–1929.
95. V. E. Gouverneur; K. N. Houk; B. Pascual-Teresa; B. Beno; K. D. Janda; R. A. Lerner, *Science* **1993**, *262*, 204–208.
96. J. T. Ylikauhaluoma; J. A. Ashley; C. H. Lo; L. Tucker; M. M. Wolfe; K. D. Janda, *J. Am. Chem. Soc.* **1995**, *117*, 7041–7047.
97. F. E. Romesberg; B. Spiller; P. G. Schultz; R. C. Stevens, *Science* **1998**, *279*, 1929–1933.
98. A. Heine; E. A. Stura; J. T. Ylikauhaluoma; C. S. Gao; Q. L. Deng; B. R. Beno; K. N. Houk; K. D. Janda; I. A. Wilson, *Science* **1998**, *279*, 1934–1940.
99. J. A. Xu; Q. L. Deng; J. G. Chen; K. N. Houk; J. Bartek; D. Hilvert; I. A. Wilson, *Science* **1999**, *286*, 2345–2348.
100. K. Auclair; A. Sutherland; J. Kennedy; D. J. Witter; J. P. Van den Heever; C. R. Hutchinson; J. C. Vederas, *J. Am. Chem. Soc.* **2000**, *122*, 11519–11520.
101. K. Watanabe; H. Oikawa; K. Yagi; S. Ohashi; T. Mie; A. Ichihara; M. Honma, *J. Biochem.* **2000**, *127*, 467–473.
102. K. Watanabe; T. Mie; A. Ichihara; H. Oikawa; M. Honma, *J. Biol. Chem.* **2000**, *275*, 38393–38401.
103. K. Watanabe; T. Mie; A. Ichihara; H. Oikawa; M. Honma, *Tetrahedron Lett.* **2000**, *41*, 1443–1446.
104. T. Ose; K. Watanabe; T. Mie; M. Honma; H. Watanabe; M. Yao; H. Oikawa; I. Tanaka, *Nature* **2003**, *422*, 185–189.
105. A. Ichihara; H. Tazaki; S. Sakamura, *Tetrahedron Lett.* **1983**, *24*, 5373–5376.
106. A. Ichihara; M. Miki; S. Sakamura, *Tetrahedron Lett.* **1985**, *26*, 2453–2454.
107. H. Oikawa; T. Yokota; A. Ichihara; S. Sakamura, *J. Chem. Soc. Chem. Commun.* **1989**, 1284–1285.
108. H. Oikawa; T. Yokota; C. Sakano; Y. Suzuki; A. Naya; A. Ichihara, *Biosci. Biotechnol. Biochem.* **1998**, *62*, 2016–2022.
109. S. S. Alam; J. M. Bilton; M. Z. Slawin; D. J. Williams; R. N. Sheppard; R. M. Strange, *Phytochemistry* **1989**, *28*, 2627–2630.
110. K. M. Jenkins; S. G. Toske; P. R. Jensen; W. Fenical, *Phytochemistry* **1998**, *49*, 2299–2304.
111. A. Ichihara; M. Miki; H. Tazaki; S. Sakamura, *Tetrahedron Lett.* **1987**, *28*, 1175–1178.
112. H. Oikawa; T. Kobayashi; K. Katayama; Y. Suzuki; A. Ichihara, *J. Org. Chem.* **1998**, *63*, 8748–8756.
113. H. Oikawa; Y. Suzuki; K. Katayama; A. Naya; C. Sakano; A. Ichihara, *J. Chem. Soc. Perkin Trans. 1* **1999**, 1225–1232.
114. K. Katayama; T. Kobayashi; H. Oikawa; M. Honma; A. Ichihara, *Biochim. Biophys. Acta* **1998**, *1384*, 387–395.
115. K. Katayama; T. Kobayashi; M. Chijimatsu; A. Ichihara; H. Oikawa, *Biosci. Biotechnol. Biochem.* **2008**, *72*, 604–607.
116. L. Raimondi; F. K. Brown; J. Gonzalez; K. N. Houk, *J. Am. Chem. Soc.* **1992**, *114*, 4796–4804.
117. R. Breslow, *Acc. Chem. Res.* **1991**, *24*, 159–164.
118. P. A. Grieco, *Aldrichim. Acta* **1991**, *24*, 59–66.
119. M. F. Ruiz-López; X. Assfeld; J. I. Garcia; J. A. Mayoral; L. Salvatella, *J. Am. Chem. Soc.* **1993**, *115*, 8780–8787.
120. K. Kasahara; T. Miyamoto; T. Ujimoto; H. Oguri; T. Tokiwano; H. Oikawa; Y. Ebizuka; I. Fujii, Symposium Papers. In *49th Symposium on the Chemistry of Natural Products*, Sapporo, 2007; pp 19–24.
121. R. N. Moore; G. Bigam; J. K. Chan; A. M. Hogg; T. T. Nakashima; J. C. Vederas, *J. Am. Chem. Soc.* **1985**, *107*, 3694–3701.
122. Y. Yoshizawa; D. J. Witter; Y. Liu; J. C. Vederas, *J. Am. Chem. Soc.* **1994**, *116*, 2693–2694.
123. J. L. Sorensen; K. Auclair; J. Kennedy; C. R. Hutchinson; J. C. Vederas, *Org. Biol. Chem.* **2003**, *1*, 50–59.
124. D. J. Witter; J. C. Vederas, *J. Org. Chem.* **1996**, *61*, 2613–2623.
125. J. Kennedy; K. Auclair; S. G. Kendrew; C. Park; J. C. Vederas; C. R. Hutchinson, *Science* **1999**, *284*, 1368–1372.
126. I. Sakurai; H. Miyajima; K. Akiyama; S. Shimizu; Y. Yamamoto, *Chem. Pharm. Bull.* **1988**, *36*, 2003–2011.
127. H. Oikawa; K. Watanabe; K. Yagi; S. Ohashi; T. Mie; A. Ichihara; M. Honma, *Tetrahedron Lett.* **1999**, *40*, 6983–6986.
128. H. Oikawa; K. Yagi; K. Watanabe; M. Honma; A. Ichihara, *J. Chem. Soc. Chem. Commun.* **1997**, 97–98.
129. J. A. Piccirilli; J. D. Rozzell; S. A. J. Benner, *J. Am. Chem. Soc.* **1987**, *109*, 8084–8085.
130. K. Watanabe; T. Mie; A. Ichihara; H. Oikawa; M. Honma, *Biosci. Biotechnol. Biochem.* **2000**, *64*, 530–538.
131. J. M. Serafimov; H. C. Lehmann; H. Oikawa; D. Hilvert, *Chem. Commun.* **2007**, 1701–1703.
132. J. M. Serafimov; T. Westfeld; B. H. Meier; D. Hilvert, *J. Am. Chem. Soc.* **2007**, *129*, 9580–9581.

133. C. R. W. Guimaraes; M. Udier-Blagovic; W. L. Jorgensen, *J. Am. Chem. Soc.* **2005**, *127*, 3577–3588.
134. S. Sekita; K. Yoshihira; S. Natori; H. Kuwano, *Tetrahedron Lett.* **1973**, 2109–2122.
135. A. Probst; C. Tamm, *Helv. Chim. Acta* **1981**, *64*, 2065–2077.
136. S. Sekita; K. Yoshihira; S. Natori, *Chem. Pharm. Bull.* **1983**, *31*, 490–498.
137. H. Oikawa; Y. Murakami; A. Ichihara, *J. Chem. Soc. Perkin Trans. 1* **1992**, 2955–2959.
138. H. Oikawa; Y. Murakami; A. Ichihara, *J. Chem. Soc. Perkin Trans. 1* **1992**, 2949–2954.
139. J. Schumann; C. Hertweck, *J. Am. Chem. Soc.* **2007**, *129*, 9564–9565.
140. E. J. Thomas; J. W. F. Whitehead, *J. Chem. Soc. Perkin Trans. 1* **1989**, 499–505.
141. L. T. Burke; D. J. Dixon; S. V. Ley; F. Rodriguez, *Org. Lett.* **2000**, *2*, 3611–3613.
142. J. W. Sims; J. P. Fillmore; D. D. Warner; E. W. Schmidt, *Chem. Commun.* **2005**, 186–188.
143. R. Muntwyler; W. Keller-Schierlein, *Helv. Chim. Acta* **1972**, *55*, 2071.
144. X. Y. Jia; Z. H. Tian; L. Shao; X. D. Qu; Q. F. Zhao; J. Tang; G. L. Tang; W. Liu, *Chem. Biol.* **2006**, *13*, 575–585.
145. W. R. Roush; R. J. Sciotti, *J. Am. Chem. Soc.* **1998**, *120*, 7411–7419.
146. H. Zhang; J. A. White-Phillip; C. E. Melancon; H. J. Kwon; W. L. Yu; H. W. Liu, *J. Am. Chem. Soc.* **2007**, *129*, 14670–14683.
147. H. A. Kirst; K. H. Michel; J. W. Martin; L. C. Creemer; E. H. Chio; R. C. Yao; W. M. Nakatsukasa; L. Boeck; J. L. Occolowitz; J. W. Paschal; J. B. Deeter; N. D. Jones; G. D. Thompson, *Tetrahedron Lett.* **1991**, *32*, 4839–4842.
148. C. Waldron; P. Matsushima; P. R. Rosteck; M. C. Broughton; J. Turner; K. Madduri; K. P. Crawford; D. J. Merlo; R. H. Baltz, *Chem. Biol.* **2001**, *8*, 487–499.
149. H. J. Kim; R. Pongdee; Q. Q. Wu; L. Hong; H. W. Liu, *J. Am. Chem. Soc.* **2007**, *129*, 14582–14585.
150. L. B. Davin; H. B. Wang; A. L. Crowell; D. L. Bedgar; D. M. Martin; S. Sarkanen; N. G. Lewis, *Science* **1997**, *275*, 362–366.

Biographical Sketch

Hideaki Oikawa was born in 1956. He obtained his Ph.D. from the Hokkaido University in 1984 under the guidance of Professor S. Sakamura). Dr. Oikawa was a postdoctoral fellow in the lab of Professor D. E. Cane at the Brown University in 1984 and with Dr. K. Isono at RIKEN in 1985. He was an assistant professor in the Department of Agricultural Chemistry, Hokkaido University (1986–99); and an associate professor (1999–2003). He later became a professor at the Division of Chemistry and Graduate School of Science at the Hokkaido University (2003–present). Professor Oikawa received an award for the encouragement of young scientists from the Japan Society of Bioscience, Biotechnology, and Agrochemistry in 1993; the Chemical Society of Japan Award for Creative Work in 2004: Research field, engineering of biosynthesis of natural products, combinatorial biochemistry, reaction mechanism on C–C bond forming enzymes.

8.09 Phosphoryl and Sulfuryl Transfer

**Tiago A. S. Brandao and Alvan C. Hengge**, Utah State University, Logan, UT, USA

8.09.1 Introduction

8.09.1.1 Summary of Biological Roles and Properties of Phosphate and Sulfate Esters

Phosphate esters and anhydrides dominate the living world[1] with crucial roles in genetic materials, coenzymes, and energy reservoirs, and as intermediates in biochemical transformations. Protein phosphorylation levels regulate a host of processes including cell division, differentiation and development, metabolism, learning, and memory. It has been estimated that one-third of all proteins in organisms undergo reversible phosphorylation, often at multiple sites.[2] The important roles of sulfate esters in biology have been less appreciated historically, but are increasingly recognized. The sulfation of a number of biological signaling molecules, including hormones, neurotransmitters, peptides, and proteins, alters their biological activity. Sulfation also serves to solubilize molecules to aid in their excretion, and sulfate monoesters are also found among many classes of natural products. This review summarizes the chemistry of nonenzymatic phosphoryl and sulfuryl transfer reactions, and discusses the kinetic and chemical mechanisms of some of the enzymes that catalyze these reactions.

Phosphoric and sulfuric acid derivatives possess crucial properties that allow them to uniquely fill their many roles in biochemistry. Phosphoric acid may be esterified to form a monoester, diester, or triester (**Figure 1**). Sulfuric acid may be esterified at one or two positions, to form a monoester or a diester. Sulfate diesters are highly reactive, and have not been found in nature; nor do phosphate triesters occur naturally. The hydrolysis of both phosphate and sulfate esters are thermodynamically favorable, but nucleophiles are repelled by the negative charge of the ionized forms. The resulting kinetic stability of phosphate monoesters and diesters, and of sulfate monoesters, is a major factor in their suitability for biological roles. For example, the half-life for hydrolysis of alkyl phosphate dianions by water is approximately 1.1×10^{12} years ($k = 2 \times 10^{-20}\,\mathrm{s}^{-1}$) at 25 °C.[3] Such species
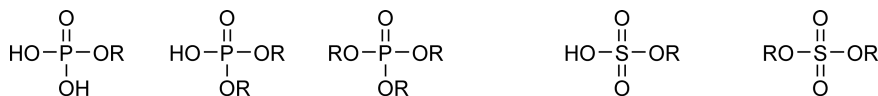
Figure 1   On the left are the structures of a phosphate monoester, diester, and triester. The first pK_a of an alkyl phosphate monoester, and the pK_a of a dialkyl diester, are typically between 1 and 2. The second pK_a of an alkyl monoester is approximately 6.8, and from one to two units lower for aryl esters. At the right are the structures of a sulfate monoester and diester. Sulfate monoesters have a very low pK_a values and as a result are always ionized at physiological pH.

are substrates for phosphatases involved in signal transduction and regulation, which produce the largest enzymatic rate enhancements ($\sim 10^{21}$-fold) that have yet been identified relative to the uncatalyzed reactions of their substrates. Analogous experiments yield estimated rate enhancements of approximately 10^{11}-fold by human alkyl sulfatases.[4]

## 8.09.1.2   Mechanistic Possibilities for Phosphoryl and Sulfuryl Transfer

Phosphoryl and sulfuryl transfer reactions are nucleophilic substitution reactions, and three limiting mechanisms can be envisioned. One is a dissociative S_N1-type mechanism, designated $D_N + A_N$ in the IUPAC nomenclature,[5] in which a discrete intermediate forms that is subsequently attacked by a nucleophile. The intermediate would be metaphosphate (PO_3^-) from a phosphate monoester and SO_3 from a sulfate monoester. Another potential stepwise transfer mechanism is an associative, two-step process ($A_N + D_N$). Here, the nucleophile adds to form a trigonal bipyramidal intermediate, called a phosphorane in the phosphate ester case, which collapses with leaving-group departure in a second step. This mechanism has been best characterized for phosphate esters, where it has been documented in some reactions of phosphate triesters and diesters, and speculated to occur in enzymatic reactions. Nucleophiles add, and leaving groups depart, from the apical positions of the trigonal bipyramid (**Scheme 1**). For this reason, a concerted (single-step) mechanism requires the nucleophile to approach from a direction directly opposite the incipient leaving group in the trigonal bipyramidal transition state; this is referred to as an in-line mechanism. If an intermediate forms, the apical and equatorial ligands of the resulting phosphorane can undergo rearrangement by a process known as pseudorotation[6,7] as shown in **Scheme 1**. In an $A_N + D_N$ mechanism, if the leaving group is initially in an equatorial position, pseudorotation to bring the leaving group into an apical position will result in a final product with net retention of stereochemistry if the reactant is chiral. Inversion of stereochemistry does not require that a reaction is concerted, but it does impose the condition that if a phosphorane intermediate forms, leaving-group departure must occur before pseudorotation can take place. This carries the requirement that the leaving group must initially reside in an apical position. Further discussions of pseudorotation can be found in other reviews, and references therein.[8-10]

The final mechanistic possibility is a concerted mechanism (A_ND_N) with no intermediate. In this mechanism bond formation to the nucleophile and bond fission to the leaving group occurs in the same step. The transition state could be loose or tight, depending upon the synchronicity between the nucleophilic attack and the leaving-group departure. In this review, a transition state is defined as loose or tight depending on the sum of the bond orders to the nucleophile and the leaving group. In any substitution reaction, this sum is unity in the reactant and in the product. In an addition intermediate this bond order sum is two. A tight transition state is one in which bond formation to the nucleophile is more advanced than leaving-group bond fission, giving a sum of >1. In a loose
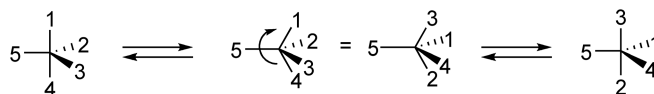
Scheme 1   On the left is a trigonal bipyramidal structure with apical ligands (1 and 4) and equatorial ligands (2, 3, and 5). During pseudorotation, a pair of equatorial ligands exchange with the apical ligands in a concerted fashion through the intermediacy of a tetragonal–pyramidal transition state. The pivot point is one of the equatorial ligands (in the example above). This process can be visualized as one in which the two apical ligands (1 and 4) undergo a motion where their bond angles reduce from 180 to 120°, and the two equatorial ligands (2 and 3) open their bond angles from 120 to 180°.
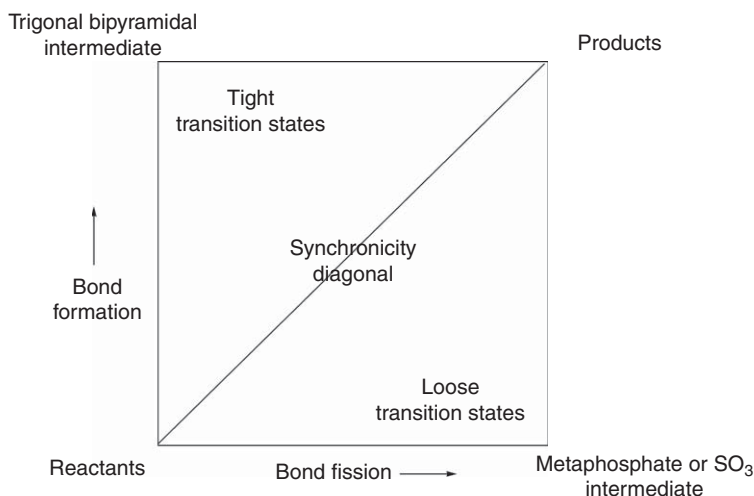
Figure 2   A loose transition state for phosphoryl or sulfuryl transfer is one in which bond fission is ahead of bond formation to the nucleophile, and resides in the lower right region of the More–O'Ferrall Jencks diagram. A tight transition state is the reverse situation, residing in the upper left region. If the sum of bond order to nucleophile plus leaving group is unity, the transition state will lie on the synchronicity diagonal.
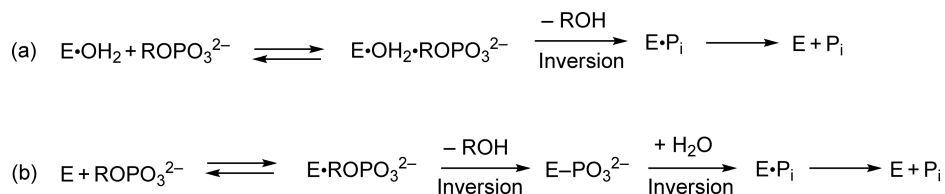
transition state, this sum is <1. This description is conveyed pictorially in a More–O'Ferrall Jencks diagram[11–13] (**Figure 2**).

Physical organic chemists have developed a number of mechanistic tools based on kinetics and stereochemistry to study the reaction mechanisms and to discern transition state characteristics. These methods have been used to study both uncatalyzed and enzyme-catalyzed reactions of phosphate and sulfate esters. Linear free energy relationships (LFERs) are measurements of the dependency of the rate of reactions on electronic characteristics of the nucleophile and/or the leaving group. In a Hammett LFER, the logarithm of the rate constant is plotted against the sigma constant associated with an electron-donating or electron-withdrawing substituent. In a Brønsted plot, the log of the rate constant is plotted against the pK_a of the nucleophile or leaving group to obtain the resulting slopes (β_{nuc} or β_{lg}, respectively). The magnitude of the slope reflects the amount of change in charge between the reactant and the transition state. A large negative β_{lg} is indicative of a transition state in which the bond to the leaving group is largely broken and a significant negative charge is developed on this group. Similarly, a large dependency on nucleophile basicity (β_{nuc}) implies a transition state with significant bond formation to the nucleophile.[13]

Kinetic isotope effects (KIEs) are also reporters of changes in bonding between the reactant and the transition state.[14,15] In an isotope effect experiment, the reactant is labeled with a light and a heavy isotope at a position of interest, often where bond formation or bond fission occurs during the reaction. The isotope effect is the ratio of the rate constants for the light isotopic isomer over that of the heavy isotope. Normal isotope effects result when bond weakening to the labeled atom occurs in the rate-limiting step, while inverse isotope effects result from bond formation to the labeled atom, and the magnitudes of KIEs can give a measure of the extent of bond fission or bond formation at the transition state. Isotope effects have been measured for a number of phosphoryl and sulfuryl transfer reactions, both enzymatic and uncatalyzed.[16] Results from studies on specific enzymes are discussed in the sections that follow.

Activation parameters such as entropy or volume of activation, and to a lesser degree the enthalpy of activation, also provide mechanistic information.[17] Entropies and enthalpies of activation are obtained from measurements of the rate constant for a reaction over a range of temperatures. Whether the rate-determining step of a reaction is unimolecular or bimolecular can have a significant effect on these parameters.

All of the methods described above are subject to limitations in their interpretation, particularly in enzymatic reactions if substrate binding is irreversible, or when a chemical step is not rate limiting. It is inadvisable to draw mechanistic conclusions from the results of any single method, but when used in combination, and sometimes

(a) $E \cdot OH_2 + ROPO_3^{2-}$ $\rightleftharpoons$ $E \cdot OH_2 \cdot ROPO_3^{2-}$ $\xrightarrow[\text{Inversion}]{- ROH}$ $E \cdot P_i$ $\longrightarrow$ $E + P_i$

(b) $E + ROPO_3^{2-}$ $\rightleftharpoons$ $E \cdot ROPO_3^{2-}$ $\xrightarrow[\text{Inversion}]{- ROH}$ $E - PO_3^{2-}$ $\xrightarrow[\text{Inversion}]{+ H_2O}$ $E \cdot P_i$ $\longrightarrow$ $E + P_i$

Scheme 2   Two potential reaction mechanisms for phosphatases or sulfatases are shown here using a phosphate ester. In (a), the phosphoryl group is transferred directly to a water molecule, which is typically bound to one or two metal ions; if the substrate is made chiral at phosphorus, the stereochemical outcome is inversion. In (b), the phosphoryl group is first transferred to an enzymatic nucleophile; $E - PO_3^{2-}$ is a covalent phosphoenzyme intermediate. In a subsequent step, this intermediate is hydrolyzed. Since each step occurs with inversion of configuration at phosphorus, the net outcome is retention. The same principles apply to sulfuryl transfer. P_i = inorganic phosphate.

with the assistance of computational studies, the tools of physical organic chemistry have proven very valuable in the decipherment of mechanisms and in the identification of transition states.[18]

   Stereochemistry is another powerful tool for determining the net reaction pathway of phosphatases and sulfatases. These enzymes catalyze the net transfer of a phosphoryl or sulfuryl group to water from a monoester, producing inorganic phosphate or sulfate. Inversion results when the reaction occurs in a single step (**Scheme 2**, pathway a). Phosphatases that transfer the phosphoryl group directly to water with inversion typically possess a binuclear metal center and the nucleophile is a metal-coordinated hydroxide. Examples of phosphatases that follow this mechanism are the purple acid phosphatases (PAPs) and the serine/threonine phosphatases (described in Sections 8.09.4.3 and 8.09.4.4.1). Net retention of stereochemistry occurs when a phosphorylated or sulfurylated enzyme intermediate is on the catalytic pathway, which is hydrolyzed by the nucleophilic addition of water in a subsequent step (**Scheme 2**, pathway b).

   Figure 3 shows how a phosphate or a sulfate monoester can be rendered chiral. The phosphoryl or sulfuryl group can be made chiral using oxygen isotopes ^{16}O, ^{17}O, and ^{18}O. A chiral phosphorothioate has one sulfur atom and two isotopes of oxygen in the nonbridging positions, and is sometimes used as a surrogate for a chiral phosphate monoester. If another nucleophile can be substituted for water in the first two examples, the product of the reaction will also be a chiral monoester, permitting the stereochemical outcome of the reaction to be determined. A number of phosphatases, and some sulfatases, have been analyzed by the stereochemical methods using chiral substrates. Summaries of stereochemical results are available in several reviews.[14,19–21]

8.09.2   Uncatalyzed Reactions of Phosphate and Sulfate Monoesters

8.09.2.1   Phosphate Monoester Dianions and Monoanions

A phosphate monoester may be neutral, a monoanion, or a dianion depending on pH. The neutral form is present only under very acidic conditions, and the reactions of this form have been subjected to less study than monoanions and dianions. Dianions are much less reactive than monoanions, except for monoesters with highly activated leaving groups, such as 2,4-dinitrophenyl phosphate.[22]
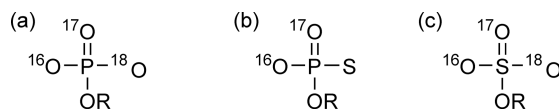
Figure 3   Chiral phosphate and sulfate esters. Negative charges have been omitted from the nonbridging oxygen atoms for clarity; under alkaline conditions, all three nonbridging oxygen atoms will have equivalent bonds due to resonance. A chiral phosphate monoester is shown in (a), with the three isotopes of oxygen in the nonbridging positions. A chiral phosphorothioate is shown in (b). This surrogate for phosphate has the advantage of yielding a chiral product when the ester group is replaced by water, if the water oxygen atom has a different oxygen isotope from the two already present. Structure (c) shows a chiral sulfate monoester.
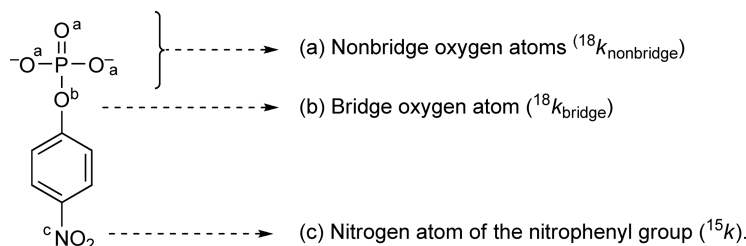
8.09.2.1.1   Reactions of monoester dianions

The compiled[21,23] experimental evidence in the form of LFER data, activation parameters, KIEs, and stereochemical results indicates that these esters undergo phosphoryl transfer through a concerted reaction with a loose transition state. The rate of reaction is little affected by nucleophile basicity but has a large dependency on the leaving group pK_a (reflected in a large negative Brønsted $\beta_{lg} = -1.2$)[22] and together with large $^{18}k_{bridge}$ isotope effects[24,25] are indicative of a transition state with significant bond fission to the leaving group and little bond formation to the nucleophile. Despite very loose transition states, there is no evidence for the formation of free metaphosphate in aqueous solution.[26] Stereochemical experiments, from a reactant in which the phosphoryl group is chiral by means of the isotopes ^{16}O, ^{17}O, and ^{18}O, demonstrate that phosphoryl transfer occurs with the inversion of configuration. In contrast, racemic t-butyl phosphate forms in the reaction of chiral p-nitrophenyl phosphate (pNPP) when t-butanol is the solvent and phosphoryl acceptor.[27,28] The latter reaction also exhibits a significantly more positive entropy of activation ($\Delta S^{\ddagger}$) of $+24$ eu,[29] compared to $+0.3$ eu[30] for the aqueous hydrolysis. The stereochemical and entropy results are consistent with a two-step $D_N + A_N$ mechanism with a metaphosphate intermediate for the reaction in t-butanol. There is little transition state difference between the aqueous hydrolysis and the reaction in t-butanol other than the degree of nucleophilic participation in the rate-limiting step. The β_{lg} value ($\beta_{lg} = -1.2$) is similar for the two reactions, and the KIEs in the nonbridging phosphoryl oxygen atom ($^{18}k_{nonbridge}$), the bridging oxygen atom ($^{18}k_{bridge}$), and in the nitrogen atom (^{15}k), corrected for the difference in temperature at which the KIEs were measured, are also similar[24] (**Table 1** and **Figure 4**). These results mean that the only difference between the transition states of the two reactions is that the small nucleophilic involvement by water in the aqueous hydrolysis is absent in the rate-limiting step of the butanolysis.

The data and mechanistic conclusions summarized above come from work with aryl phosphomonoesters. The large negative β_{lg} leads to the expectation of very slow rates for the hydrolysis of alkyl ester dianions. This has been confirmed by an investigation of the hydrolysis of methyl phosphate, that found the rate of the dianion reaction to be below the threshold of detectability, with an estimated rate constant of 2×10^{-20} s^{-1} at 25 °C.[3] This value is close to the rate predicted from an extrapolation of the Brønsted plot of aryl phosphomonoester dianions, suggesting that the alkyl and aryl esters likely follow a similar hydrolysis mechanism.

A number of groups have used computational methods to investigate phosphate ester hydrolysis, and have reached disparate conclusions, depending on the method used.[31–41] A review of the computational literature on this topic is beyond the scope of this article; however, a majority of the investigations of dianion hydrolysis have come to conclusions consistent with those from experiments, of a loose transition state in a concerted reaction.

Table 1 Data for phosphoryl transfer reactions from pNPP to water and to t-butanol

Reaction	β_{lg}	$\Delta S^{\ddagger}$	^{15}k	$^{18}k_{bridge}$	$^{18}k_{nonbridge}$
Aqueous hydrolysis (95 °C)	−1.2	+0.3 eu	1.0028 ± 0.0002	1.0189 ± 0.0005	0.9994 ± 0.0005
Reaction in t-butanol (30 °C)	−1.2	+24 eu	1.0039 ± 0.0003	1.0202 ± 0.0008	0.9997 ± 0.0016

(a) Nonbridge oxygen atoms ($^{18}k_{nonbridge}$)

(b) Bridge oxygen atom ($^{18}k_{bridge}$)

(c) Nitrogen atom of the nitrophenyl group (^{15}k).

Figure 4   Diagram of pNPP showing positions of KIE measurement for reactions in water and in t-butanol. KIEs are designated by a leading superscript showing the heavier isotope; thus, $^{18}k_{bridge}$ represents the oxygen-18 isotope effect in the bridging oxygen position.

8.09.2.1.2   *Reactions of monoester monoanions*

The uncatalyzed hydrolysis reactions of phosphomonoesters are fastest around pH 4, where concentration of the monoanion is highest. The monoanion undergoes more facile hydrolysis by virtue of the beneficial effect of leaving-group protonation, which is accomplished by a proton transfer from the phosphoryl group, either directly or through an intervening water molecule. Neutralization of the leaving group in the transition state is reflected in the less negative β_{lg} of -0.27,[22] indicating only modest difference in effective charge on the leaving group from the ground state to the transition state. Normal nonbridge-^{18}O KIEs on the hydrolysis of the monoanions of *p*-NPP[24] (1.0184 ± 0.0005) and *m*-nitrobenzyl phosphate (*m*NBP)[42] (1.0151 ± 0.0002) are very close to previously measured[43] [18]O equilibrium isotope effects on the deprotonation of phosphate esters, consistent with proton transfer from the phosphoryl group during the reaction. The precise timing of proton transfer and leaving-group departure depends on the ester; it has been proposed[22] that protonation is concerted with leaving-group departure for less basic leaving groups, while for more basic leaving groups, proton transfer occurs first followed by rate-limiting P–O bond fission. Variations in the magnitudes of deuterium solvent isotope effects are consistent with this hypothesis.[22,42,44]

Similar to the hydrolysis of dianions, monoanion hydrolysis is thought to be concerted, without the intermediacy of a metaphosphate intermediate.[45] A stereochemical study carried out on the monoanion of phenyl phosphate found inversion of stereochemistry, ruling out the existence of a free metaphosphate intermediate.[27]

8.09.2.2   **Sulfate Monoester Anions**

Sulfuryl transfer has available the same mechanistic pathways previously discussed for phosphoryl transfer (**Figure 2**). If a monoester follows a fully dissociative ($D_N + A_N$) mechanism, a sulfur trioxide intermediate would form, analogous to metaphosphate in the phosphoryl system. An addition–elimination ($A_N + D_N$) mechanism would form a pentacoordinate sulfurane intermediate.

Experimental evidence indicates that the reactions of aryl sulfate monoesters are mechanistically similar to those of phosphate monoesters. Hydrolysis is faster under strongly acidic or basic conditions, with a broad pH-independent region between pH 4 and pH 12.[46,47] Tracer studies with [18]O have shown that aryl sulfate monoesters undergo hydrolysis by S–O bond fission in the pH-independent reaction.[46] LFERs[46–49] and KIEs in the sulfuryl group and in the leaving group[50] are similar to those for reactions of phosphate monoester dianions, indicating that they react by a similar mechanism with a loose transition state in which the sulfuryl group resembles SO_3. Sulfur isotope effects on the hydrolysis of *p*-nitrophenyl sulfate (*p*NPS) are also consistent with such a mechanism.[51] Stereochemical studies of sulfuryl transfer from phenyl [(*R*)-^{16}O, [17]O, ^{18}O] sulfate to an acceptor alcohol show that the reaction proceeds with inversion, ruling out the existence of a free SO_3 intermediate.[52]

The reactive species under acidic conditions is the neutral ester. This reaction is believed to proceed by transfer of the proton from the sulfuryl group to the leaving group,[53] as in reactions of phosphate monoester monoanions. A reduced value for β_{lg}[47] and a solvent deuterium isotope effect of 2.43[54] are consistent with proton transfer to the leaving group in the transition state. The intermediacy of free SO_3 in the acid hydrolysis is sometimes assumed, but has not been proven.

Under alkaline conditions (pH ≥ 13), where the rate of hydrolysis is linear with [HO$^-$], both aryl and alkyl sulfate esters undergo hydrolysis through C–O bond cleavage.[55,56] In the hydrolysis of SDS (sodium dodecyl sulfate) under neutral conditions, the pH-independent reaction proceeds by water attack at sulfur, evidenced by the absence of labeled dodecanol from reactions carried out in $H_2^{18}O$.[57] However, the reactions of other alkyl sulfate monoesters in the pH-independent region differ from SDS, and from their aryl counterparts, in that attack at the carbon with C–O bond fission is the dominant pathway.[4,58] Surprisingly, methyl sulfate is an efficient methyl donor to oxygen and nitrogen nucleophiles, accomplishing such reactions more rapidly than alkyl sulfonium ions such as *S*-adenosylmethionine, the usual methyl donor in biochemistry.[4]

8.09.3   Key Features of Enzymatic Catalysis, and Structural Similarities between Sulfatases and Phosphatases

The potential catalytic interactions that enzymes might use to promote phosphoryl or sulfuryl transfer are common to hydrolytic reactions in general. One source of catalysis is a means for neutralization of the leaving group, either through protonation by a general acid or coordination to a metal ion. General base assistance for the deprotonation of a nucleophile is another common characteristic of many enzymes, including phosphatases and sulfatases. A common feature of metallohydrolase catalysis is metal ion coordination of the nucleophile, whether it be an enzymatic amino acid side chain or a water molecule, facilitating its deprotonation to form the more nucleophilic alkoxide or hydroxide.

The general applicability of these interactions to catalysis probably accounts for the observation of phosphatase activity in a large number of enzyme families. A search of the Protein Families database[59] for phosphatases yields more than 40 enzyme families with reported phosphatase activity (see http://pfam.sanger.ac.uk/search/keyword?-query=phosphatases). A large group of enzymes known as the haloalkanoic acid dehalogenase superfamily (HADSF) is dominated by phosphotransferases.[60] The majority of these enzymes are hydrolases (phosphatases and ATPases) and a smaller number are hexose phosphate mutases, including phosphoglucomutase (PGM), described in Section 8.09.4.5. The HADSF phosphotransferases share a core domain that binds a Mg^{2+} ion and the phosphate ester reactant in an orientation that allows attack by an Asp nucleophile, with general acid–base catalysis by a second Asp positioned nearby. The aspartylphosphate intermediate undergoes phosphoryl transfer to a water molecule (in the phosphatases) or to a sugar hydroxyl group (in the mutases).

The uncatalyzed reactions of sulfate and phosphate monoesters proceed by essentially identical mechanisms and transition states; the only difference is in the charge of the transferring group. Comparisons of the structures of enzymes that catalyze phosphoryl and sulfuryl transfer show that the primary similarity at the active site is the presence of positive charge, either in the form of metal ion(s) and/or positively charged amino acid side chains. Given the mechanistic similarity of the reactions, it seems reasonable that an active site with complementarity to the transition state of the reaction of either a phosphate or a sulfate monoester should be endowed with some degree of complementarity for the other. However, catalytic promiscuity is the exception rather than the rule, and when it is found, significant preference for the natural substrate is common.

Of the known examples of crossover activity, that of alkaline phosphatase (AP) is the most studied.[61] Several sulfatases, including arylsulfatases A and B (ASA and ASB) and PAS from *Pseudomonas aeruginosa*,[62] share the common core of the AP superfamily enzymes.[63] Comparison of the *Escherichia coli* AP and human ASB structures shows that the secondary structure homology consists of a central arrangement formed by eight well-aligned β-sheets flanked on both sides by five α-helices (**Figure 5(a)**). The active-site residues involved in the binding of the metals, Zn_{II} in AP and Ca in ASB, are held in structural conserved loops emerging from the central β-sheets (**Figure 5(b)**). The structures show close superposition of sources of positive charge (**Figure 5**). One of the two Zn ions of AP aligns with a calcium ion in AS, while the second Zn ion aligns closely with Lys318. Despite these structural similarities, AP catalyzes the hydrolysis of *p*-NPP approximately 10^9-fold more efficiently than *p*NPS.[61]

Other examples of crossover activity are known. ASA has been shown to possess cyclic phosphodiesterase activity.[67] Adenylate kinase transfers the γ-phosphoryl group of ATP, but can also accept as a substrate the sulfuryl analogue γ-sulfuryl-ADP.[68] Sulfoenolpyruvate, the sulfate analogue of phosphoenolpyruvate, is a substrate for pyruvate kinase, producing pyruvate and adenosine-5′-sulfatopyrophosphate.[69] Many other examples probably remain undiscovered.

Lewis acid (electrophilic) activation of the transferring group toward nucleophilic attack by interactions with hydrogen bond donors or metal ions is intuitively more likely in an associative mechanism rather than with a dissociative transition state. However, electrostatic interactions between the nonbridging oxygens can also stabilize a loose, metaphosphate-like or sulfur trioxide-like transition state, if a conformational change of the protein or the geometric change from tetrahedral to trigonal bipyramidal in the transition state enhances these interactions. For phosphatases, which catalyze the hydrolysis of phosphomonoesters, it has been suggested that metal ions or cationic side-chain sites might change the normally loose transition state into a more associative process by promoting electron withdrawal from the phosphorus atom, thus promoting nucleophilic attack.[70,71]

(a)

(b)
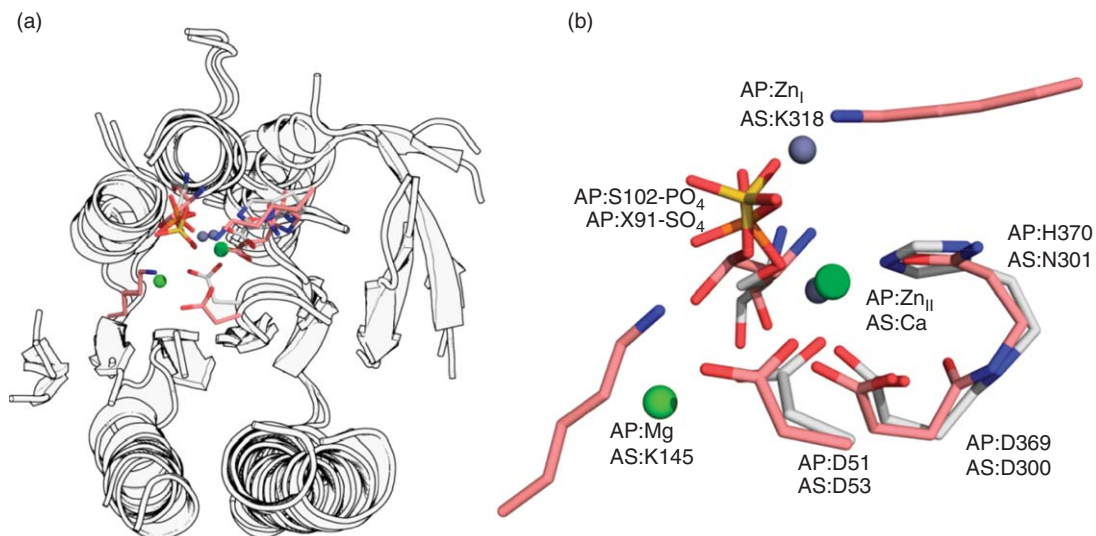
Figure 5 Superposition of homologous *Escherichia coli* AP (AP, in cyan) and human ASB (AS, in gray). (a) Conserved structural core of the AP superfamily enzymes. (b) Alignment for selected active-site residues: the position of the phosphorylated serine of AP (AP:S102-PO$_4$) coincides with that of the sulfurylated FGly of AS (AS:X91-SO$_4$), Zn$_{II}$ (AP) with Ca (AS), Zn$_I$ (AP) with K318Nξ (AS), Mg (AP) with K145Nξ (AS), and the metal ligand H370 (AP) with the metal ligand N301 (AS). The alignment was generated by DALI server[64] using the X-ray structures PDB codes 1HJK[65] and 1FSU[66] for H331Q AP and AS, respectively, and the representations were drawn in PyMol.

This would be analogous to the change that results from alkyl substitution; that is, transition states become more associative in the continuum from monoesters to triesters. Although relatively few phosphatases have been subjected to serious scrutiny of their transition states, in the cases that have been reported, this prediction has not been borne out. The reactions catalyzed by AP proceeds through loose transition states that are not significantly altered from those in solution, both in its phosphatase and in its promiscuous sulfatase activities.[72–74] Results with λ-phosphatase[75] and with calcineurin,[76] which both catalyze phosphoryl transfer to a metal-coordinated hydroxide nucleophile, also provide no evidence of a significantly different transition state. Protein tyrosine phosphatases (PTPs), which do not contain metal ion cofactors but have a conserved arginine residue and proceed via a phosphocysteine intermediate, similarly catalyze phosphoryl transfer via a transition state similar to the one in solution.[16]

8.09.4   Enzymes that Catalyze the Hydrolysis of Phosphate Monoesters

All phosphatases catalyze the same net reaction, the hydrolysis of a phosphate monoester to inorganic phosphate and the alcohol or phenol from the ester group. As already mentioned, despite the thermodynamic favorability of this reaction, the kinetic barrier is formidable. A number of the enzymes that catalyze this reaction have been characterized. Phosphatases vary in their preference for the charge state of the substrate (either the monoanion or the dianion), in the presence or absence of a metal center, and in the utilization of a phosphoenzyme intermediate versus direct attack by water. Even among metallophosphatases, there are variations in the means by which the dinuclear metal center participates in binding and catalysis.

8.09.4.1   Alkaline Phosphatases

APs are nonspecific metalloenzymes that hydrolyze phosphate monoesters optimally at alkaline pH. They liberate inorganic phosphate from many types of molecules including nucleotides, proteins and other biomolecules, and synthetic alkyl and aryl phosphate esters. Probably the most-studied phosphatase of all is the AP from
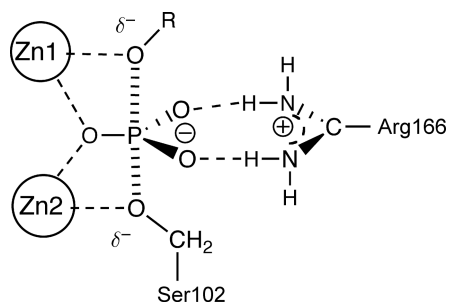
Figure 6   A model of the transition state interactions in the reaction catalyzed by APs, based on X-ray structures.[77,80,82] The residue numbering is that for the *Escherichia coli* enzyme, but the critical Arg and Ser are conserved among all known APs.

E. coli.[73,77–80] The enzyme contains two Zn^{2+} ions and one Mg^{2+} ion in the active site.[77,80] The zinc ions play the most direct roles in catalysis, while the Mg^{2+} has been suggested to either serve a structural role, or to provide a Mg-coordinated hydroxide that deprotonates the serine nucleophile.[81] All known APs have this conserved three-metal ion center, and an arginine (Arg166 in *E. coli* AP) that plays a role in binding and transition state stabilization (**Figure 6**).

In the first step of the AP-catalyzed reaction a serine residue (Ser102 in *E. coli* AP) is phosphorylated. The hydrolysis of this intermediate by water to produce inorganic phosphate competes with phosphoryl transfer to other acceptors, such as alcohols or nucleophilic buffers, when these are present in solution. The overall rate-limiting step at pH > 7 is release of inorganic phosphate product, while at pH < 7 the rate-limiting step is hydrolysis of the phosphoserine (pSer) intermediate.[83]

During the phosphoryl transfer reaction the leaving-group oxygen atom coordinates to one of the zinc ions, designated as Zn1 in **Figure 6**, and resides opposite from the nucleophilic serine. Coordination of the nucleophilic oxygen atom to Zn2 facilitates its deprotonation to form the more nucleophilic alkoxide, while Zn1 stabilizes the negative charge arising from P–O bond fission. In a study by Kim and Wyckoff,[80] the replacement of the zinc ions by Cd^{2+} enabled the isolation and structural determination of the pSer inter-mediate at 2.5 Å resolution. Hydrolysis of the intermediate is assisted by coordination of a nucleophilic water molecule to Zn1, while P–O bond fission is assisted by coordination with Zn2. The roles of the two zinc ions are exactly reversed in the two phosphoryl transfer steps.

For aryl phosphate ester substrates a nonchemical step, such as binding or a conformational change, is rate limiting for k_{cat}/K_m (the portion of the kinetic mechanism consisting of all the steps from substrate binding up to the first irreversible step). This is manifested in a viscosity dependence on k_{cat}/K_m,[85] and in the absence of significant isotope effects on the AP-catalyzed hydrolysis of *p*-NPP.[24] For alkyl monoesters phosphorylation of Ser102 is slower, making the chemical step rate limiting yielding a β_{lg} value of −0.6.[86]

Catalysis by AP is subject to product inhibition by inorganic phosphate with a K_i of approximately 1 µmol l^{-1}. Reactions run under conditions conducive to traditional spectrophotometric assay methods quickly generate concentrations of inorganic phosphate in excess of this, which affects the observed kinetic constants. The AP mechanism has been reexamined using a sensitive ^{32}P-based radioactivity assay for measuring the hydrolysis of alkyl phosphomonoester substrates that avoids this problem.[73] The results confirmed that the chemical step of phosphoryl transfer is rate limiting for k_{cat}/K_m with alkyl ester substrates. The value of −0.85 ± 0.1[73] for β_{lg} is somewhat more negative than the earlier value of −0.6 obtained using spectrophotometric methods.[86] The steep leaving-group dependence reflected in these β_{lg} values indicates considerable P–O bond fission in the transition state. The pH dependence of the kinetic data indicate that the nucleophilic serine hydroxyl group has a $pK_a \leq 5.5$ in the free enzyme. Coordination to Zn1 is primarily responsible for the considerable reduction in this value from the solution pK_a of approximately 16.[73]

The guanidinium group of Arg166 plays a role in binding and in transition state stabilization, but it is not essential for catalysis. Replacement of this residue by either alanine or serine results in weaker binding and reduced k_{cat}.[87–89] Results from LFER experiments with the R166S mutant confirmed the notion that the transition state for the AP-catalyzed reaction is similar to that for the uncatalyzed hydrolysis of dianions.[89]

AP catalyzes reactions of substrates having poor leaving groups with significantly greater catalytic proficiency than those with good leaving groups. For example, the difference in catalytic proficiency between pNPP and methyl phosphate is 10^8-fold,[74] a difference reflected in the divergence between the β_{lg} values of the uncatalyzed and the AP-catalyzed reactions, which are -1.2 and -0.85, respectively. KIE data for pNPP^{2-} and mNBP^{2-} indicate that there are strong interactions between the leaving-group oxygen and an active-site Zn^{2+} ion, and inequalities in the strength of this interaction probably contribute to the different catalytic proficiencies for alkyl and aryl substrates.[74]

AP also catalyzes the hydrolysis of sulfate esters, though its native phosphatase activity is nine orders of magnitude more efficient than the sulfatase reaction.[61,72] Examinations of the sulfatase activity of AP have provided additional insights into the roles of particular parts of the AP active site. KIEs on the sulfatase reaction showed that, like the phosphatase reaction, it proceeds with a loose transition state not distinguishable from the uncatalyzed hydrolysis, and that metal ion interactions with the leaving group are similar in the two reactions. In contrast, AP perturbs the KIE in the nonbridging oxygen atoms of a phosphate ester substrate but not a sulfate one.[72] This is consistent with previous studies indicating that strong interactions between nonbridging oxygen atoms and the zinc ions are particularly important, and are weaker in the sulfate ester substrate due to the lower charge on these atoms.[90]

8.09.4.2   Acid (Histidine) Phosphatases

A number of phosphatase families have acidic pH optima, including PAPs, PTPs, low-molecular weight acid phosphatases, and high-molecular weight acid phosphatases. The latter comprises a distinct family that displays full activity near pH 2.5, have dimeric structures with high-molecular weight subunits in the range of 40–60 kDa, and include human lysosomal, human prostatic, yeast, and plant phosphatases. They utilize an active-site histidine as a nucleophile, and as a result are sometimes referred to as histidine phosphatases. Similar to APs, the acid (or histidine) phosphatases are nonspecific and widespread in nature and their precise functions are uncertain. The utilization of a phosphohistidine intermediate separates these enzymes from other phosphatases that also have acidic pH optima (such as the PAPs and PTPs, which are discussed in subsequent sections).

These enzymes differ from APs in the absence of metal ions. The bovine liver enzyme was used to carry out the transphosphorylation from phenyl (R)-[^{16}O, ^{17}O, ^{18}O] phosphate to (S)-propane-1,2-diol, and demonstrated that the reaction proceeds with net retention of stereochemistry.[91] This indicates that the catalytic reaction proceeds via formation of an intermediate, with two inversions of configuration at phosphorus. The nucleophilic residue was identified as histidine by trapping experiments with p-nitrophenyl [^{32}P]-phosphate followed by denaturation.[92] The nucleophilic histidine residue is part of a characteristic amino acid sequence RHGXRXP[93] (using the amino acid single-letter codes where X represents amino acid residues that are not conserved). The acid (or histidine) phosphatases have not been subjected to as much study as APs, but some further mechanistic information has been obtained from X-ray structures,[94–97] mutagenesis studies,[98,99] and LFER analyses.[99]

The overall structures and the active sites of the human and the rat enzymes are similar, as exemplified by the structure of the rat enzyme with bound vanadate (**Figure 7**). Implications from the structure regarding the
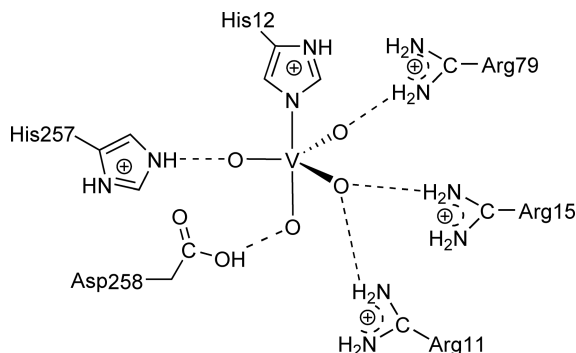
Figure 7   Interactions with active-site residues found in the X-ray structure of rat acid (histidine) phosphatase with bound vanadate.[97]
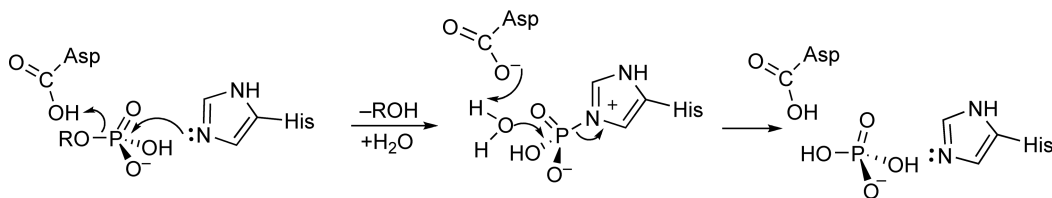
Figure 8 The reaction mechanism of acid (histidine) phosphatases implicated by structural, kinetic, and stereochemical data.

participation of catalytic residues are consistent with kinetic data from site-directed mutants that point to the importance of a conserved histidine and aspartic acid. In the *E. coli* acid phosphatase, His303 and Asp304 correspond to His257 and Asp258 in the rat enzyme. LFER experiments using the *E. coli* enzyme show that the dependence of catalysis on leaving-group pK_a is small for both the native enzyme ($\beta_{lg} = -0.08$) and the H303A mutant (-0.13). In contrast, the β_{lg} for catalysis by the D304A mutant is -0.51, indicating substantial negative charge resides on the leaving group.[99] This result is consistent with a role for Asp304 as a general acid, and the corresponding residue in prostatic acid phosphatase (Asp258) fills the same role.[98,100]

The structural and kinetic data suggest a mechanism in which the first phosphoryl transfer step occurs by attack of the nitrogen atom of histidine accompanied by general acid assistance by the Asp carboxyl group. A logical assumption is that the carboxylate form of the Asp residue then acts as a general base to deprotonate a nucleophilic water molecule in the second step, which is hydrolysis of the phosphohistidine intermediate (**Figure 8**).

8.09.4.3   Purple Acid Phosphatases

PAPs catalyze the hydrolysis of phosphate monoesters with mildly acidic pH optima (5–7) utilizing a binuclear metal center containing a ferric ion and a divalent metal ion. PAPs are also characterized by their purple color, the result of a tyrosine to Fe^{3+} charge transfer transition at about 560 nm.[101] The mammalian enzymes are monomers of approximately 35 kDa with a binuclear $Fe^{3+}-Fe^{2+}$ center, and are highly conserved with 85% or greater sequence identity. Plant PAPs have also been identified from several sources, and most are dimeric. Like the mammalian enzymes, the plant PAPs are fairly homologous, sharing 65% or more sequence identity. The mammalian and plant PAPs, however, share <20% sequence identity with each other. Plant PAPs also differ in the identity of the divalent metal ion, which is most often Zn^{2+} and, less often, Mn^{2+}.[102]

The biological roles for PAPs are uncertain, but several functions have been proposed and are dependent on the organism. In mammals, PAP is secreted into the bone resorptive space, and may be required for normal bone turnover.[103] It has been speculated that PAP participates in bone resorption by dephosphorylating bone matrix proteins such as osteopontin.[104] Such a connection is supported by the fact that PAP levels are elevated in patients with metabolic bone diseases such as osteoporosis and cancers with bone metastases.[105] Because of its high-expression levels, PAP is used as a histochemical marker for osteoporosis.[103] Plant PAPs may serve to liberate phosphate from phosphate esters during germination. Identification of the roles of PAPs in plants is complicated by the fact that plant genomes contain a number of genes that encode for different PAP isoforms. It may be that different isoforms are expressed in particular plant tissues in response to environmental factors.[106,107]

The crystal structures for kidney bean PAP[108] and the PAP from rat bone[109] reveal that despite a sequence similarity of only 18%, the plant and mammalian enzymes share very similar catalytic sites. All PAPs that have been sequenced share a conserved set of active-site residues, and all structures have a similar metal center geometry.[102] These are typified by the kidney bean PAP, in which the two metal ions are separated by 3.1 Å with monodentate bridging Asp carboxylate. These and other residues involved in metal coordination can be seen in **Figure 9**. In addition to the metal ligands, most PAPs have two histidine residues in the region of the metal center that have been shown to be important for substrate positioning and, potentially, as a general acid in catalysis.[102]
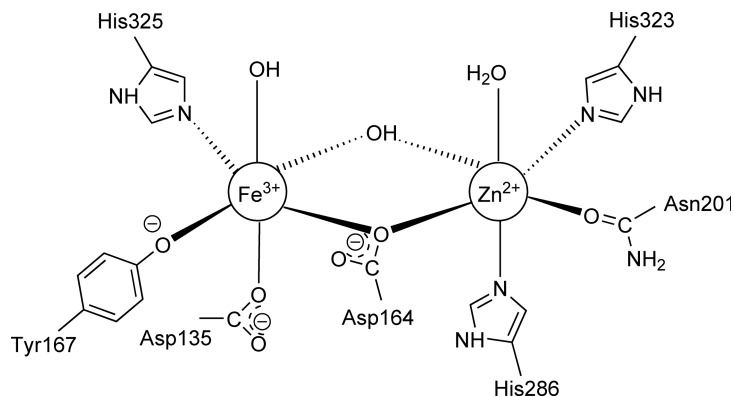
Figure 9 The active site of kidney bean PAP.[108]

The PAP-catalyzed reaction occurs with inversion of stereochemistry at phosphorus, supporting the direct phosphoryl transfer from substrate to a metal-coordinated water or hydroxide nucleophile.[110] Inspection of the active site (**Figure 9**) reveals two candidates, a monodentate hydroxo/aquo moiety coordinated to the ferric ion, and a bridging hydroxide. Debate surrounds the identity of the nucleophile and of the substrate-binding mode; three mechanistic proposals are summarized in **Scheme 3**.[111–114]

All three mechanisms begin with initial coordination of the substrate to the divalent ion. Pathway a shows nucleophilic attack by the Fe^{3+}-coordinated hydroxide. Another proposed pathway is subsequent bidentate coordination, followed by nucleophilic attack of the bridging hydroxide (pathway b). Subsequent addition of water then gives the bidentately coordinated product shown. A third possibility is a reaction of the monodentately coordinated substrate in which the Fe^{3+}-bound hydroxide acts as a general base for nucleophilic attack by water (pathway c).
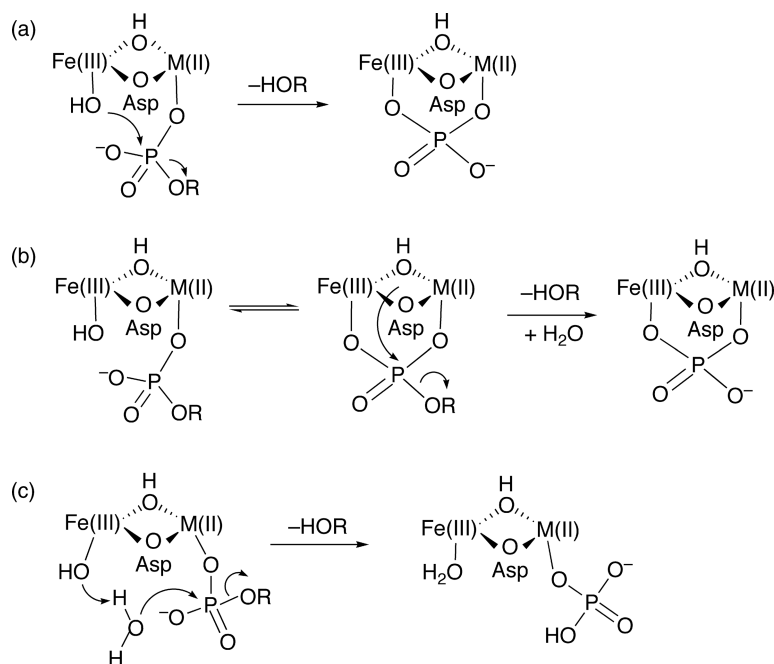
Scheme 3 Three possible roles for metal-assisted phosphoryl transfer to water by PAPs.[115] In (a), a Fe^{3+}-coordinated hydroxide is the nucleophile. In (b), a bridging hydroxide (or oxide) is the nucleophile, forming an intermediate that is subsequently opened by attack of water. In (c), a Fe^{3+}-coordinated hydroxide acts as a general base to facilitate attack by water.

The kinetics of PAPs exhibit a bell-shaped pH–rate dependency, typical of acid–base catalysis.[116–118] NMR data obtained with recombinant human PAP between pH 5.5 and pH 7.1 indicates that pK_{a2} does not involve a metal ligand, and may instead be due to the ionization of one of the two conserved histidine residues near the active site.[119] It has been proposed that one of these histidine residue acts as a general acid in protonation of the leaving group.[120] The other histidine residue (H92 in the human PAP[121] and H202 in the kidney bean PAP[122]) has been suggested to assist in substrate positioning.

The deprotonation of the nucleophile is believed to be responsible for pK_{a1} in the pH profile. The identity of the divalent metal ion affects pK_{a1}, consistent with the proposal that the nucleophilic hydroxide bridges the metals,[123] as in pathway b of **Scheme 3**. An electron paramagnetic resonance (EPR) study with the Fe^{3+}–Zn^{2+} form of bovine spleen PAP showed that the spectroscopic characteristics of the Fe^{3+} ion are unaffected by the addition of phosphate at pH 6.5, the optimal pH for k_{cat}, suggesting that phosphate binds only to the spectroscopically silent Zn^{2+}. In contrast, phosphate does perturb the Fe^{3+} site at lower pH, suggesting that the mode of phosphate binding is pH dependent.[114] At lower pH, the Fe-coordinated ligand would be the more easily displaced water rather than hydroxide, thus facilitating the conversion of the monodentate into the bidentate-bound phosphate, such as shown in the first step of pathway b in **Scheme 3**.

LFER reveal differences between the reactions catalyzed by PAPs from different sources. The pig and red kidney bean PAPs display similar Brønsted correlations for (k_{cat}/K_m) with β_{lg} values of −0.6 and −0.4, respectively.[118] This is less than the maximum value for the hydrolysis of a dianionic monoester of −1.3, and less than the −1.2 value in the uncatalyzed hydrolysis of aryl diesters (see Section 8.09.2.1.1).[13] Such a reduction would be expected from charge neutralization by the putative general acid. These values are in contrast to the near-zero slope of the Brønsted correlations with sweet potato PAP[118] and a human PAP from bone.[124] The sweet potato enzyme is also unusual in its requirement for Mn^{2+} in the divalent site.[118] Furthermore, rather than the standard two histidine residues in the region of the ester group of the substrate, the sweet potato PAP has Glu365 and His295 in this region. The Glu residue has been suggested to function as the general acid in the sweet potato PAP, and an Asp has been proposed for this role in the human bone PAP. The ability of these two PAPs to hydrolyze substrates of low and high leaving-group pK_a values with similar efficiency may be a result of more efficient general acid catalysis by the carboxylic acid groups in these enzymes.

8.09.4.4   Phosphoprotein Phosphatases

The substrates of phosphoprotein phosphatases range from simple phosphopeptides to phosphorylated enzymes involved in many biological processes. Based on substrate specificity and sequence homology, these enzymes fall into two large families, the PTPs and the protein serine/threonine phosphatases (PPs).[125,126] As their name implies, the PPs dephosphorylate pSer and phosphothreonine (pThr) residues. These enzymes have binuclear metal centers and catalyze a phosphoryl transfer from the substrate directly to water. The PPs are distinguished from the PTPs, which do not contain metal ions and dephosphorylate phosphotyrosine (pTyr) residues through a phosphoenzyme intermediate.

8.09.4.4.1   Protein serine/threonine phosphatases

Enzymes in this family are metallophosphatases that hydrolyze phosphomonoesters by means of nucleophilic attack by metal-coordinated hydroxide ions. The products are inorganic phosphate and the respective alcohol, serine or threonine *in vivo*, or an alcohol/phenol when alternate substrates are used *in vitro*.

Based on their catalytic domains, phosphatases in this family are grouped into two subfamilies, designated PPP (protein phosphatase P) and PPM (protein phosphatase M).[125,126] The PPP family shares the conserved signature sequence $DXH(X)_{23–26}GDXXDR(X)_{20–26}GNH(D/E)$.[127,128] Members of the PPP family include PP1, PP2A, PP2B (also called calcineurin), and the lambda protein phosphatase (λPP). Smaller than other PPPs, the λPP was isolated from the bacteriophage lambda. Its similarities to other PPPs include the N-terminal region of the catalytic domain, a binuclear metal center and all the conserved residues at the active site.[129] The PPPs differ along their noncatalytic N- and C-termini and in the regulatory subunits, which affords a diversity of biological roles.[125] For example, PP1 is involved in an assortment of cellular processes including cell cycle progression, protein synthesis, carbohydrate metabolism, transcription, and neural signaling.[130,131] The PP2B enzyme consists of an A catalytic subunit, and can bind to the regulatory subunits PP2B B and calmodulin. Its activity is regulated by the concentration of Ca^{2+}.[125,132]

The complete PP2B unit is responsible for the activation of T-cells mediated by the Ca^{2+} signaling cascade,[133] and it is an important immunosuppressive target.[132,134] Both PP1 and PP2A are potently and selectively inhibited by a number of natural occurring toxins, including okadaic acid, a shellfish poison and strong tumor promoter,[135,136] and microcystin, a cyanobacterial ('blue-green algae') toxin that targets liver cells.[137] In contrast, PP2B is only weakly inhibited by these toxins.[125]

The PPM family is typified by the protein phosphatase 2C (PP2C) subfamily, a group of binuclear metallophosphatases dependent on Mn^{2+} or Mg^{2+}. In contrast to the neutral histidines and amidic carbonyls in PPPs, the active site of PPMs is comprised mostly of carboxylates that coordinate the metal ions, which are 4 Å apart.[138] The PP2C subfamily has no significant sequence homology with members of the PPP family, although they do display similarities in the overall protein fold and both share a dinuclear metal center.[138] However, enzymes in the PPM family show no sensitivity to inhibitors of the PPP family, such as okadaic acid.[135,136] Human PP2C homologues have been identified in several other eukaryotes as well as in prokaryotes.[139] They have a conserved role in negatively regulating stress responses. For example, the plant PP2C analogue MP2C is a negative regulator of the stress induced by MAPK pathway, which is activated by dehydration, cold, wounding, and physical contact.[140]

The crystal structures of several PPPs are known, including: the apo- and tungstate-bound forms of PP1;[141] the microcystin and okadaic acid-bound forms of PP1[142,143] and PP2A;[144] PP2B in apo form and complexed with immunosuppressants;[145–147] and λPP bound to sulfate.[110] More recently, structures for PPP complexes with their regulatory subunits have also been published, for example, PP1/myosin phosphatase (or MYPT1), that regulates smooth muscle relaxation,[148] and PP2B/NFAT, a T-cell activation switch.[149]

The identities of the metal ions at the active sites of PP1 and PP2A *in vivo* are uncertain. The determination of near-stoichiometric amounts of Fe^{3+} and Zn^{2+} combined with EPR studies indicate that calcineurin (PP2B) possesses a Fe^{3+}–Zn^{2+} center,[150] although studies *in vitro* show that the best activators are Mn^{2+} and Ni^{2+}.[132] Structural data show that the metals at the catalytic site are separated by approximately 3.5 Å, with Fe^{3+} in octahedral coordination, in which the axial positions are occupied by a water molecule and an aspartate residue (Asp90). A water molecule and a histidine side chain occupy equatorial positions. The bridging ligands are a hydroxide ion and an aspartate, which, together with another histidine moiety, form the trigonal core of the distorted bipyramid coordination of the Zn^{2+}. The axial ligands of the Zn^{2+} ion are the amide side chain of an asparagine and another histidine residue (**Figure 10**).[147] This coordination environment is identical to that
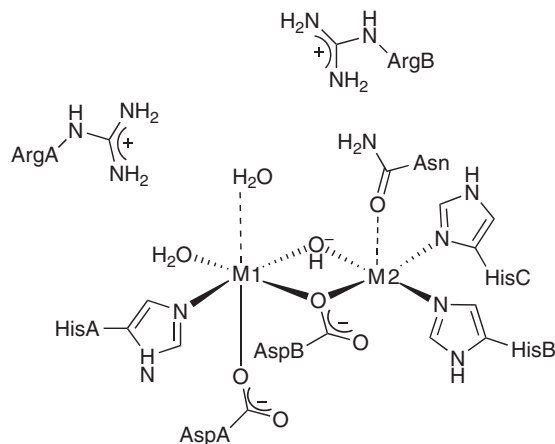
Metals and residues positions at the active site PPPs

|	AspA	HisA	AspB	ArgA	Asn	HisB	ArgB	HisC	M1	M2
PP1	64	66	92	96	124	173	221	248	Mn(II)	Mn(II)
PP2A	57	59	85	89	117	167	214	241	Mn(II)	Mn(II)
PP2B	90	92	118	112	150	199	254	281	Fe(III)	Zn(II)
λPP	20	22	49	53	75	139	162	186	Mn(II)	Mn(II)

Figure 10   Schematic representation of the active site of PPPs based on the crystal structure of calcineurin.[147] Metals listed for each PPP member are those used most frequently in reported crystal structures.

observed in PAPs (see Section 8.09.4.3), except that the axial Fe^{3+}-coordinated water molecule is replaced by a tyrosine residue. The structural similarity suggests the likelihood of mechanistic parallels for PPP enzymes.[125,126]

The catalytic reactions of PPPs proceed via a single direct phosphoryl transfer without participation of a phosphoenzyme intermediate. Supporting evidence includes the finding that calcineurin does not exhibit phosphotransferase activity toward nucleophiles,[151] that activity is inhibited by the products of the reaction with pNPP,[152] and that the enzyme cannot exchange ^{18}O-labeled water with phosphate.[153] Further, the stereochemical course of the reaction catalyzed by the structurally similar PAPs proceeds with inversion of configuration at phosphorus.[110]

The bridging hydroxide ion has been proposed to be the nucleophile based on the crystal structure of PP1 complexed to tungstate.[141] The mechanism is envisaged to follow an A_ND_N mechanism, in which nucleophilic attack is concerted and with departure of the leaving group (**Figure 11**). An essentially identical mechanism has been postulated for calcineurin[145] and λPP.[129] Substrate binding involves interactions with the binuclear metal center and with conserved arginine residues at the top of the active site (Arg96 and Arg221 in PP1), that bring the phosphate group into position for catalysis. Mutations of these arginine residues in PP1,[154,155] PP2B,[156] and λPP[128] dramatically reduce k_{cat}. The phosphate esters that are the natural substrates for the PPPs have poor leaving groups with pK_a of approximately 14, and enzymes that catalyze the hydrolysis of any type of ester typically aid departure of such leaving groups with an electrophilic metal or a general acid catalyst. In PPPs, the metal ions are not directly involved in leaving-group departure, but general acid catalysis has been commonly proposed. A conserved histidine residue (His151 in PP2B and His76 in λPP) is the most likely candidate for general acid catalysis in PPPs.[145] Mutation of this residue in both PP2B and λPP causes a significant reduction in k_{cat}.[156,157] However, catalysis by these His mutants show less than a threefold reduction in rate when the leaving-group basicity increases by more than 3 pK_a units.[157] A much greater sensitivity should result from loss of general acid catalysis. For example, when the general acid is mutated in PTPs, the near-zero β_{lg} of the native enzymes changes into a strong negative β_{lg} of approximately -1.3.[158] Thus, a difference of 3 units in leaving-group pK_a should result in a decrease in the rate constant of approximately 3900-fold if general acid catalysis is abrogated by mutation. Given this, the role of histidine in these enzymes remains inconclusive.

Solvent kinetic isotope effects (SKIEs) in H_2O/D_2O mixtures on the reaction of pNPP catalyzed by calcineurin gave a small normal value of 1.35. Proton inventory and fractionation data are consistent with a mechanism involving a single proton transfer from a metal-bound water, although due to the small KSIE value and the inherent experimental error of the proton inventory technique, the participation of a second proton could not be excluded.[153] Further information has been furnished by heavy-atom isotope effects. Reaction of pNPP catalyzed by λPP shows that phosphoryl transfer is fully rate limiting.[75] However, for calcineurin the
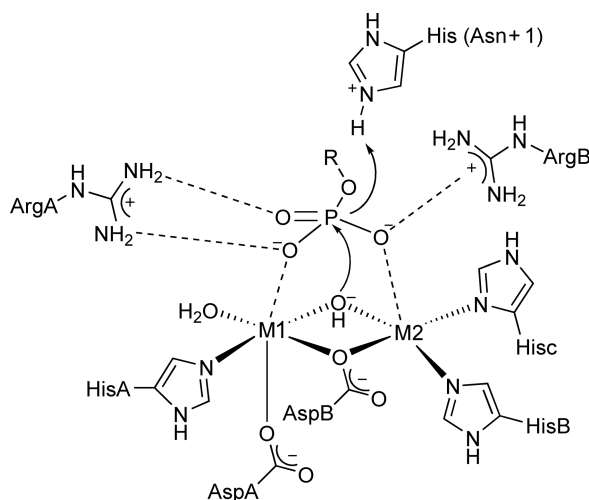
Figure 11 A possible mechanism for phosphate monoester hydrolysis catalyzed by PPPs.[126]

phosphoryl transfer from pNPP is only partially rate limiting up to pH 8, where it becomes more rate limiting.[76,159] The $^{18}(V/K)_{\text{nonbridge}}$ KIE for both λPP and calcineurin are inverse and consistent with a loose transition state similar to the uncatalyzed reaction. The $^{18}(V/K)_{\text{bridge}}$ KIEs indicate partial neutralization of the negative charge on the leaving group. The mutation H76N in λPP causes an increase of $^{18}(V/K)_{\text{bridge}}$ KIE, and the $^{15}(V/K)$ also increases, from 1.0006 ± 0.0003 to 1.0016 ± 0.0003 upon mutation.[75] These values are consistent with charge development on the leaving group, as expected if general acid catalysis is lost. The modest $^{15}(V/K)$ increase, however, implies that only half of a negative charge is present on the leaving group; by comparison, $^{15}(V/K)$ KIE increases to approximately 1.0030 in other phosphatases when general acid catalysis is removed by mutation. This raises the question of whether other stabilizing forces may be present at the active site of PPPs.

The PP2C is the best-characterized member of the PPM family. PP2C is generally described as a Mg^{2+}- or Mn^{2+}-dependent phosphatase.[126] Metal ion binding is weak at the active site and the identity of the binuclear center *in vivo* is uncertain. The K_{metal} dissociation constants obtained kinetically for several divalent ions range from $1.42 \, \text{mmol} \, L^{-1}$ for Mn^{2+} to $20.6 \, \text{mmol} \, L^{-1}$ for Mg^{2+}.[160] Many other divalent metal ions activate PP2C, some resulting in higher enzymatic activity than the likely *in vivo* metals Mn^{2+} and Mg^{2+}. Interestingly, Fe^{2+} results in $k_{\text{cat}}/K_{\text{metal}}$ values that are >3000-fold higher than with Mg^{2+}, but it is improbable that such a strong Fenton reaction catalyst is present *in vivo* with a weak binding site such as PP2C.

The kinetically determined K_{metal} values are surprisingly high in light of physiological metal concentrations. This raises the question of whether metal binding precedes the binding of the substrate or if it occurs simultaneously, with the second metal ion brought to the active site by cooperative interactions with the substrate. On the other hand, isothermal titration calorimetry (ITC) studies with MspP, a PP2C homologue found in the saprophyte *Mycobacterium smegmatis*, obtained dissociation constants for binding of two Mn^{2+} ions in the nano- and micromolar range, much smaller than the kinetic values obtained with PP2C. The ITC data are within the range that would suggest full occupancy of the active site at physiological metal concentrations.[161]

The binding mode of the phosphate ester in the active sites of PPMs may vary with experimental conditions. An X-ray structure of PP2C with bound phosphate shows the oxyanion not bound directly to the binuclear center, but indirectly through hydrogen bonds with metal-bound water molecules.[138] It has been noted that this crystallization was carried out at pH 5.0; at this pH, the enzyme is improperly protonated and exhibits only 0.1% of the optimal activity observed at higher pH.[160,162] Hence, this crystal may not represent the catalytically active structure. Moreover, at pH 5 the major form of phosphate ($pK_a = 7.2$) is the monoanion, which favors interactions mediated by water rather than directly to the metal center.[163] Recently, a crystal structure with phosphate bound was obtained for MspP at higher pH where the enzyme is more active.[163] In this structure the phosphate oxygen atoms bind directly to the metal center in a fashion similar to that in the structure of PP1 bound to tungstate.[141] The phosphorus atom is placed in a favorable position for the nucleophilic attack of a metal-bridging hydroxide ion.

The pH–rate profile of k_{cat}/K_m using Mn^{2+} or Mg^{2+} as metal ions for the PP2C-catalyzed reaction of pNPP shows a bell-shaped behavior typical of acid–base catalysis. Two critical ionizations are required for catalysis, one on the acidic limb due to a residue of $pK_a \sim 7.5$ and another on the basic limb from the deprotonation of a residue of $pK_a = 9$.[160] It has been proposed that the pK_a on the acidic limb belongs to the bridge metal-bound water molecule, which must be deprotonated to generate a nucleophilic hydroxide ion similar to the mechanism of PPPs. The identity of the acid catalyst is not certain, although it has been proposed that eukaryotic PPMs use a conserved histidine residue for this role (His62 in PP2C).[162] Interestingly, in prokaryotic PPMs this histidine is replaced by phenylalanine, and these enzymes possess a binding site for a third metal ion on the opposite side of the active site. This additional metal ion may provide general acid catalysis via a coordinated water molecule (**Figure 12**).[164]

Site-directed mutagenesis reveals further details about general-acid catalysis in PPMs. The mutation H62Q in PP2C causes a 20-fold decrease in k_{cat}.[162] However, the pH-dependence for k_{cat} is still bell-shaped, though the basic limb is absent in the profile for k_{cat}/K_m. A Brønsted plot for the reaction of PP2C with several aryl phosphomonoesters revealed that at pH 7.0 phosphoryl transfer is rate limiting only for substrates with leaving-group $pK_a > 7$. Under these conditions a $\beta_{\text{lg}} = -0.32$ was obtained. The H62Q mutation yielded a $\beta_{\text{lg}} = -0.84$, consistent with the proposed general-acid catalysis from His62.[162] In prokaryotic PPMs, mutations of residues involved in binding of the third metal do not result in decreased catalysis; however, these experiments were
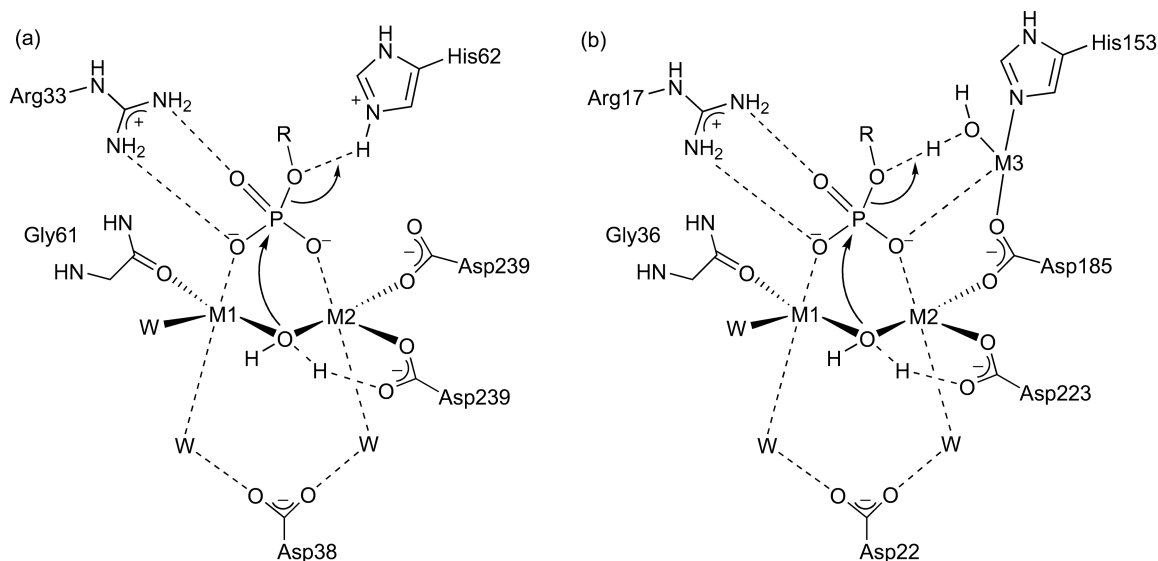
Figure 12 Possible mechanisms for phosphate monoester hydrolysis catalyzed by eukaryotic (a) and prokaryotic (b) PPMs. The representation for eukaryotic PPMs is based on the crystal structure of PP2Cα bound to phosphate (PDB code 1A6Q).[138] In the published structure the phosphate anion is not directly coordinated to the metal ions and it has been modified here to represent a catalytic complex analogous with that implicated by the crystal structure of the homologue MspP bound to phosphate (PDB code 2JFR);[163] see text for details. The octahedral coordination of metals M1 and M2 are completed by bridging carboxylate residues Asp60 in PP2Cα, and Asp35 in MspP (not shown for clarity). Water molecules complete the octahedral coordination at third metal site of the prokaryotic MspP.

carried out in the presence of excessively high Mn^{2+} concentrations that may have resulted in occupancy despite the weakened coordination site.[165] It has also been postulated that the third metal ion may mediate phosphate binding and substrate selectivity rather than participate in catalysis.[161,164]

8.09.4.4.2 *Protein Tyrosine Phosphatases*

The PTPs catalyze the hydrolysis of phosphorylated tyrosine residues in proteins, to yield the free tyrosine side chains and inorganic phosphate. They are classified according to substrate specificity: (1) tyrosine-specific PTPs, such as the *Yersinia* PTP (YopH) and the mammalian PTP1B and PTP1, which *in vivo* hydrolyze only pTyr residues; as well as (2) the dual-specificity phosphatases (DSPs), such as the human VHR and Cdc25, which hydrolyze pTyr and pSer and pThr residues of protein substrates.[166] Based on their cellular localization, PTPs are classified as receptor-like or intracellular.[167]

In concert with protein tyrosine kinases (PTKs), the action of PTPs comprises a crucial system for the regulation of the biological activity of other proteins.[125,168,169] In eukaryotes, phosphorylation occurs primarily on serine or threonine, with tyrosine accounting for only 0.01–0.05% of the total. However, upon oncogenesis or growth factor stimulation, tyrosine phosphorylation increases to 1–2% of the total protein phosphorylation in the cell.[168] Tyrosine phosphorylation is also a key process in insulin action. Propagation of the insulin signal is triggered by the phosphorylation of a tyrosine residue on the insulin receptor, while negative regulation is promoted by PTPs, such as PTP-1B, LAR, and PTPα. Thus, inhibitors for such PTPs can result in cell hypersensitivity to insulin providing viable therapy for diabetes.[170] PTPs may also represent novel targets for antibiotic development. Parasitic protozoa species from the genus *Leishmania* cause leishmaniasis, a severe cutaneous disease, responsible for over 1.5 million cases a year in tropical and subtropical countries. These organisms induce the PTP SHP-1 in the host, consequently causing the dephosphorylation of critical regulatory proteins involved in macrophage activation.[171] The genomes of bacteria examined to date do not reveal the presence of PTKs, suggesting that the PTPs produced by bacteria serve as pathogenic factors. Examples of bacteria that utilize PTPs to disrupt immune responses in the host include the typhus agent
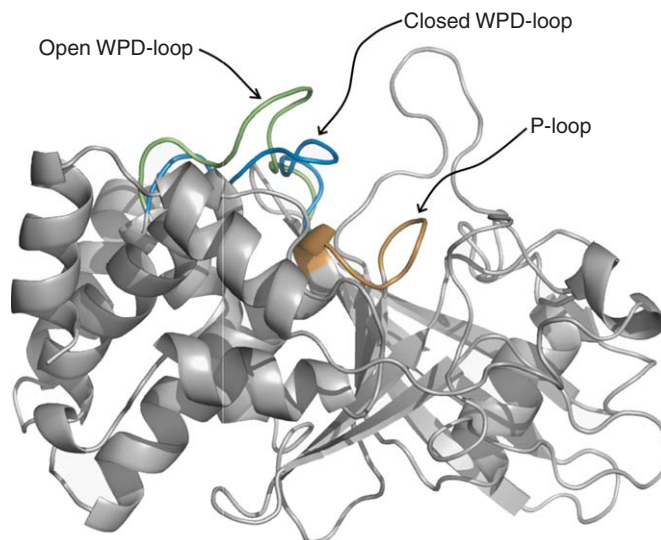
Figure 13 Structures of PTPs include two important motifs, the P-loop that bears the cysteine nucleophile within the general signature motif (H/V)C(X)$_5$R(S/T), and the WPD-loop, which includes an important aspartic acid, a general acid–base catalyst. Substrate binding by the P-loop promotes a change of the WPD-loop conformation from an open, inactive to a closed, active conformation in which the aspartic acid completes the catalytic ensemble used for catalysis. The representation in this figure was created using PyMol from the PTP1B structures in apo-bound (PDB 2CM2)[187] and inhibitor-bound (PDB 1BZJ)[188] forms.

Salmonella typhimurium, and the genus *Yersinia*, which includes three species responsible for human diseases from gastrointestinal conditions to bubonic plague.[172,173]

Mechanistic information has been obtained from studies with a number of members of this group, including the tyrosine-specific PTPs YopH[174–178] and PTP1B,[179–182] and the DSPs Stp1[183] and VHR.[184–186] Conclusions are based on a number of protein crystal structures and from mutagenesis and kinetic studies. These studies reveal that binding and catalysis involve two central motifs: the P-loop, which includes a nucleophilic Cys; and the WPD-loop, a flexible loop found in most PTPs, which contains a catalytically important Asp residue (**Figure 13**).

Kinetic data indicate a typical ping-pong mechanism.[189] In the first step, nucleophilic attack of the cysteine thiolate on the phosphate moiety of the substrate occurs with release of the ester leaving group and formation of a phosphoenzyme intermediate. The next step is attack of a water molecule on this intermediate to yield inorganic phosphate and the regenerated enzyme. In both steps, catalysis is enhanced by participation of the conserved aspartic acid that functions as a general acid catalyst in the first step, and as a general base catalyst in the second step (**Figure 14**).

The P-loop is the central binding site for the substrate as well as for competitive inhibitors, which include tetrahedral phosphate-like oxyanions such as sulfate, tungstate and arsenate.[158,174,175,177] The P-loop has the general signature motif (H/V)C(X)$_5$R(S/T) and adopts a common architecture also found in several ATP and GTP binding proteins. A series of backbone amide N–H bonds point toward a central region forming a
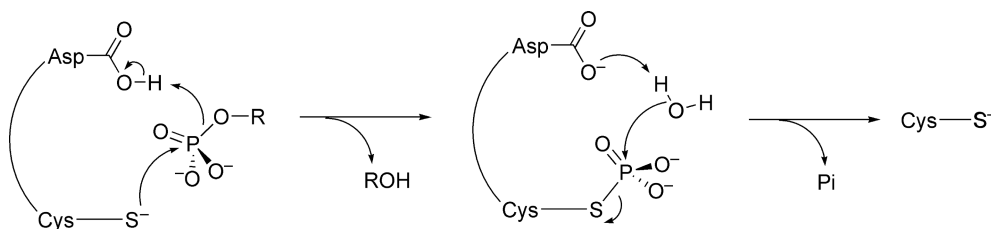
Figure 14 General mechanism for hydrolysis of phosphate monoesters catalyzed by PTPs.
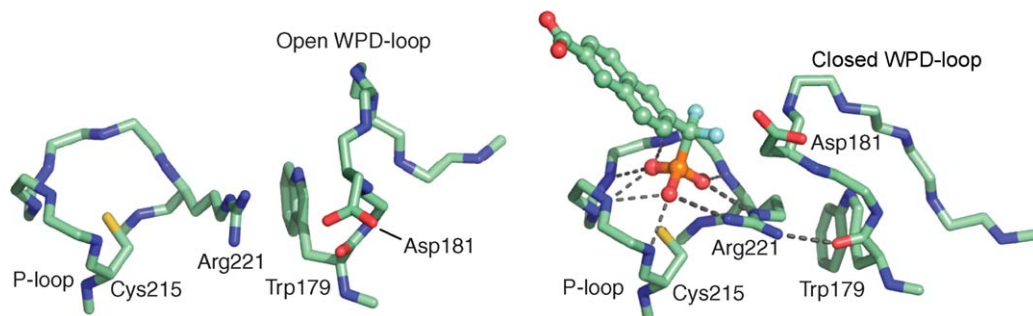
Figure 15 The binding of substrate and competitive inhibitors in the active site of PTPs causes conformational changes of important active-site residues related to catalysis. This figure shows the active site of PTP1B in the unbound (left, PDB 2CM2)[187] and a phosphonate inhibitor bound states (right, PDB 1BZJ).[188] Hydrogen bonds are present between the oxyanion and the backbone amide groups of the P-loop and with Arg221. This residue rotates to form two hydrogen bonds to the oxyanion, resulting in a new hydrogen bond with the carbonyl oxygen of Trp179. Associated movement of the WPD-loop brings Asp181 into position to function as a general acid during the first catalytic step. Except for the carbonyl oxygen of Trp179, all carbonyl oxygen atoms were omitted for clarity.

horseshoe structure at the N terminus of an α-helix. This creates a structure well suited for phosphate binding, due to the positively charged environment and network of hydrogen bonds, sometimes referred to as a 'giant anion hole' or 'nest'.[190]

The WPD loop is a flexible β-turn found in all tyrosine-specific PTPs, and includes the conserved aspartic acid residue that serves as a general acid–base catalyst. Substrate binding thermodynamically favors the closed, catalytically active conformation, where the aspartic acid is in position for catalysis (**Figure 15**).[175] The DSPs also share a conserved aspartic acid in this catalytic role. However, except for VHZ, a recently purified DSP which may possess a flexible IPD loop,[191] the aspartic acid in DSPs is located on a rigid structure. Consequently, no conformational change analogous to WPD loop movement in PTPs seems to be associated with catalysis for most DSPs.

Substrate recognition in PTPs is determined in part by the depth of the active-site cavity and by the P-loop architecture. The active site crevice is approximately 9 Å deep in tyrosine-specific PTPs and approximately 6 Å for the DSPs.[186] Hence, the deeper active site for tyrosine-specific PTPs can select exclusively for pTyr, while the shallower active site in the DSPs can productively bind either pSer/pThr and pTyr substrates. Additional substrate recognition, by discrimination of the amino acid sequence neighboring the phosphorylated residue, occurs in regions flanking the active site. The physiological substrates for PTPs are phosphoproteins, and an understanding of the molecular basis for substrate recognition by PTPs is an area of ongoing research. Most of the information regarding substrate recognition by PTPs has been obtained from studies using synthetic phosphopeptides. Crystal structures with bound substrate analogues have been determined for YopH,[192,193] PTP1B,[194,195] and VHR.[196] Studies using combinatorial libraries show that tyrosine-specific PTPs display more modest specificity for their substrates than DSPs. For example, PTP1B is rather promiscuous and dephosphorylates a variety of peptides and proteins with similar activity.[195,197,198,199] In contrast, the DSPs MKP3 and Cdc25 exhibit strong preference for their native substrates.[200,201]

As a matter of simplicity and to permit comparisons between PTPs, kinetic studies *in vitro* are often carried out using simple aryl phosphomonoesters, mostly with chromophoric or fluorophoric characteristics. The pH–rate profiles for V/K in PTPs are invariably bell shaped with a maximum between 5 and 6, slopes of $+2$ on the acidic limb and -1 on the basic limb. The acidic limb results from the deprotonation of the monoanion of the substrate ($pK_a \sim 5.1$ for aryl phosphate substrates) to give the active dianionic substrate, and from the deprotonation of the cysteine thiol ($pK_a \sim 4.7$ in YopH[202] and 5.5 in VHR[185]) to produce the active thiolate nucleophile. On the basic limb, the slope of -1 is assigned to the aspartic acid, $pK_a \sim 5.1$ in YopH.[202] The nucleophilic cysteine thiol in PTPs and DSPs has an unusually low pK_a, about three units below the thiol group in free cysteine. Several interactions in the active site stabilize the cysteine thiolate. In YopH, the His402 residue provides the primary stabilization of the Cys403 thiolate. Mutagenesis of His402 to asparagine or alanine increases the pK_a of Cys403 by 1.3 or 2.7 pH units, respectively. In crystal structures, His402 does not
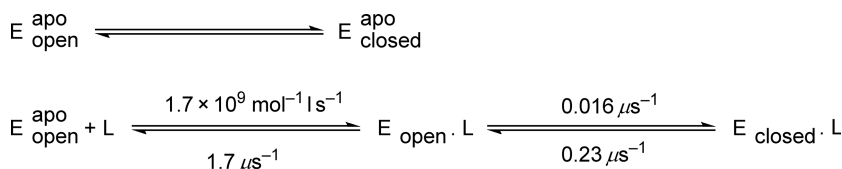
hydrogen bond directly to Cys403, but participates in a network of hydrogen bonds in the active site. This network is proposed to effectively polarize NH groups in the P-loop creating a positive microenvironment that stabilizes the cysteine thiolate.[202] Additionally, Thr410 can hydrogen bond to Cys403 and it is responsible for another reduction of about 0.6 pH units.[203]

The binding of substrate is associated with crucial dynamic events important for catalysis in PTPs. The P-loop brings the phosphoryl group of the substrate into position for nucleophilic attack by the cysteine thiolate. The guanidinium side chain of the conserved arginine rotates to a new conformation that interacts with the substrate, and a new hydrogen bond forms with the carbonyl oxygen of the Trp at the WPD-loop (**Figure 15**). This event helps stabilize the WPD-loop in the closed conformation, with important consequences for catalysis.[174,175] In the first step, culminating with phosphoenzyme formation, the aspartic acid at the top of the loop swings 8 Å into the active site and into position for general acid catalysis. Proton transfer to the leaving-group oxygen occurs, which is in-line with the nucleophilic attack of cysteine. In the second step, the aspartate carboxylate can activate a nucleophilic water molecule. The directionality of water attack is also believed to be in-line through an $A_N D_N$ mechanism.[177,182] In agreement with the importance of the P-loop residues Cys and Arg, their mutagenesis causes severe effects on catalysis.[204,205] Catalysis is also impaired by mutations to the WPD-loop residues Trp and Asp in PTPs,[158,176,206] and to the Asp in DSPs.[185]

Studies of the WPD-loop dynamics have been carried out to obtain a deeper understanding of the role of protein motions in catalysis by PTPs. Experimental techniques that have been employed include time-resolved fluorescence anisotropy, steady-state fluorescence, ultraviolet resonance Raman (UVRR) spectroscopy,[207] relaxation kinetics,[208] H/D exchange, and electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry,[209] and WPD-loop motion has been addressed theoretically by molecular dynamics (MD) simulations.[210] These studies all verify that the WPD-loop in apo YopH alternates between open and partially closed conformations, although estimates of the frequency vary from the nanosecond to the millisecond range. Substrate binding does not change the rate constant for WPD-loop open/closure, but the reopening rate is significantly slowed, pushing the equilibrium toward the closed form (**Scheme 4**).[208]

Details about the transition state were revealed by KIEs for the reactions of the wild-type and mutant enzymes of the tyrosine-specific PTPs YopH,[206,211] PTP1,[212] PTP1B (T. A. S. Brandao *et al.*, unpublished observations) and the DSPs VHR[213,214] and Stp1.[215] The small to inverse $^{18}(V/K)_{nonbridge}$ KIEs suggest that the transition states are loose and similar to uncatalyzed phosphoryl transfer. Except for Stp1, $^{15}(V/K)$ values for all PTPs are negligible indicating that proton transfer is concerted with P–O bond cleavage and the leaving group remains neutral. For Stp1, a small normal $^{15}(V/K)$ KIE is found indicating that proton transfer lags slightly behind P–O cleavage resulting in a partial negative charge on the leaving group in the transition state. The primary $^{18}(V/K)_{bridge}$ KIEs exhibited magnitudes consistent with a late transition state in which P–O bond fission and proton transfer are well advanced. All mutations that impair general acid catalysis result in magnitudes of $^{15}(V/K)$ that are considerably increased, denoting significant negative charge on the leaving group. This was confirmed by the accompanying increase in $^{18}(V/K)_{bridge}$ when the inverse contribution from protonation is absent.

Consistent with the KIE results, LFER studies showed that k_{cat} for the hydrolysis of aryl phosphomonoesters by native YopH exhibits almost no dependence on the basicity of the leaving group between pK_{lg} 7 and 15. In contrast, a strong negative $\beta_{lg} = -1.3$ is found in reactions of mutants in which general acid catalysis is disabled.[158] The fact that alcohols as well as water can dephosphorylate the phosphoenzyme intermediate was utilized to evaluate the transition state of this step. It was found that, for native Stp1 and the YopH mutant Q446A, the $\beta_{nuc} \sim 0.15$, indicative of little nucleophilic participation and, presumably, a loose transition state.[189,216]

$$E\ {}^{apo}_{open} \rightleftharpoons E\ {}^{apo}_{closed}$$

$$E\ {}^{apo}_{open} + L \underset{1.7\ \mu s^{-1}}{\overset{1.7 \times 10^9\ mol^{-1}\,l\,s^{-1}}{\rightleftharpoons}} E_{open}\cdot L \underset{0.23\ \mu s^{-1}}{\overset{0.016\ \mu s^{-1}}{\rightleftharpoons}} E_{closed}\cdot L$$

Scheme 4   A substrate binding and WPD loop movement model for YopH with rate constants at 35 °C obtained from temperature-jump experiments. L is the ligand *p*-nitrocatechol sulfate.[208]

It is interesting that despite the mechanistic and structural similarities in PTPs, k_{cat} values in this family differ by more than 10^3-fold.[217] The reasons are uncertain, but such disparities in rate might reflect the significant participation of conformational changes involving the WPD-loop and perhaps other regions of these proteins. Estimated rates of WPD-loop movement in YopH are faster than catalysis. This suggests that not all instances of loop closure result in a catalytically productive conformation. Either variations in the rate of loop opening/closing or the efficiency with which loop closure leads to a catalytic event might be responsible for the differences in catalytic efficiency within the PTP family.
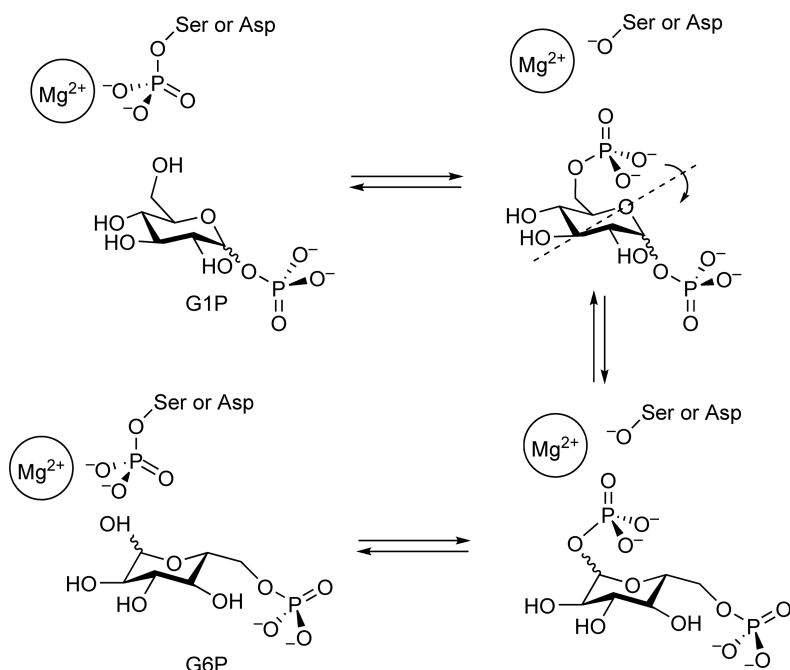
8.09.4.5   Phosphoglucomutases

The α- and β-phosphoglucomutases are enzymes found in all plant and animal cells that catalyze the reversible isomerization of glucose-1-phosphate (G1P) and glucose-6-phosphate (G6P) (**Scheme 5**). The enzyme assists in metabolism by converting the G1P, released from glycogen by glycogen phosphorylase, to G6P, which then travels down the pentose phosphate pathway. When concentrations of G6P are high, the reverse reaction produces G1P, which is converted into UDP-glucose by a number of steps. If activated by insulin, glycogen synthase will then incorporate G1P into glycogen.

PGMs are named for the anomer of G1P on which they act; α-phosphoglucomutase (α-PGM) acts on the α-C(1) anomer of G1P, while β-PGM catalyzes the reaction of the β-C(1) anomer. Both mutases employ Mg^{2+} as a cofactor.

Stereochemical analysis of the products using $[^{16}O, ^{17}O, ^{18}O]$-phosphate ester methods shows that the PGM reaction proceeds with overall retention of configuration at phosphorus,[218] indicating that an even number of phosphoryl transfers occur, and thus a phosphoenzyme intermediate. The enzymatic nucleophile in α-PGM is serine, while β-PGM uses an aspartic acid. The phosphorylated PGM binds either G1P or G6P and transfers the phosphoryl group to the C(6)OH or C(1)OH, respectively (**Scheme 5**).

The overall isomerization reaction presumably requires the intermediate glucose-1,6-diphosphate to either dissociate from the enzyme and rebind in a different configuration, or to reorient itself within the active site.

Scheme 5   Representations of the four enzyme–substrate complexes on the reaction pathway of phosphoglucomutases (PGMs). G1P, glucose-1-phosphate; G6P, glucose-6-phosphate. The α-PGMs utilize the α-anomers at C1, and a serine as the phosphoryl carrier. The β-PGMs utilize the β-anomers at C1 and an aspartate as the phosphoryl carrier.
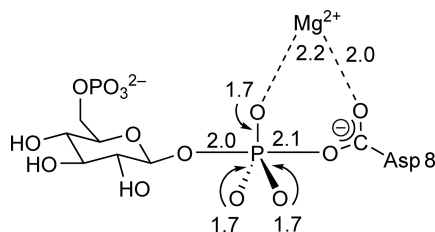
Figure 16   Geometry of the putative pentacoordinate phosphorane observed in the X-ray structure of β-PGM from *Lactococcus lactis*. Distances are given in angstroms.[222]

Substrate reorientation within the active site occurs at a faster rate than dissociation with the PGMs from rabbit muscle and from rat.[219] In contrast, dissociation is more rapid with the PGMs from *Bacillus cereus* and *Micrococcus lysodekticus.* The X-ray structure of the β-phosphoglucomutase from *Lactococcus lactis* reveals a small active-site cavity, seemingly too small to permit reorientation. The observation of two neighboring Asp residues in this active site, Asp8 and Asp10, led to the proposal that both might function in catalysis with one as the phosphoryl acceptor and the other as the donor, eliminating the need for dissociation and rebinding or for reorientation of the intermediate.[220]

On the basis of spectroscopic studies using the transition state analogue glucose-1-phosphate 6-vanadate bound to the enzyme, the transition state for phosphoryl transfer was concluded to be S_N2-like in character, and tighter than the transition states for uncatalyzed reactions of phosphomonoester dianions.[221] A 1.2 Å resolution X-ray structure[222] of β-PGM from *Lactococcus lactis* revealed a seeming intermediate in the reaction (**Figure 16**). There has been some question[223–225] and response[226,227] as to whether the pertinent 4-atom moiety observed in the crystal structure is PO_3^- or MgF_3^-. Since magnesium and fluoride ions are necessary to obtain crystals, either possibility is tenable. The species was originally described[222] as a pentacovalent phosphorane, but if it is a phosphorus species, it is certainly an unusual one. Pentacovalent phosphoranes exhibit typical P–O bond lengths of 1.7 Å.[7,8] The elongated apical P–O bond lengths in the crystal structure would correspond to bond orders of only 0.2–0.3 using Pauling's rule.[228]

In *ab initio* calculations, the species was modeled separately with MgF_3^- and with PO_3^- as the central moiety.[224] In the case of MgF_3^-, an energy minimum was obtained with a geometry that corresponded well with the X-ray structure. When modeled as PO_3^- the species is unstable, and resembles instead a transition state for phosphoryl transfer (with a moderate barrier of $14\,kcal\,mol^{-1}$) from substrate G6P to the product glucose-1,6-diphosphate. The computationally modeled reaction is concerted with no phosphorane intermediate.[224]

^{19}F NMR and kinetic analysis were used to study the complex formed when β-PGM is added to a solution containing $MgCl_2$ and NH_4F. In addition to the resonances for free fluoride ion and MgF^+, three new ^{19}F resonances result upon addition of β-PGM, indicating three distinct sites for fluoride binding in the complex. The three F ions do not interchange with one another, nor with fluoride ion in solution, on the time scale of the NMR experiments (10 s), and nuclear Overhauser effect (NOE) experiments permitted assignments to be made for each of the three fluoride ions within the active-site structure.[225] Consistent with the notion that MgF_3^- is a transition state mimic of PO_3^-, the reaction is inhibited by fluoride with a K_i in the low millimolar range (**Figure 17**).[225]
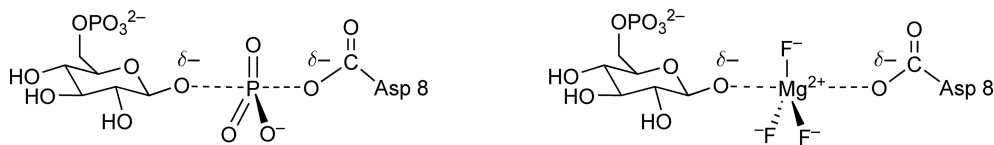
Figure 17   A transition state for a concerted phosphoryl transfer between the aspartylphosphate carrier and G6P (left), and a transition state analogue formed from β-G6P, magnesium, and fluoride (right).[225]

8.09.5 Sulfatases

Sulfatases comprise a diverse family of enzymes.[229,230] Depending on the particular enzyme, the catalytic mechanism for sulfate ester hydrolysis proceeds by cleavage of either the S–O or the C–O bond, yielding in both cases inorganic sulfate. The S–O cleavage reaction is more common, and occurs in the sulfate ester hydrolysis reactions catalyzed by all arylsulfatases (ASs) and those of some alkyl sulfatases. The C–O cleavage pathway has been observed only for alkyl sulfatases, and proceeds with inversion of configuration at the carbon center (**Scheme 6**). An interesting group of alkyl sulfatases are the nonheme Fe^{2+}, α-ketoglutarate-dependent dioxygenases; enzymes in this family carry out the C–O bond cleavages of alkyl sulfate esters to their respective aldehydes and inorganic sulfate using a radical mechanism.

Sulfatases are involved in several processes integral to human health and diseases.[230] Sulfatases have been implicated in hormone regulation, gamete interactions, and bone and cartilage formation.[231,232] High levels of sulfatase activity have been found in several adenomas,[233] including prostate and breast cancers.[234,235] Many sulfatases have been implicated as pathogenic factors of various bacteria.[236,237] Alkylsulfatases that catalyze C–O bond cleavage are used industrially to carry out the enantiomeric enrichment of secondary alcohols, since these enzymes often exhibit affinity for a particular stereoisomer of chiral secondary sulfate esters.[238] In addition, several alkyl sulfatases may decompose sulfate esters in wastewater and have been considered for use in sewage treatment.[239]

8.09.5.1 Arylsulfatases

ASs have been found in several prokaryotes and eukaryotes. Despite sharing only from 20 to 60% sequence identity, AS tertiary structures exhibit high homology, and residues at the catalytic site are highly conserved in both eukaryotic and prokaryotic ASs (**Figure 18**).[230] It has been suggested that AS members share a common origin based on an ancestral gene.[240]

The activity of ASs depends upon the posttranslational modification of highly conserved residues: cysteine in eukaryotes[241,242] and in some prokaryotes,[243,244] and serine in other prokaryotes.[245] This posttranslational modification is enzymatically mediated and results in the oxidation of the cysteine or serine to yield an aldehyde residue, referred to as a formylglycine (FGly).[246] The inability to perform this step results in the occurrence of multiple sulfatase deficiency in humans, an autosomal recessive disease in which the activities of all sulfatases are sharply reduced.[241]

The crystal structures of a number of ASs have been reported.[62,247–251] The human ASA is the most studied enzyme in this group. The structure reveals that the aldehyde FGly residue is in a hydrated *gem*-diol form,[248] as seen in the 1.3 Å structure of the *P. aeruginosa* arylsulfatase (PAS).[62] Surrounding the hydrated aldehyde moiety many polar residues form a hydrogen bond network (**Figure 19**). The FGlyOγ1 oxygen atom is coordinated to a divalent metal ion, proposed to be calcium *in vivo*.[62,251] The FGlyOγ2 oxygen atom hydrogen bonds to conserved residues Arg, His, and Lys.

ASs are able to hydrolyze a variety of sulfate esters. Depending on the enzyme specificity the physiological substrates include, among others, small mono- and disaccharide sulfates, hydrophobic steroids, and amphiphilic sulfated carbohydrates found in glycosaminoglycans (GAGs).[230] However, small synthetic aryl sulfates are the
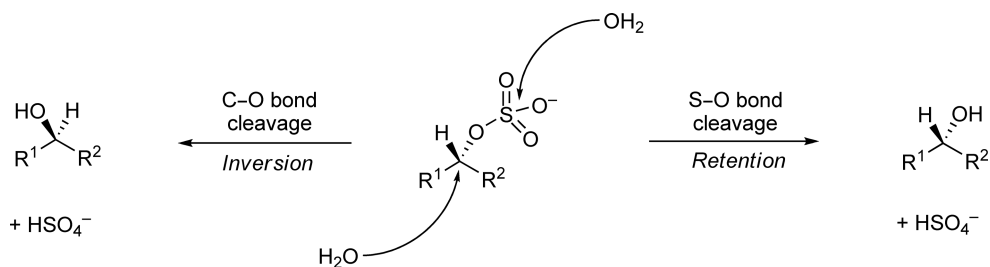
Scheme 6 The potential stereochemical outcomes at the chiral secondary carbon atom of a substrate in the hydrolysis of sulfate esters catalyzed by sulfatases.

Figure 18 Conserved active-site residues in ASs. Numbering is for human ASA. Highlighted are highly conserved residues, underlined residues are involved in metal binding, and in red are the cysteine residues that undergo posttranslational modification to FGly. The secondary structure DSSP assignment is according to the ASA structure (PDB code 1AUK): E = beta strand, H = helix, and L = loop. The codes refer to the ExPASy accession numbers.
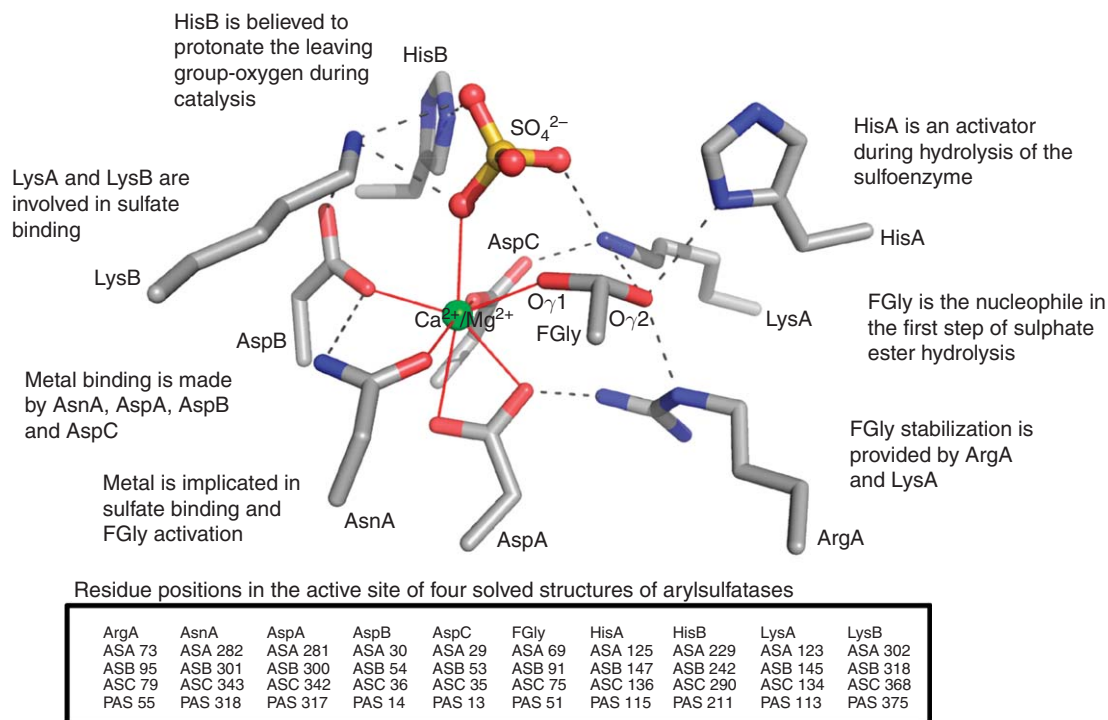
Figure 19 Sulfate anion interactions in the active site of ASs. The representation was drawn from the 1.3 Å sulfate bound structure of the *P. aeuroginosa* AS (PAS, PDB code 1HDH).[62] The numbering of residues is according to Hanson *et al*.[230]

most-often used substrates in kinetic and crystallographic studies. Two distinct recognition sites form the binding sites for sulfate esters in ASs, one to the substrate ester moiety and the other to the sulfuryl group. Among ASs, structurally unrelated and variable architectures recognize the ester moiety, but the sulfuryl group is held in an oxyanion hole with features similar among all ASs.[252] For example, the competitive inhibitor inorganic sulfate adopts a position in the PAS active site[62] similar to that occupied by the sulfuryl moiety of *p*-nitrocatechol sulfate (*p*NCS) in the inactive mutant of ASA C69A (ASA C69A, in which the Cys precursor to the functional FGly residue is mutated to Ala).[249] The sulfuryl recognition site involves interactions of the nonbridge oxygen atoms to the metal center, hydrogen bonds to the main chain NH group of FGly, and electrostatic/hydrogen bonds to the positively charged side chains of the conserved residues LysA, LysB, and HisA (**Figure 19**). As a result, the protonation states of these residues are responsible for the pH-dependent increase in K_m at pH > 7.[253,254]

A growing body of evidence on the mechanism of ASs has been collected over the years. Labeling experiments show that sulfate ester hydrolysis by ASs proceeds via S–O bond fission, and ^{18}O is incorporated from water into the inorganic sulfate product.[255,256] The stereochemical course of the reaction catalyzed by the AS from *Aspergillus oryzae* was tested by examining the sulfuryl transfer from phenyl $[(R)\text{-}^{16}O, {}^{17}O, {}^{18}O]$sulfate to *p*-cresol. The reaction was shown to proceed with retention of configuration at sulfur, providing evidence for the involvement of a sulfoenzyme intermediate.[257] Mutagenesis experiments provided further evidence for an intermediate; when the FGly residues in ASA and ASB were mutated to Ser the sulfate group remained covalently bound to the enzyme after hydrolysis of $[^{35}S]$-*p*NCS, while in the reaction by the native enzyme the sulfate anion was completely released.[258] The sulfoenzyme has also been observed in crystallographic studies.[247,252] It is now accepted that hydrolysis of sulfate esters by ASs is a two-step, ping-pong mechanism (**Figure 20**).[259] **Figure 19** gives a brief summary of the roles of conserved residues in the active sites of ASs.

Crystallographic studies indicate that in the first step, formation of the sulfoenzyme, the sulfate ester leaving group is opposite to the FGly residue and its $O\gamma1$ oxygen atom is oriented for nucleophilic attack on the sulfur atom. Hence, the reaction is believed to proceed via an S_N2 (A_ND_N) mechanism, where the FGlyO$\gamma1$ and the leaving-group oxygen occupy the apical positions in the pentacoordinate transition state (**Figure 20(a)**).[62,248,249,251] An alternative mechanism was initially proposed for the first step of catalysis by ASB (**Figure 20(b)**),[247] in which the FGly aldehyde suffered nucleophilic attack from the oxygen atom of the sulfuryl group, yielding a sulfodiester intermediate. Subsequent hydrolysis of this intermediate would give a second sulfoenzyme, the same as that generated by the catalytic pathway in **Figure 20(a)**. However, this mechanism does not explain the sulfoenzyme intermediate found in the C91S mutant of ASB, in which the Cys precursor to the functional FGly was mutated to Ser.[258]

For the second step, decomposition of the sulfoenzyme intermediate, two mechanisms have been proposed. The first is the elimination, shown in **Figure 20(c)**, in which deprotonation of the FGlyO$\gamma2$ hydroxyl group and S–O bond cleavage yields an aldehyde and inorganic sulfate. Subsequent regeneration of the active *gem*-diol state occurs by addition of water to the aldehyde.[62] The second proposed mechanism is hydrolysis of the sulfoenzyme to give inorganic sulfate and the hydrated *gem*-diol FGly directly (**Figure 20(d)**).[250] This pathway is supported by the
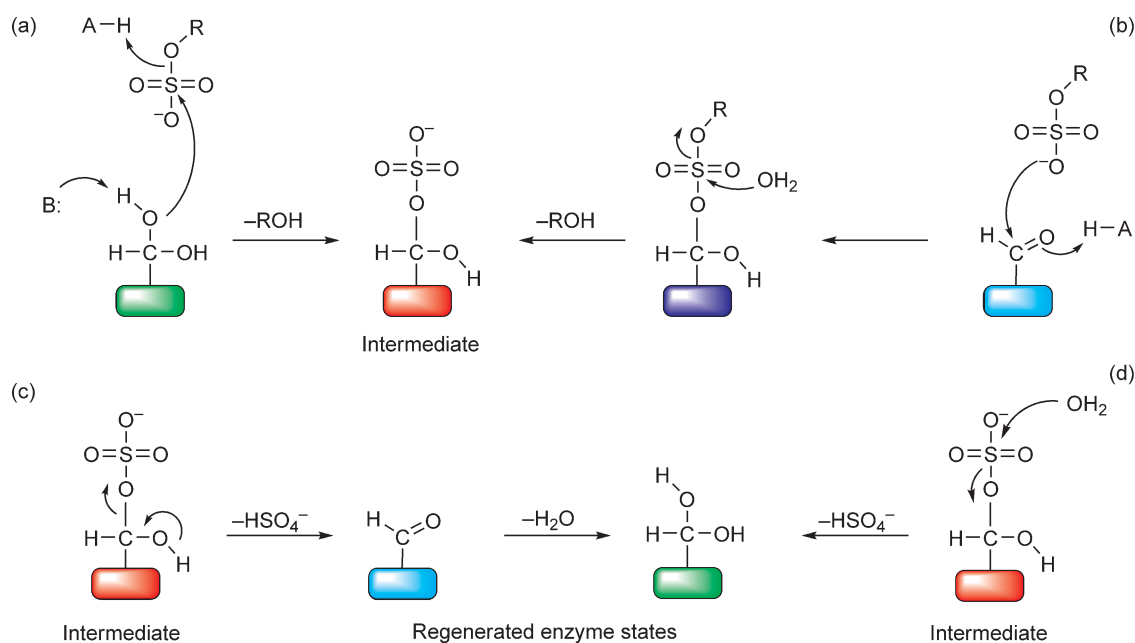
Figure 20   Possible mechanisms for sulfate ester hydrolysis catalyzed by members of the AS family. Formation of the sulfoenzyme: (a) nucleophilic attack of the hydrated FGly oxygen on the substrate; (b) nucleophilic attack of the sulfate oxygen of the substrate on the aldehyde FGly yields a diester sulfoenzyme, which hydrolyzes into the sulfoenzyme monoester. Enzyme desulfurylation: (c) elimination of inorganic sulfate leaves an aldehyde moiety, which is hydrated to give a *gem*-diol; (d) nucleophilic attack of a nucleophile, such as water, yields the hydrated *gem*-diol FGly.

sulfuryl transfer reaction from chiral phenyl sulfate to p-cresol catalyzed by the AS of *A. oryzae* described above.[257] However, this observation may not be pertinent to ASs in general, as the *A. oryzae* enzyme differs from most ASs in functioning as a much more effective sulfotransferase than a sulfatase in the presence of phenolic acceptors.[259] It has not been shown that this is a general property of sulfatases. A sulfurylated FGly has been observed in the crystal structures of ASB and ASC, suggesting that this sulfoenzyme may be the resting state of ASs.[247,252]

The pH–rate profiles have been determined for many ASs, and show a bell-shaped dependence that implies the participation of acid–base catalysis.[254,260,261] ASs fall into two groups according to their pH optima for V_{max}/K_m. Such pH–rate profiles reflect ionizable groups on free enzyme and free substrate that are important for binding and catalysis. Ionizations of the substrate will not appear in the pH profiles of sulfatases because of the very low pK_a value of the sulfuryl group, which is approximately -3. As a result, only the anion is present in solution under normal experimental conditions.

The lysosomal ASs, such as ASA and ASB, exhibit optimal activity between pH 4.0 and 5.7. A second group exhibits maximum activity between pH 7 and 8, which includes ASC (also known as steroid sulfatase, STS) and PAS. The pH-rate data indicate that, depending on the enzyme, the first pK_a varies from 3 to 7.1, while the second pK_a, presumably for the general acid, can range from 6.3 to 8.8.[253,254,256] Considering the fact that ASs share similar active-site residues and architectures, this variation is surprising and has not been explained. To date, the residues that function as acid and base catalysts have not been identified. Crystallographic studies have proposed that the same residues implicated in sulfuryl group binding may also be involved as general acid–base catalysts. In ASA, the conjugate base of Asp281 has been suggested to deprotonate the nucleophilic FGlyOγ1 atom, and His229 to accomplish proton transfer to the leaving-group oxygen. General acid catalysis was also implicated in the hydrolysis of substituted sulfate esters by the AS from *A. oryzae*, since V_{max} exhibited no dependence on leaving-group basicity.[262] However, a site-directed mutagenesis study of nine active-site residues in ASA, including the D281A and H229A mutants, found that V_{max} was decreased to between 1 and 26% of the wild-type activity by these as well as other mutations, and no specific role for any of the mutated residues could be assigned.[260]

Details about the transition state of the reaction catalyzed by ASs have been obtained through KIE experiments. For the hydrolysis of pNPS by the AS from *Helix pomatia*, there is no $^{15}(V/K)$ KIE and $^{18}(V/K)_{bridge} = 1.0136$. These magnitudes are consistent with a transition state in which S–O bond fission and protonation of the leaving group are both far advanced. A small normal value for $^{18}(V/K)_{nonbridge}$ of 1.0024 in the enzymatic reaction contrasts with the inverse value of 0.9951 in the uncatalyzed hydrolysis of pNPS, suggesting more nucleophilic participation than in the uncatalyzed reaction.[254]

8.09.5.2 Alkyl sulfatases

Alkyl sulfatases have been found in several lower eukaryotes and prokaryotes, and exhibit no sequence similarity with ASs. The alkyl sulfatases do not have an FGly residue, but bear in the active site a dinuclear Zn^{2+} center, and their sequences are related to those of the metallo-β-lactamase (MBL) superfamily (**Figure 21**).[263] Depending on the enzyme and catalytic mechanism the alkyl sulfatases can either cleave the

aa SdsA1	169 171 173 174	280	299	306 307 310 312	317	344	405	Accession code
Alkylsulfatases								
Pseudomonas aeruginosa (SdsA1)	..SHAHADHFGG..	..TESPAEMNIWLPRQKALLMAE	NVVGTL	HNLYTLRGAEVRDA...	VHNWP..	.GYHGSV..		Q9I5I9/2CFU
Vibrio vulnificus	..SHSHLDHFGG..	..TEAPTEINTWFPDKKALWMAE	NSTNTM	HNILTLRGAQVRDA...	SHHWP..	.GYYGTL..		Q7MCA6
Shewanella amazonensis	..SHSHVDHFGG..	..TEAPAEMNTLFPDFKALWMAE	NTTNTM	HNILTLRGAQVRDA...	SHHWP..	.GYYGTL..		A1S2W5
Ralstonia eutropha	..THSHGDHFGG..	..TEAPAEMLYFPQWKALCAAE	DATHNL	HNLYTIRGAQVRDA...	QHHWP..	.DYYGTV..		Q0K5J3
Synechococcus sp.	..SHSHADHFGG..	..TEAPAEMNTWFPQFKTFWAAE	NITGTI	HNIYTLRGALIRDP...	SHSWP..	.SYHGSE..		O7U9Y9
Saccharomyces cerevisiae (BdsA1)	..THSHSDHYGG..	..SEAPSEMLIYMPQQRVLNMAE	DVTHHM	HNLYALRGVEVRDG...	QHHWP..	.GYYGTL..		Q08347
Hydroxyacylglutathione hydrolase								
Homo sapiens	..THHHWDHAGG..	..CHTSGH..		..FTGDTLF..		..GHEYT..		Q16775/1QH5
Methyl parathion hydrolase								
Pseudomonas sp.	..THMHPDHVGG..	..GHTPGH..		..LLGDLIL..		..SHLSF..		Q841S6/1P9E
AiiA-like protein								
Bascillus thuringiensis	..SHDHFDHAGG..	..GHSPGH..		..LTIDASY..		..GHDIE..		Q7B8C3/2A7M
Beta-lactamase type II								
Bacteroides fragilis	..NHWHGDCIGG..	..GHATDN..		..FGGCMLK..		..GHGDY..		P25910/1A7T
Rubredoxin-oxygen oxidoreductase								
Desulfovibrio gigas	..QHLHLDHAGA..	..LHWPDS..		..ISNDIFG..		..DHGVI..		Q9F0J6/1E5D

Figure 21 Sequence alignment for conserved active-site residues in alkylsulfatases and homology to enzymes in the MBL superfamily. The numbering is that for SdsA1 from *Pseudomonas aeruginosa*. Red background indicates conserved residues involved in metal binding. Gray background indicates the residues involved in sulfate binding according to the structure of SdsA1 (see **Figure 22**). The accession codes refer to the ExPASy/PDB accession numbers.

S–O or C–O bond.[238] The C–O bond cleavage pathway has been demonstrated by labeling experiments using $H_2^{18}O$. When an alkyl sulfate ester was hydrolyzed by alkyl sulfatases, the ^{18}O was fully incorporated in the alcohol, while in the reactions of ASs incorporation occurs only in the inorganic sulfate product.[264,265] The cleavage of the C–O bond has been used for several biotransformation methods involving *sec*-alkyl sulfate esters.[266,267] The attack of water or hydroxide ion on the carbon atom involves inversion of configuration at carbon, in contrast to the attack on sulfur, which gives retention (**Scheme 6**). Thus, the goal is to achieve interconversion and enantiomeric excess of a chiral alcohol product. It has been speculated that the mechanism of C–O bond cleavage and inversion of configuration might involve a dinuclear Zn^{2+} center.[238] The classification of these enzymes and determination of the mechanism await additional sequence information and further kinetic studies.

A recent study of the extracellular alkylsulfatase SdsA1 from *P. aeruginosa* shows that the binuclear Zn^{2+} center is situated at the bottom of a narrow, approximately 23 Å deep hydrophobic groove. This alkylsulfatase cannot hydrolyze arylsulfate esters, but has the ability to hydrolyze long-chain alkyl esters such as the detergent SDS.[263] It has been postulated that the ability of *Pseudomonas* sp. to grow in SDS as the only sulfur source is due to the presence of this enzyme. Based on crystal structures, one with a bound substrate analogue and one with products, the mechanism has been proposed to occur by S–O bond cleavage. The sulfonyl moiety in this structure hydrogen bonds to several polar residues at the bottom of the groove, but is not coordinated to the metal ions. A mechanism was proposed involving nucleophilic attack from an activated water molecule in the second hydration shell of one of the Zn^{2+} atoms (W2 in **Figure 22**).[263] However, the crystal structure indicates that the nucleophilic water molecule is adjacent to the leaving group, not in-line, and thus would require an addition–elimination ($A_N + D_N$) mechanism. If sulfur follows the same rules in this regard as phosphorus, this implies the reaction should occur through a pentacoordinate intermediate involving pseudorotation (see Section 8.09.1.2). It is possible that the crystal structures do not represent a catalytically active state. Studies to date have investigated only primary sulfate esters, and thus the stereochemical features in relation to retention or inversion at the carbon center (which would reveal the presence of a C–O pathway; **Scheme 6**) are unknown.

An unusual sulfatase mechanism is observed for alkyl sulfatases from the nonheme Fe^{2+}, α-ketoglutarate-dependent dioxygenase superfamily.[268–270] These enzymes cleave a variety of sulfate esters by a complex radical mechanism, yielding the corresponding aldehyde and inorganic sulfate. The crystal structure of AtsK from *Pseudomonas putida*, a member of this group of enzymes, has been solved and a mechanism proposed
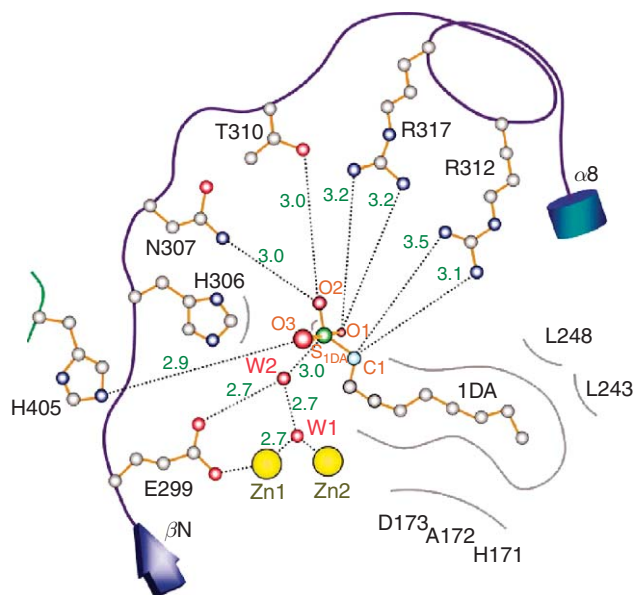
Figure 22 Schematic representation of the active site of the alkylsulfatase SdsA1 from *Pseudomonas aeruginosa* bound with 1-decane-sulfonate. H bonds and salt bridges are marked by dotted lines and hydrophobic interactions by gray arcs. The distances are in angstroms.[263] Reproduced with permission from G. Hagelueken; T. M. Adams; L. Wiehlmann; L. Widow; H. Kolmar; B. Tummler; D. W. Heinz; W. D. Schubert, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7631–7636.
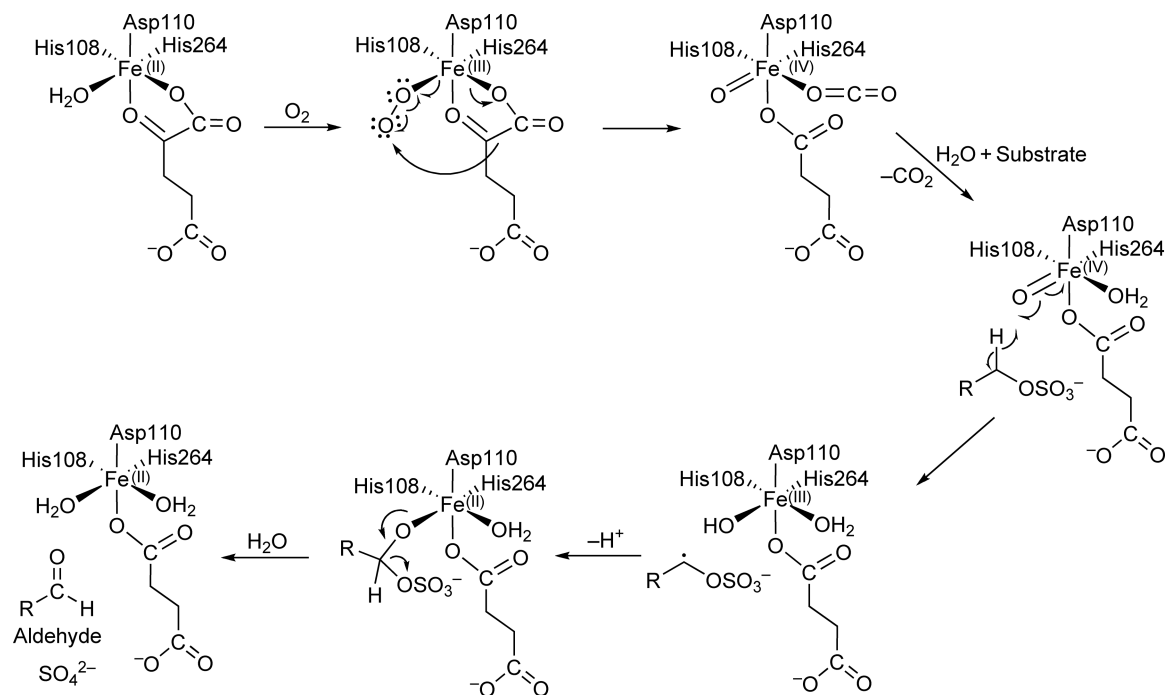
Figure 23 A proposed catalytic mechanism of sulfate ester hydrolysis by the alkysulfatase AtsK from *Pseudomonas putida*.[270]

(**Figure 23**). The iron center in the active site is able to coordinate molecular oxygen, which through a radical pathway reacts with the metal-bound α-ketoglutarate giving carbon dioxide, succinate, and a highly reactive Fe^{4+} species. The subsequent reaction occurs by a radical mechanism that ultimately yields an aldehyde, inorganic sulfate, and the native Fe^{2+} state of the enzyme. This mechanism is not well characterized and only a few studies have been conducted.[268–270]

8.09.6 Conclusions

The importance of phosphatases in biological chemistry has been appreciated for many years. This awareness has grown even more since the 1980s with the realization of the role that phosphoryl transfer on proteins, mediated by phosphatases and kinases, performs in the regulation of a host of biochemical processes. While some of these enzymes are fairly well understood mechanistically, knowledge of biological substrates and the roles of particular phosphatases remain sketchy. Considerably less work has been done on sulfatases, and a great deal remains to be learned about their chemistry and biochemistry. It is interesting that despite the resemblance between the phosphoryl and sulfuryl group, and the high degree of mechanistic similarity in reactions of phosphate and sulfate monoesters, disparate enzymes have evolved a high degree of specificity for the two activities. Future inquiry by chemists and biochemists into this phenomenon might provide insights into how enzymes evolve new activities and specificity for their substrates.

Abbreviation

EPR	electron paramagnetic resonance
KIE	kinetic isotope effect
LFER	linear free energy relationships
mNBP	*m*-nitrobenzyl phosphate
pNPP	*p*-nitrophenyl phosphate

References

1. F. H. Westheimer, *Science* **1987**, *235*, 1173–1178.
2. P. Cohen, *Eur. J. Biochem.* **2001**, *268*, 5001–5010.
3. C. Lad; N. H. Williams; R. Wolfenden, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5607–5610.
4. R. Wolfenden; Y. Yuan, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 83–86.
5. R. D. Guthrie; W. P. Jencks, *Acc. Chem. Res.* **1989**, *22*, 343–349.
6. R. S. Berry, *J. Chem. Phys.* **1960**, *32*, 933–938.
7. N. V. Timosheva; A. Chandrasekaran; R. R. Holmes, *Inorg. Chem.* **2006**, *45*, 3113–3123, 10836–10848.
8. R. R. Holmes, *J. Am. Chem. Soc.* **1978**, *100*, 433–446.
9. F. H. Westheimer, *Acc. Chem. Res.* **1968**, *1*, 70–78.
10. R. R. Holmes, *Acc. Chem. Res.* **2004**, *37*, 746–753.
11. R. A. Moore O'Ferrall, *J. Chem. Soc. B* **1970**, 274–277.
12. W. P. Jencks, *Chem. Rev.* **1972**, *72*, 705–718.
13. A. Williams, *Free Energy Relationships*; Royal Society of Chemistry: Cambridge, UK, 2003.
14. W. W. Cleland; A. C. Hengge, *Chem. Rev.* **2006**, *106*, 3252–3278.
15. L. Melander; W. H. Saunders, *Reaction Rates of Isotopic Molecules*; Robert E. Krieger: Malabar, FL, 1987.
16. A. C. Hengge, *Acc. Chem. Res.* **2002**, *35*, 105–112.
17. E. V. Anslyn; D. A. Dougherty, *Modern Physical Organic Chemistry*; University Science Books: Sausalito, CA, 2006.
18. J. A. Gutierrez; M. Luo; V. Singh; L. Li; R. L. Brown; G. E. Norris; G. B. Evans; R. H. Furneaux; P. C. Tyler; G. F. Painter; D. H. Lenz; V. L. Schramm, *ACS Chem. Biol.* **2007**, *2*, 725–734.
19. P. A. Frey; J. P. Richard; H.-T. Ho; R. S. Brody; R. D. Sammons; K.-F. Sheu, Stereochemistry of Selected Phosphotransferases and Nucleotidyltransferases. In *Methods of Enzymology*; D. L. Purich, Ed.; Academic Press: New York, 1982; Vol. 87, pp 213–235.
20. F. Eckstein; P. J. Romaniuk; B. A. Connolly, Stereochemistry of Enzymic Phosphoryl and Nucleotidyl Transfer. In *Methods of Enzymology*; D. L. Purich, Ed.; Academic Press: New York, 1982; Vol. 87, pp 197–212.
21. C. Hengge. In *Comprehensive Biological Catalysis: A Mechanistic Reference*; M. Sinnott, Ed.; Academic Press: San Diego, CA, 1998; Vol. 1, pp 517–542.
22. A. J. Kirby; A. G. Varvoglis, *J. Am. Chem. Soc.* **1967**, *89*, 415–423.
23. G. R. J. Thatcher; R. Kluger, *Adv. Phys. Org. Chem.* **1989**, *25*, 99–265.
24. A. C. Hengge; W. A. Edens; H. Elsing, *J. Am. Chem. Soc.* **1994**, *116*, 5045–5049.
25. D. G. Gorenstein; Y.-G. Lee; D. Kar, *J. Am. Chem. Soc.* **1977**, *99*, 2264–2267.
26. D. Herschlag; W. P. Jencks, *J. Am. Chem. Soc.* **1989**, *111*, 7579–7586.
27. S. L. Buchwald; J. M. Friedman; J. R. Knowles, *J. Am. Chem. Soc.* **1984**, *106*, 4911–4916.
28. J. M. Friedman; S. Freeman; J. R. Knowles, *J. Am. Chem. Soc.* **1988**, *110*, 1268–1275.
29. R. H. Hoff; A. C. Hengge, *J. Org. Chem.* **1998**, *63*, 6680–6688.
30. A. J. Kirby; W. P. Jencks, *J. Am. Chem. Soc.* **1965**, *87*, 3209–3216.
31. J. Aqvist; K. Kolmodin; J. Florian; A. Warshel, *Chem. Biol.* **1999**, *6*, R71–80.
32. M. Bianciotto; J.-C. Barthelat; A. Vigroux, *J. Am. Chem. Soc.* **2002**, *124*, 7573–7587.
33. M. Bianciotto; J.-C. Barthelat; A. Vigroux, *J. Phys. Chem. A* **2002**, *106*, 6521–6526.
34. A. Dejaegere; M. Karplus, *J. Am. Chem. Soc.* **1993**, *115*, 5316–5317.
35. J. Florian; J. Aqvist; A. Warshel, *J. Am. Chem. Soc.* **1998**, *120*, 11524–11525.
36. J. Florian; A. Warshel, *J. Phys. Chem. B* **1998**, *102*, 719–734.
37. C.-H. Hu; T. Brinck, *J. Phys. Chem. A* **1999**, *103*, 5379–5386.
38. X. Lopez; A. Dejaegere; M. Karplus, *J. Am. Chem. Soc.* **2001**, *123*, 11755–11763.
39. Y. N. Wang; I. A. Topol; J. R. Collins; S. K. Burt, *J. Am. Chem. Soc.* **2003**, *125*, 13265–13273.
40. L. Zhang; D. Xie; D. Xu; H. Guo, *Chem. Commun.* **2007**, 1638–1640.
41. M. Klahn; E. Rosta; A. Warshel, *J. Am. Chem. Soc.* **2006**, *128*, 15310–15323.
42. P. K. Grzyska; P. G. Czyryca; J. Purcell; A. C. Hengge, *J. Am. Chem. Soc.* **2003**, *125*, 13106–13111.
43. W. B. Knight; P. M. Weiss; W. W. Cleland, *J. Am. Chem. Soc.* **1986**, *108*, 2759–2761.
44. C. A. Bunton; D. R. Llewellyn; K. G. Oldham; C. A. Vernon, *J. Chem. Soc.* **1958**, 3574–3587.
45. S. J. Admiraal; D. Herschlag, *J. Am. Chem. Soc.* **2000**, *122*, 2145–2148.
46. S. J. Benkovic; P. A. Benkovic, *J. Am. Chem. Soc.* **1966**, *88*, 5504–5511.
47. E. J. Fendler; J. H. Fendler, *J. Org. Chem.* **1968**, *33*, 3852–3859.
48. N. Bourne; A. Hopkins; A. Williams, *J. Am. Chem. Soc.* **1985**, *107*, 4327–4331.
49. P. D'Rozario; R. L. Smyth; A. Williams, *J. Am. Chem. Soc.* **1984**, *106*, 5027–5028.
50. R. H. Hoff; P. Larsen; A. C. Hengge, *J. Am. Chem. Soc.* **2001**, *123*, 9338–9344.
51. B. T. Burlingham; L. M. Pratt; E. R. Davidson; V. J. J. Shiner; J. Fong; T. S. Widlanski, *J. Am. Chem. Soc.* **2003**, *125*, 13036–13037.
52. C. L. L. Chai; T. W. Hepburn; G. Lowe, *J. Chem. Soc. Chem. Commun.* **1991**, *19*, 1403–1405.
53. S. Burstein; S. Lieberman, *J. Am. Chem. Soc.* **1958**, *80*, 5235–5239.
54. J. L. Kice; J. M. Anderson, *J. Am. Chem. Soc.* **1966**, *88*, 5242–5245.
55. H. K. Garner; H. J. Lucas, *J. Am. Chem. Soc.* **1950**, *72*, 5497–5501.
56. E. T. Kaiser; M. Panar; F. H. Westheimer, *J. Am. Chem. Soc.* **1963**, *85*, 602–607.
57. D. Bethell; R. E. Fessey; E. Namwindwa; D. W. Roberts, *J. Chem. Soc. Perkin Trans.* **2001**, *2*, 1489–1495.
58. R. L. Burwell, *J. Am. Chem. Soc.* **1952**, *74*, 1462–1466.
59. R. D. Finn; J. Mistry; B. Schuster-Bockler; S. Griffiths-Jones; V. Hollich; T. Lassmann; S. Moxon; M. Marshall; A. Khanna; R. Durbin; S. R. Eddy; E. L. Sonnhammer; A. Bateman, *Nucleic Acids Res.* **2006**, *34*, D247–251.
60. Z. Lu; D. Dunaway-Mariano; K. N. Allen, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5687–5692.

61. P. J. O'Brien; D. Herschlag, *J. Am. Chem. Soc.* **1998**, *120*, 12369–12370.
62. I. Boltes; H. Czapinska; A. Kahnert; R. von Bulow; T. Dierks; B. Schmidt; K. von Figura; M. A. Kertesz; I. Uson, *Structure* **2001**, *9*, 483–491.
63. M. Y. Galperin; M. J. Jedrzejas, *Proteins* **2001**, *45*, 318–324.
64. L. Holm; C. Sander, *Science* **1996**, *273*, 595–602.
65. J. E. Murphy; B. Stec; L. Ma; E. R. Kantrowitz, *Nat. Struct. Biol.* **1997**, *4*, 618–622.
66. C. S. Bond; P. R. Clements; S. J. Ashby; C. A. Collyer; S. J. Harrop; J. J. Hopwood; J. M. Guss, *Structure* **1997**, *5*, 277–289.
67. T. Uchida; F. Egami; A. B. Roy, *Biochim. Biophys. Acta* **1981**, *657*, 356–363.
68. J. A. Peliska; M. H. O'Leary, *Biochemistry* **1991**, *30*, 1049–1057.
69. J. A. Peliska; M. H. O'Leary, *Biochemistry* **1989**, *28*, 1604–1611.
70. A. Hasset; W. Blattler; J. R. Knowles, *Biochemistry* **1982**, *21*, 6335–6340.
71. J. Knowles, *Annu. Rev. Biochem.* **1980**, *49*, 877–919.
72. I. Catrina; P. J. O'Brien; J. Purcell; I. Nikolic-Hughes; J. G. Zalatan; A. C. Hengge; D. Herschlag, *J. Am. Chem. Soc.* **2007**, *129*, 5760–5765.
73. P. J. O'Brien; D. Herschlag, *Biochemistry* **2002**, *41*, 3207–3225.
74. J. G. Zalatan; I. Catrina; R. Mitchell; P. K. Grzyska; P. J. O'Brien; D. Herschlag; A. C. Hengge, *J. Am. Chem. Soc.* **2007**, *129*, 9789–9798.
75. R. H. Hoff; P. Mertz; F. Rusnak; A. C. Hengge, *J. Am. Chem. Soc.* **1999**, *121*, 6382–6390.
76. A. C. Hengge; B. L. Martin, *Biochemistry* **1997**, *36*, 10185–10191.
77. K. M. Holtz; E. R. Kantrowitz, *FEBS Lett.* **1999**, *462*, 7–11.
78. J. E. Coleman, *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 441–483.
79. T. W. Reid; I. B. Wilson, *E.Coli* Alkaline Phosphatase. In *The Enzymes*; P. D. Boyer, Ed.; Academic Press: New York, 1971; pp 373–415.
80. E. E. Kim; H. W. Wyckoff, *J. Mol. Biol.* **1991**, *218*, 449–464.
81. B. Stec; K. M. Holtz; E. R. Kantrowitz, *J. Mol. Biol.* **2000**, *299*, 1303–1311.
82. K. M. Holtz; B. Stec; E. R. Kantrowitz, *J. Biol. Chem.* **1999**, *274*, 8351–8354.
83. J. Bale; C. Huang; P. Chock, *J. Biol. Chem.* **1980**, *255*, 8431–8436.
84. W. Bloch; M. J. Schlesinger, *J. Biol. Chem.* **1973**, *248*, 5794–5805.
85. T. T. Simopoulos; W. P. Jencks, *Biochemistry* **1994**, *33*, 10375–10380.
86. R. Han; J. E. Coleman, *Biochemistry* **1995**, *34*, 4238–4245.
87. A. Chaidaroglou; D. J. Brezinski; S. A. Middleton; E. R. Kantrowitz, *Biochemistry* **1988**, *27*, 8338–8343.
88. J. E. Butler-Ransohoff; S. E. Rokita; D. A. Kendall; J. A. Banzon; K. S. Carano; E. T. Kaiser; A. R. Matlin, *J. Org. Chem.* **1992**, *57*, 142–145.
89. P. J. O'Brien; D. Herschlag, *J. Am. Chem. Soc.* **1999**, *121*, 11022–11023.
90. I. Nikolic-Hughes; J. P. O'Brien; D. Herschlag, *J. Am. Chem. Soc.* **2005**, *127*, 9314–9315.
91. M. S. Saini; S. L. Buchwald; R. L. Van Etten; J. R. Knowles, *J. Biol. Chem.* **1981**, *256*, 10453–10455.
92. R. L. Van Etten; M. E. Hickey, *Arch. Biochem. Biophys.* **1977**, *183*, 250–259.
93. K. Ostanin; E. H. Harms; P. E. Stevis; R. Kuciel; M. M. Zhou; R. L. Van Etten, *J. Biol. Chem.* **1992**, *267*, 22830–22836.
94. K. Ishikawa; Y. Mihara; K. Gondoh; E. Suzuki; Y. Asano, *EMBO J.* **2000**, *19*, 2412–2423.
95. D. Kostrewa; M. Wyss; A. D'Arcy; A. P. van Loon, *J. Mol. Biol.* **1999**, *288*, 965–974.
96. C. G. Jakob; K. Lewinski; R. Kuciel; W. Ostrowski; L. Lebioda, *Prostate* **2000**, *42*, 211–218.
97. Y. Lindqvist; G. Schneider; P. Wihko, *Eur. J. Biochem.* **1994**, *221*, 139–142.
98. K. S. Porvari; A. M. Herrala; R. M. Kurkela; P. A. Taavitsainen; Y. Lindqvist; G. Schneider; P. T. Vihko, *J. Biol. Chem.* **1994**, *269*, 22642–22646.
99. K. Ostanin; R. L. Van Etten, *J. Biol. Chem.* **1993**, *268*, 20778–20784.
100. M. S. Saini; R. L. Van Etten, *Biochim. Biophys. Acta* **1979**, *568*, 370–376.
101. J. B. Vincent; G. L. Olivier-Lilley; B. A. Averill, *Chem. Rev.* **1990**, *90*, 1447–1467.
102. N. Mitic; S. J. Smith; A. Neves; L. W. Guddat; L. R. Gahan; G. Schenk, *Chem. Rev.* **2006**, *106*, 3338–3363.
103. G. W. Oddie; G. Schenk; N. Z. Angel; N. Walsh; L. W. Guddat; J. de Jersey; A. I. Cassady; S. E. Hamilton; D. A. Hume, *Bone* **2000**, *27*, 575–584.
104. B. Ek-Rylander; M. Flores; M. Wendel; D. Heinegard; G. Andersson, *J. Biol. Chem.* **1994**, *269*, 14853–14856.
105. D. W. Moss; F. D. Raymond; D. B. Wile, *Crit. Rev. Clin. Lab. Sci.* **1995**, *32*, 431–467.
106. H. Liao; F. L. Wong; T. H. Phang; M. Y. Cheung; W. Y. Li; G. Shao; X. Yan; H. M. Lam, *Gene* **2003**, *318*, 103–111.
107. P. Zimmermann; B. Regierer; J. Kossmann; E. Frossard; N. Amrhein; M. Bucher, *Plant Biol.* **2004**, *6*, 519–528.
108. N. Sträter; T. Klabunde; P. Tucker; H. Witzel; B. Krebs, *Science* **1995**, *268*, 1489–1492.
109. Y. Lindqvist; E. Johansson; H. Kaija; P. Vihko; G. Schneider, *J. Mol. Biol.* **1999**, *291*, 135–147.
110. E. G. Mueller; M. W. Crowder; B. A. Averill; J. R. Knowles, *J. Am. Chem. Soc.* **1993**, *115*, 2974–2975.
111. N. Sträter; W. N. Lipscomb; T. Klabunde; B. Krebs, *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 2024–2055.
112. F. Rusnak; L. Yu; P. Mertz, *J. Biol. Inorg. Chem.* **1996**, *1*, 388–396.
113. S. K. Smoukov; L. Quaroni; X. Wang; P. E. Doan; B. M. Hoffman; L. Que, Jr., *J. Am. Chem. Soc.* **2002**, *124*, 2595–2603.
114. M. Merkx; M. W. Pinkse; B. A. Averill, *Biochemistry* **1999**, *38*, 9914–9925.
115. M. Merkx; B. A. Averill, *J. Am. Chem. Soc.* **1999**, *121*, 6683–6689.
116. E. G. Funhoff; C. H. Klaassen; B. Samyn; J. Van Beeumen; B. A. Averill, *Chembiochem* **2001**, *2*, 355–363.
117. M. Valizadeh; G. Schenk; K. Nash; G. W. Oddie; L. W. Guddat; D. A. Hume; J. de Jersey; T. R. Burke, Jr.; S. Hamilton, *Arch. Biochem. Biophys.* **2004**, *424*, 154–162.
118. G. Schenk; L. R. Gahan; L. E. Carrington; N. Mitic; M. Valizadeh; S. E. Hamilton; J. de Jersey; L. W. Guddat, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 273–278.
119. A. Dikiy; E. G. Funhoff; B. A. Averill; S. Ciurli, *J. Am. Chem. Soc.* **2002**, *124*, 13974–13975.
120. T. Klabunde; N. Sträter; R. Frohlich; H. Witzel; B. Krebs, *J. Mol. Biol.* **1996**, *259*, 737–748.

121. E. G. Funhoff; Y. Wang; G. Andersson; B. A. Averill, *FEBS J.* **2005**, *272*, 2968–2977.
122. N. T. Truong; J. I. Naseri; A. Vogel; A. Rompel; B. Krebs, *Arch. Biochem. Biophys.* **2005**, *440*, 38–45.
123. E. G. Funhoff; J. Ljusberg; Y. Wang; G. Andersson; B. A. Averill, *Biochemistry* **2001**, *40*, 11614–11622.
124. N. Mitic; M. Valizadeh; E. W. Leung; J. de Jersey; S. Hamilton; D. A. Hume; A. I. Cassady; G. Schenk, *Arch. Biochem. Biophys.* **2005**, *439*, 154–164.
125. D. Barford; A. K. Das; M. P. Egloff, *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 133–164.
126. M. D. Jackson; J. M. Denu, *Chem. Rev.* **2001**, *101*, 2313–2340.
127. D. L. Lohse; J. M. Denu; J. E. Dixon, *Structure* **1995**, *3*, 987–990.
128. S. Q. Zhuo; J. C. Clemens; R. L. Stone; J. E. Dixon, *J. Biol. Chem.* **1994**, *269*, 26234–26238.
129. W. C. Voegtli; D. J. White; N. J. Reiter; F. Rusnak; A. C. Rosenzweig, *Biochemistry* **2000**, *39*, 15365–15374.
130. P. T. W. Cohen, *J. Cell Sci.* **2002**, *115*, 241–256.
131. J. B. Aggen; A. C. Nairn; R. Chamberlin, *Chem. Biol.* **2000**, *7*, R13–R23.
132. F. Rusnak; P. Mertz, *Physiol. Rev.* **2000**, *80*, 1483–1521.
133. G. R. Crabtree; N. A. Clipstone, *Annu. Rev. Biochem.* **1994**, *63*, 1045–1083.
134. F. J. Dumont, *Curr. Med. Chem.* **2000**, *7*, 731–748.
135. M. C. Louzao; M. R. Vieytes; L. M. Botana, *Mini Rev. Med. Chem.* **2005**, *5*, 207–215.
136. J. J. Fernandez; M. L. Candenas; M. L. Souto; M. M. Trujillo; M. Norte, *Curr. Med. Chem.* **2002**, *9*, 229–262.
137. L. Herfindal; F. Selheim, *Mini Rev. Med. Chem.* **2006**, *6*, 279–285.
138. A. K. Das; N. R. Helps; P. T. W. Cohen; D. Barford, *EMBO J.* **1996**, *15*, 6798–6809.
139. P. Bork; N. P. Brown; H. Hegyi; J. Schultz, *Protein Sci.* **1996**, *5*, 1421–1425.
140. A. Schweighofer; H. Hirt; L. Meskiene, *Trends Plant Sci.* **2004**, *9*, 236–243.
141. M. P. Egloff; P. T. W. Cohen; P. Reinemer; D. Barford, *J. Mol. Biol.* **1995**, *254*, 942–959.
142. J. Goldberg; H. B. Huang; Y. G. Kwon; P. Greengard; A. C. Nairn; J. Kuriyan, *Nature* **1995**, *376*, 745–753.
143. T. Maynes; K. S. Bateman; M. M. Cherney; A. K. Das; H. A. Luu; C. F. B. Holmes; M. N. G. James, *J. Biol. Chem.* **2001**, *276*, 44078–44082.
144. Y. Xing; Y. H. Xu; Y. Chen; P. D. Jeffrey; Y. Chao; Z. Lin; Z. Li; S. Strack; J. B. Stock; Y. G. Shi, *Cell* **2006**, *127*, 341–353.
145. J. P. Griffith; J. L. Kim; E. E. Kim; M. D. Sintchak; J. A. Thomson; M. J. Fitzgibbon; M. A. Fleming; P. R. Caron; K. Hsiao; M. A. Navia, *Cell* **1995**, *82*, 507–522.
146. Q. Huai; H. Y. Kim; Y. D. Liu; Y. D. Zhao; A. Mondragon; J. O. Liu; H. M. Ke, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12037–12042.
147. C. R. Kissinger; H. E. Parge; D. R. Knighton; C. T. Lewis; L. A. Pelletier; A. Tempczyk; V. J. Kalish; K. D. Tucker; R. E. Showalter; E. W. Moomaw; L. N. Gastinel; N. Habuka; X. H. Chen; F. Maldonado; J. E. Barker; R. Bacquet; J. E. Villafranca, *Nature* **1995**, *378*, 641–644.
148. M. Terrak; F. Kerff; K. Langsetmo; T. Tao; R. Dominguez, *Nature* **2004**, *429*, 780–784.
149. K. Takeuchi; M. H. A. Roehrl; Z. Y. J. Sun; G. Wagner, *Structure* **2007**, *15*, 587–597.
150. L. Yu; A. Haddy; F. Rusnak, *J. Am. Chem. Soc.* **1995**, *117*, 10147–10148.
151. B. L. Martin; C. J. Pallen; J. H. Wang; D. J. Graves, *J. Biol. Chem.* **1985**, *260*, 14932–14937.
152. B. L. Martin; D. J. Graves, *J. Biol. Chem.* **1986**, *261*, 14545–14550.
153. B. L. Martin; D. J. Graves, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1994**, *1206*, 136–142.
154. H. B. Huang; A. Horiuchi; J. Goldberg; P. Greengard; A. C. Nairn, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 3530–3535.
155. J. Zhang; Z. J. Zhang; K. Brew; E. Y. C. Lee, *Biochemistry* **1996**, *35*, 6276–6282.
156. A. Mondragon; E. C. Griffith; L. Sun; F. Xiong; C. Armstrong; J. O. Liu, *Biochemistry* **1997**, *36*, 4934–4942.
157. P. Mertz; L. Yu; R. Sikkink; F. Rusnak, *J. Biol. Chem.* **1997**, *272*, 21296–21302.
158. Y. F. Keng; L. Wu; Z. Y. Zhang, *Eur. J. Biochem.* **1999**, *259*, 809–814.
159. B. L. Martin; L. A. Jurado; A. C. Hengge, *Biochemistry* **1999**, *38*, 3386–3392.
160. C. C. Fjeld; J. M. Denu, *J. Biol. Chem.* **1999**, *274*, 20336–20343.
161. A. Wehenkel; M. Bellinzoni; F. Schaeffer; A. Villarino; P. M. Alzari, *J. Mol. Biol.* **2007**, *374*, 890–898.
162. M. D. Jackson; C. C. Fjeld; J. M. Denu, *Biochemistry* **2003**, *42*, 8513–8521.
163. M. Bellinzoni; A. Welhenkel; W. Shepard; P. M. Alzari, *Structure* **2007**, *15*, 863–872.
164. M. K. Rantanen; L. Lehtio; L. Rajagopal; C. E. Rubens; A. Goldman, *FEBS J.* **2007**, *274*, 3128–3137.
165. K. E. Pullen; H. L. Ng; P. Y. Sung; M. C. Good; S. M. Smith; T. Alber, *Structure* **2004**, *12*, 1947–1954.
166. Z. Y. Zhang, *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 209–234.
167. L. W. Li; J. E. Dixon, *Semin. Immunol.* **2000**, *12*, 75–84.
168. Z. Y. Zhang, *Crit. Rev. Biochem. Mol. Biol.* **1998**, *33*, 1–52.
169. A. Alonso; J. Sasin; N. Bottini; I. Friedberg; I. Friedberg; A. Osterman; A. Godzik; T. Hunter; J. Dixon; T. Mustelin, *Cell* **2004**, *117*, 699–711.
170. T. Hunter, *Cell* **2000**, *100*, 113–127.
171. D. Nandan; N. E. Reiner, *Clin. Immunol.* **2005**, *114*, 266–277.
172. R. DeVinney; O. Steele-Mortimer; B. B. Finlay, *Trends Microbiol.* **2000**, *8*, 29–33.
173. L. Bialy; H. Waldmann, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 3814–3839.
174. J. A. Stuckey; H. L. Schubert; E. B. Fauman; Z. Y. Zhang; J. E. Dixon; M. A. Saper, *Nature* **1994**, *370*, 571–575.
175. H. L. Schubert; E. B. Fauman; J. A. Stuckey; J. E. Dixon; M. A. Saper, *Protein Sci.* **1995**, *4*, 1904–1913.
176. Z. Y. Zhang; Y. A. Wang; J. E. Dixon, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1624–1627.
177. E. B. Fauman; C. Yuvaniyama; H. L. Schubert; J. A. Stuckey; M. A. Saper, *J. Biol. Chem.* **1996**, *271*, 18780–18788.
178. Z. Y. Zhang; W. P. Malachowski; R. L. Vanetten; J. E. Dixon, *J. Biol. Chem.* **1994**, *269*, 8140–8145.
179. D. Barford; A. J. Flint; N. K. Tonks, *Science* **1994**, *263*, 1397–1404.
180. A. K. Pedersen; X. L. Guo; K. B. Moller; G. H. Peters; H. S. Andersen; J. S. Kastrup; S. B. Mortensen; L. F. Iversen; Z. Y. Zhang; N. P. H. Moller, *Biochem. J.* **2004**, *378*, 421–433.
181. A. K. Pedersen; G. H. Peters; K. B. Moller; L. F. Iversen; J. S. Kastrup, *Acta Crystallogr., Sect D: Biol. Crystallogr.* **2004**, *60*, 1527–1534.

182. A. D. B. Pannifer; A. J. Flint; N. K. Tonks; D. Barford, *J. Biol. Chem.* **1998**, *273*, 10454–10462.
183. L. Wu; Z. Y. Zhang, *Biochemistry* **1996**, *35*, 5426–5434.
184. J. M. Denu; J. E. Dixon, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5910–5914.
185. J. M. Denu; G. C. Zhou; Y. P. Guo; J. E. Dixon, *Biochemistry* **1995**, *34*, 3396–3403.
186. P. J. Yuvaniyama; J. M. Denu; J. E. Dixon; M. A. Saper, *Science* **1996**, *272*, 1328–1331.
187. P. J. Ala; L. Gonneville; M. C. Hillman; M. Becker-Pasha; M. Wei; B. G. Reid; R. Klabe; E. W. Yue; B. Wayland; B. Douty; P. Polam; Z. Wasserman; M. Bower; A. P. Combs; T. C. Burn; G. F. Hollis; R. Wynn, *J. Biol. Chem.* **2006**, *281*, 32784–32795.
188. M. R. Groves; Z. J. Yao; P. P. Roller; T. R. Burke; D. Barford, *Biochemistry* **1998**, *37*, 17773–17783.
189. Y. L. Zhang; F. Hollfelder; S. J. Gordon; L. Chen; Y. F. Keng; L. Wu; D. Herschlag; Z. Y. Zhang, *Biochemistry* **1999**, *38*, 12111–12123.
190. A. K. H. Hirsch; F. R. Fischer; F. Diederich, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 338–352.
191. R. Agarwal; S. K. Burley; S. Swaminathan, *J. Biol. Chem.* **2008**, *283*, 8946–8953.
192. J. Phan; K. Lee; S. Cherry; J. E. Tropea; T. R. Burke; D. S. Waugh, *Biochemistry* **2003**, *42*, 13113–13121.
193. M. I. Ivanov; J. A. Stuckey; H. L. Schubert; M. A. Saper; J. B. Bliska, *Mol. Microbiol.* **2005**, *55*, 1346–1356.
194. Z. C. Jia; D. Barford; A. J. Flint; N. K. Tonks, *Science* **1995**, *268*, 1754–1758.
195. A. Salmeen; J. N. Andersen; M. P. Myers; N. K. Tonks; D. Barford, *Mol. Cell* **2000**, *6*, 1401–1412.
196. M. Sarmiento; Y. A. Puius; S. W. Vetter; Y. F. Keng; L. Wu; Y. Zhao; D. S. Lawrence; S. C. Almo; Z. Y. Zhang, *Biochemistry* **2000**, *39*, 8171–8179.
197. M. A. Schumacher; J. L. Todd; A. E. Rice; K. G. Tanner; J. M. Denu, *Biochemistry* **2002**, *41*, 3009–3017.
198. Z. Y. Zhang; A. M. Thiemesefler; D. Maclean; D. J. Mcnamara; E. M. Dobrusin; T. K. Sawyer; J. E. Dixon, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4446–4450.
199. Z. Y. Zhang; D. Maclean; D. J. Mcnamara; T. K. Sawyer; J. E. Dixon, *Biochemistry* **1994**, *33*, 2285–2290.
200. J. Rudolph, *Biochemistry* **2007**, *46*, 3595–3604.
201. Y. Zhao; Z. Y. Zhang, *J. Biol. Chem.* **2001**, *276*, 32382–32391.
202. Z. Y. Zhang; J. E. Dixon, *Biochemistry* **1993**, *32*, 9340–9345.
203. Z. Y. Zhang; B. A. Palfey; L. Wu; Y. Zhao, *Biochemistry* **1995**, *34*, 16389–16396.
204. K. L. Guan; J. E. Dixon, *J. Biol. Chem.* **1991**, *266*, 17026–17030.
205. R. H. Hoff; L. Wu; B. Zhou; Z. Y. Zhang; A. C. Hengge, *J. Am. Chem. Soc.* **1999**, *121*, 9514–9521.
206. R. H. Hoff; A. C. Hengge; L. Wu; Y. F. Keng; Z. Y. Zhang, *Biochemistry* **2000**, *39*, 46–54.
207. L. J. Juszczak; Z. Y. Zhang; L. Wu; D. S. Gottfried; D. D. Eads, *Biochemistry* **1997**, *36*, 2227–2236.
208. M. Khajehpour; L. Wu; S. J. Liu; N. Zhadin; Z. Y. Zhang; R. Callender, *Biochemistry* **2007**, *46*, 4370–4378.
209. F. Wang; W. Q. Li; M. R. Emmett; C. L. Hendrickson; A. G. Marshall; Y. L. Zhang; L. Wu; Z. Y. Zhang, *Biochemistry* **1998**, *37*, 15289–15299.
210. X. Hu; C. E. Stebbins, *Biophys. J.* **2006**, *91*, 948–956.
211. D. F. McCain; P. K. Grzyska; L. Wu; A. C. Hengge; Z. Y. Zhang, *Biochemistry* **2004**, *43*, 8256–8264.
212. A. C. Hengge; G. A. Sowa; L. Wu; Z. Y. Zhang, *Biochemistry* **1995**, *34*, 13982–13987.
213. P. K. Grzyska; Y. Kim; M. D. Jackson; A. C. Hengge; J. M. Denu, *Biochemistry* **2004**, *43*, 8807–8814.
214. A. C. Hengge; J. M. Denu; J. E. Dixon, *Biochemistry* **1996**, *35*, 7084–7092.
215. A. C. Hengge; Y. Zhao; L. Wu; Z. Y. Zhang, *Biochemistry* **1997**, *36*, 7928–7936.
216. Y. Zhao; Z. Y. Zhang, *Biochemistry* **1996**, *35*, 11797–11804.
217. Z. Y. Zhang; J. C. Clemens; H. L. Schubert; J. A. Stuckey; M. W. F. Fischer; D. M. Hume; M. A. Saper; J. E. Dixon, *J. Biol. Chem.* **1992**, *267*, 23759–23766.
218. G. Lowe; B. V. L. Potter, *Biochem. J.* **1981**, *199*, 693–698.
219. W. J. Ray, Jr.; J. W. Burgner, II; C. B. Post, *Biochemistry* **1990**, *29*, 2770–2778.
220. S. D. Lahiri; G. Zhang; D. Dunaway-Mariano; K. N. Allen, *Biochemistry* **2002**, *41*, 8351–8359.
221. H. Deng; W. J. Ray, Jr.; J. W. Burgner, II; R. Callender, *Biochemistry* **1993**, *32*, 12984–12992.
222. S. D. Lahiri; G. Zhang; D. Dunaway-Mariano; K. N. Allen, *Science* **2003**, *299*, 2067–2071.
223. G. M. Blackburn; N. H. Williams; S. J. Gamblin; S. J. Smerdon, *Science* **2003**, *301*, 5637.
224. C. E. Webster, *J. Am. Chem. Soc.* **2004**, *126*, 6840–6841.
225. N. J. Baxter; L. F. Olguin; M. Golicnik; G. Feng; A. M. Hounslow; W. Bermel; G. M. Blackburn; F. Hollfelder; J. P. Waltho; N. H. Williams, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 14732–14737.
226. K. N. Allen; D. Dunaway-Mariano, *Science* **2003**, *301*, 5637.
227. L. W. Tremblay; G. Zhang; J. Dai; D. Dunaway-Mariano; K. N. Allen, *J. Am. Chem. Soc.* **2005**, *127*, 5298–5299.
228. L. Pauling, Interatomic Distances. In *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960; pp 255–260.
229. M. A. Kertesz, *FEMS Microbiol. Rev.* **1999**, *24*, 135–175.
230. S. R. Hanson; M. D. Best; C. H. Wong, *Angew. Chem. Int. Ed.* **2004**, *43*, 5736–5763.
231. M. J. Reed; A. Purohit; L. W. L. Woo; S. P. Newman; B. V. L. Potter, *Endocr. Rev.* **2005**, *26*, 171–202.
232. G. Diez-Roux; A. Ballabio, *Annu. Rev. Genomics Hum. Genet.* **2005**, *6*, 355–379.
233. M. Matusiewicz; M. Krzystek-Korpacka; D. Diakowska; K. Grabowski; K. Augoff; L. Paradowski; I. Kustrzeba-Wojcicka; M. Piast; T. Banas, *Int. J. Colorectal Dis.* **2008**, *23*, 383–387.
234. Y. Nakamura; T. Suzuki; T. Fukuda; A. Ito; M. Endo; T. Moriya; Y. Arai; H. Sasano, *Prostate* **2006**, *66*, 1005–1012.
235. S. J. Stanway; P. Delavault; A. Purohit; L. W. L. Woo; C. Thurieau; B. V. L. Potter; M. J. Reed, *Oncologist* **2007**, *12*, 370–374.
236. J. D. Mougous; R. E. Green; S. J. Williams; S. E. Brenner; C. R. Bertozzi, *Chem. Biol.* **2002**, *9*, 767–776.
237. A. M. Roberton; R. Wiggins; P. J. Horner; R. Greenwood; T. Crowley; A. Fernandes; M. Berry; A. P. Corfield, *J. Clin. Microbiol.* **2005**, *43*, 5504–5508.
238. P. Gadler; K. Faber, *Trends Biotechnol.* **2007**, *25*, 83–88.
239. A. J. Ellis; S. G. Hales; N. G. A. Ur-Rehman; G. F. White, *Appl. Environ. Microbiol.* **2002**, *68*, 31–36.
240. C. Peters; B. Schmidt; W. Rommerskirch; K. Rupp; M. Zuhlsdorf; M. Vingron; H. E. Meyer; R. Pohlmann; K. Vonfigura, *J. Biol. Chem.* **1990**, *265*, 3374–3381.
241. B. Schmidt; T. Selmer; A. Ingendoh; K. von Figura, *Cell* **1995**, *82*, 271–278.

242. T. Selmer; A. Hallmann; B. Schmidt; M. Sumper; K. von Figura, *Eur. J. Biochem.* **1996**, *238*, 341–345.
243. T. Dierks; C. Miech; J. Hummerjohann; B. Schmidt; M. A. Kertesz; K. von Figura, *J. Biol. Chem.* **1998**, *273*, 25560–25564.
244. O. Berteau; A. Guillot; A. Benjdia; S. Rabot, *J. Biol. Chem.* **2006**, *281*, 22464–22470.
245. C. Miech; T. Dierks; T. Selmer; K. von Figura; B. Schmidt, *J. Biol. Chem.* **1998**, *273*, 4835–4837.
246. K. von Figura; B. Schmidt; T. Selmer; T. Dierks, *Bioessays* **1998**, *20*, 505–510.
247. C. S. Bond; P. R. Clements; S. J. Ashby; C. A. Collyer; S. J. Harrop; J. J. Hopwood; J. M. Guss, *Structure* **1997**, *5*, 277–289.
248. G. Lukatela; N. Krauss; K. Theis; T. Selmer; V. Gieselmann; K. von Figura; W. Saenger, *Biochemistry* **1998**, *37*, 3654–3664.
249. R. von Bulow; B. Schmidt; T. Dierks; K. von Figura; I. Uson, *J. Mol. Biol.* **2001**, *305*, 269–277.
250. F. G. Hernandez-Guzman; T. Higashiyama; W. Pangborn; Y. Osawa; D. Ghosh, *J. Biol. Chem.* **2003**, *278*, 22989–22997.
251. M. Chruszcz; P. Laidler; M. Monkiewicz; E. Ortlund; L. Lebioda; K. Lewinski, *J. Inorg. Biochem.* **2003**, *96*, 386–392.
252. D. Ghosh, *Cell. Mol. Life Sci.* **2007**, *64*, 2013–2022.
253. P. Bojarova; E. Denehy; I. Walker; K. Loft; D. P. De Souza; L. W. L. Woo; B. V. L. Potter; M. J. McConville; S. J. Williams, *ChemBioChem* **2008**, *9*, 613–623.
254. S. G. Gibby; J. M. Younker; A. C. Hengge, *J. Phys. Org. Chem.* **2004**, *17*, 541–547.
255. B. Spencer, *Biochem. J.* **1958**, *69*, 155–159.
256. E. J. Sampson; E. V. Vergara; J. M. Fedor; M. O. Funk; S. J. Benkovic, *Arch. Biochem. Biophys.* **1975**, *169*, 372–383.
257. C. L. L. Chai; W. A. Loughlin; G. Lowe, *Biochem. J.* **1992**, *287*, 805–812.
258. R. Recksiek; T. Selmer; T. Dierks; B. Schmidt; K. von Figura, *J. Biol. Chem.* **1998**, *273*, 6096–6103.
259. G. R. Burns; E. Galanopoulou; C. H. Wynn, *Biochem. J.* **1977**, *167*, 223–227.
260. A. Waldow; B. Schmidt; T. Dierks; R. von Bulow; K. von Figura, *J. Biol. Chem.* **1999**, *274*, 12284–12288.
261. A. Knaust; B. Schmidt; T. Dierks; R. von Bulow; K. von Figura, *Biochemistry* **1998**, *37*, 13941–13946.
262. S. J. Benkovic; E. V. Vergara; R. C. Hevey, *J. Biol. Chem.* **1971**, *246*, 4926–4933.
263. G. Hagelueken; T. M. Adams; L. Wiehlmann; L. Widow; H. Kolmar; B. Tummler; D. W. Heinz; W. D. Schubert, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7631–7636.
264. B. Bartholomew; K. S. Dodgson; G. W. Matcham; D. J. Shaw; G. F. White, *Biochem. J.* **1977**, *165*, 575–580.
265. J. M. Cloves; K. S. Dodgson; D. E. Games; D. J. Shaw; G. F. White, *Biochem. J.* **1977**, *167*, 843–846.
266. P. Gadler; K. Faber, *Eur. J. Org. Chem.* **2007**, 5527–5530.
267. S. R. Wallner; M. Batter; C. Wurdernann; P. Wecker; F. O. Glockner; K. Faber, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 6381–6384.
268. A. Kahnert; M. A. Kertesz, *J. Biol. Chem.* **2000**, *275*, 31661–31667.
269. I. Muller; C. Stuckl; J. Wakely; M. Kertesz; I. Uson, *J. Biol. Chem.* **2005**, *280*, 5716–5723.
270. I. Muller; A. Kahnert; T. Pape; G. M. Sheldrick; W. Meyer-Klaucke; T. Dierks; M. Kertesz; I. Uson, *Biochemistry* **2004**, *43*, 3075–3088.

Biographical Sketches

Tiago A. S. Brandao is from Porto Alegre, Brazil. He received his B.S. in pharmacy from Federal University of Santa Catarina (Florianópolis, Brazil) in 2000 and his Ph.D. in chemistry from the same University in 2007. During the course of his Ph.D. he worked in the development of models for PPs under the supervision of Professor Faruk Nome. Since 2007, he has been a postdoctoral fellow in the group of Professor Alvan C. Hengge at the Utah State University, where he has been working on a project that aims to understand how protein movement in protein tyrosine PPs is associated with catalysis. His major research interests are in the area of mechanism and catalysis of phosphoric and carboxylic ester reactions by enzymatic systems and their models.

Alvan C. Hengge was born in Cincinnati, Ohio, USA. After his B.S. degree from the University of Cincinnati in 1974, he taught high school chemistry and physics from 1975 to 1982. He then returned to the Graduate School at the University of Cincinnati, and earned Ph.D. in organic chemistry in 1987 under the direction of R. Marshall Wilson studying the reactions of triazolinedione ylides. This was followed by an NIH postdoctoral fellowship in the laboratory of W. W. Cleland at the University of Wisconsin Institute for Enzyme Research, where he studied the biochemistry of phosphoryl and acyl transfer. After several subsequent years as an assistant scientist in the Cleland laboratory, he joined the faculty at Utah State University in 1996, where he is now a professor in the Department of Chemistry and Biochemistry. His research focuses on investigations of the mechanisms of biologically important reactions, particularly phosphate and sulfate ester chemistry.

8.10 Catalytic Mechanism of DNA Polymerases

Michelle P. Roettger, Marina Bakhtina, and Sandeep Kumar, The Ohio State University, Columbus, OH, USA

Ming-Daw Tsai, Academia Sinica, Taipei, Taiwan

8.10.1   Introduction

8.10.1.1   Basic Polymerase Function

DNA-dependent DNA polymerases are responsible for directing the synthesis of new DNA from deoxyribonucleotide triphosphates (dNTPs) opposite an existing DNA template, which contains the genetic information critical to an organism's survival. To properly preserve this information, during each round of catalysis, a polymerase must accurately select and catalyze the insertion of a complementary nucleotide (dNTP) substrate, from a pool of four structurally similar molecules, into a nascent DNA strand. Present across all three domains of life, including Archaea, Bacteria, and Eukaryota, polymerases are necessarily and diversely utilized during DNA replication, recombination, repair, and translesion synthesis (TLS).

8.10.1.2   Polymerase Fidelity and Relationship to Biological Function

To date, there are at least 15 identified human DNA polymerases possessing a myriad of functions (for reviews see Shcherbakova *et al.*,[1] Hubscher *et al.*,[2] Pavlov *et al.*,[3] and McCulloch and Kunkel[4]). These polymerases possess fidelities broadly ranging from 10^1 to 10^6, plus an additional fidelity enhancement of 10^1–10^2 when intrinsic exonuclease proofreading function is considered. Fidelity, or base substitution error, can be qualitatively regarded as a measure of the frequency by which a polymerase incorporates a correct nucleotide versus an incorrect nucleotide. A replicative polymerase replicates nondamaged DNA, and functionally requires high fidelity in order to accurately preserve genetic information, as well as to prevent mutations that may promote human diseases such as cancer.[5,6] Repair polymerases also possess a moderately high fidelity. The more recently discovered low fidelity polymerases are also of importance to human survival, as a number of them possess the capacity to bypass replication stalling lesions during replication of damaged DNA, and some may play a large role in developing the DNA sequence diversity required to effect proper immune response.

8.10.1.3   Objective

Studies examining the energetics of base pairing in solution have estimated the free energy difference between correct and incorrect base pairing to be 1–3 kcal mol^{-1}.[7,8] This translates into a fidelity of 10^1–10^2 if a polymerase were to offer no selective preference during nucleotide incorporation catalysis.[9] Clearly, the observation of higher fidelity owned by the majority of polymerases indicates that these enzymes possess a unique ability to provide significant selectivity enhancement during catalysis. Based upon the differences in free energy for correct and incorrect base pairing in solution,[8] Pol β's moderate fidelity of 10^4–10^5 translates into a selectivity enhancement of 10^2–10^3 in Pol β's active site.[10,11] After decades of research, the question remains: what is the nature of the mechanism by which such selectivity is amplified in the active site of a polymerase? This work aims to examine the catalytic mechanism of DNA polymerases, with a large focus on DNA polymerase β (Pol β) as a model system. The chapter encompasses the contributions of structural, kinetic, and computational studies in the advancement of our understanding of this mechanism. When appropriate, meaningful contributions from studies on other polymerases are discussed for comparison purposes. Some issues pertinent to the overall subject of polymerase fidelity, which include base–base hydrogen bonding, water exclusion from the active site, geometric selection, mismatch extension, exonuclease proofreading, and substrate misalignment, are briefly addressed here and we refer readers to a comprehensive review of these topics in Kunkel[12] and Kunkel and Bebenek,[13] and references therein.

8.10.2   Polymerase Families

The discovery of the first DNA polymerase, *E. coli* DNA polymerase I, by Kornberg and co-workers occurred over 50 years ago.[14,15] After extensive research spanning multiple decades, there are currently seven families of polymerases (A, B, C, D, X, Y, and RT) which are classified according to primary sequence homology and structural similarity.[16–19] Eukaryotic polymerases belong to four of these families (A, B, X, and Y) (reviewed in Shcherbakova *et al.*,[1] Hubscher *et al.*,[2] Pavlov *et al.*,[3] and McCulloch and Kunkel[4]). Defects in proper regulation

or expression of the genes encoding many of these polymerases are directly linked to phenotypic manifestation of human diseases (reviewed in Sweasy *et al.*[20]).

8.10.2.1   Family A

Belonging to Family A are *Escherichia coli* DNA polymerase I (Pol I), and eukaryotic DNA polymerases, Pol γ, Pol θ, and Pol ν. *E. coli* Pol I, the prototypical member of this family, possesses both $3' \rightarrow 5'$ and $5' \rightarrow 3'$ exonuclease activity and participates in nucleotide excision repair (NER) and Okazaki fragment processing.[21,22] The high fidelity Pol γ is the mitochondrial replicase and repair polymerase (for recent reviews see Kaguni[23] and Graziewicz *et al.*[24]). Recombinant expression of the large Pol γ subunit reveals both $3' \rightarrow 5'$ exonuclease activity and $5'$ deoxyribosephosphate (dRP) lyase activity.[25,26] The roles of the more recently discovered Pol θ and Pol ν are less clear, though both demonstrate TLS capabilities.[27–31] Pol θ has lately been implicated in the somatic hypermutation (SHM) of immunoglobulin (Ig) genes.[32–34] Among other members of this family is bacteriophage replicative DNA polymerase T7.

8.10.2.2   Family B

Notable members comprising the Family B polymerases include the prototypical *E. coli* Pol II, as well as eukaryotic polymerases, Pol α, Pol δ, Pol ε, and Pol ζ. In humans, the coordinated efforts of Pol α, Pol δ, and Pol ε are responsible for leading and lagging strand synthesis during nuclear chromosomal replication prior to cell division (for review see Pospiech and Syvaoja,[35] Hindges and Hubscher,[36] and Lehman and Kaguni[37]). During this process, Pol α acts as a primase, generating short tracts of hybrid RNA/DNA primers, which are subsequently elongated by Pol δ and Pol ε (for review see Johnson and O'Donnell[38]). Both high-fidelity Pol δ and Pol ε possess strong $3' \rightarrow 5'$ exonuclease activity, and have also been implicated in a variety of mammalian repair pathways, including NER, base excision repair (BER), mismatch repair (MMR), and double-strand break (DSB) repair. The more recently identified Pol ζ possesses proficient mismatch and lesion extension capability possibly used in conjunction with other DNA polymerases for TLS during DSB repair, interstrand cross-link repair (ICL), and SHM (recently reviewed in Lawrence[39] and Gan *et al.*[40]). Bacteriophage polymerases T4 and RB69 also belong to this family.

8.10.2.3   Families C and D

The major *E. coli* replicative polymerase Pol III belongs to Family C (for review see Kelman and O'Donnell[41] and O'Donnell *et al.*[42]). The D Family is found only in archaebacteria, and includes *Pyrococcus furiosus* (*Pfu*) Pol D.[43]

8.10.2.4   Family X

The X Family of DNA polymerases houses Pol β, the prototypical polymerase of this family. Pol β serves as a good model system for polymerase mechanism due to its small size, lack of intrinsic exonuclease activity, and stability under a variety of conditions. Studies both *in vivo* and *in vitro* have unequivocally demonstrated Pol β's important role in single-nucleotide gap filling during the short-patch mammalian BER process.[25,44–46] This pathway is modeled as follows (for review see Parikh *et al.*,[47] Wilson,[48] and Lindahl and Wood[49]). A DNA glycosylase creates an apurinic/apyridinic (AP) site through removal of a damaged base in duplex DNA through breakage of the *N*-glycosidic bond. An AP endonuclease then cleaves the phosphodiester backbone $5'$ to the sugar, thus generating a $3'$-hydroxyl group and a dRP flap. Pol β then fills the gap with an undamaged nucleotide, and afterward Pol β's 8-kDa lyase domain is responsible for the removal of the dRP through a β-elimination mechanism.[50] Finally, the nicked duplex DNA is sealed by either DNA ligase I or DNA ligase III.[51,52]

   Other members of this family include eukaryotic terminal deoxynucleotidyl transferase (TdT), Pol μ, Pol λ, and Pol σ. TdT is known to function in antigen receptor diversification during V(D)J recombination through addition of nucleotides (N additions) to gene segment junctions in a template-independent manner[53–55] (reviewed in Fowler and Suo[56]). The more recently discovered Pol λ and Pol μ have been implicated in

general DSB repair, but the specific context of their roles remains to be established (for review see Nick *et al.*[57] and Moon *et al.*[58]). Pol σ is involved in the establishment of sister chromatid cohesion.[59,60] The African swine fever virus (ASFV) Pol X also belongs to this family. ASFV Pol X is the smallest of all known polymerases, and plays a gap-filling role during BER of viral DNA.[61]

8.10.2.5   Family Y

Members of the newly discovered low-fidelity Y Family are exonuclease deficient, possess TLS abilities, and include eukaryotic Pol η, Pol ι, Pol κ, and REV1, and *Sulfolobus solfataricus* DNA polymerase IV (Dpo4). Pol η, Pol ι, and Pol κ have all been shown to interact with the deoxycytidyltransferase, REV1, which may act as a scaffold during polymerase switching, the process by which the aforementioned TLS polymerases are coordinated in their projected participation in either DSB repair (Pol η), BER (Pol ι), NER (Pol κ), and SHM (Pol η and Pol ι) (reviewed in Rattray and Strathern,[62] Prakash *et al.*,[63] Lehmann *et al.*,[64] Lehmann,[65,66] Kannouche and Lehmann,[67] Prakash and Prakash,[68] Kunkel *et al.*,[69] and Vaisman *et al.*[70]).

8.10.2.6   Family RT

HIV-1 reverse transcriptase (RT) possesses DNA and RNA template-dependent polymerase activity, as well as an endonucleolytic ribonuclease H (RNase H) activity responsible for degradation of the RNA in an RNA/DNA duplex. Because RT plays a critical role in the HIV life cycle, it remains a primary therapeutic target for anti-HIV drug development.[71–73]

8.10.3   Structural Requirements for Polymerase Catalysis

8.10.3.1   General Domain/Subdomain Architecture of DNA Polymerases

Although all DNA polymerases do not share significant primary sequence homology, they do share several architectural features in common. Pol β topology consists of two domains: an N-terminal 8-kDa lyase domain and a C-terminal 31-kDa polymerase domain. The N-terminal domain binds the 5′-phosphate of the downstream primer in short gapped DNA substrates,[74] and possesses dRP lyase activity that proves essential for successful completion of BER.[75,76] Upon functional alignment, the C-terminal domain of Pol β,[77,78] possessing DNA polymerization capability, resembles the overall structure of other polymerases, such as HIV-1 reverse transcriptase,[79] bacteriophage T7 DNA polymerase,[80] and Klenow fragment (KF) of *E. coli* DNA polymerase I,[81] in maintaining the canonical polymerase architecture which resembles a hand and consists of fingers, palm, and thumb subdomains, according to the nomenclature suggested by Steitz *et al.*[82] The fingers, palm, and thumb subdomains of all polymerases functionally correspond to actions of nascent base pair binding (N-subdomain), catalysis (C-subdomain), and double-strand DNA binding (D-subdomain), respectively (**Figure 1**). In this chapter, we use this functionally based nomenclature[83] to describe subdomain motions. The spatial orientation of three conserved acidic amino acid residues in the C-subdomain which coordinate the two metal ions required for catalysis (Section 8.10.3.3) are superimposable among polymerase families. Members of the Y-family of DNA polymerases possesses an additional little finger subdomain, or polymerase-associated domain (PAD), which assists in DNA binding.[84]

8.10.3.2   Polymerases Undergo a Global Conformational Change upon dNTP Binding

Multiple crystal structures of polymerases solved in unliganded and various liganded states are of significant importance to polymerase mechanism studies. For example, the plethora of available structures for Pol β include: free enzyme,[85,86] binary complexes of enzyme with gapped, nicked, or blunt-ended DNA,[78,87] and ternary complexes of enzyme, DNA, and correct incoming dNTP.[78,86,88–90] Comparison of Pol β's binary gapped DNA complex with the ternary gapped DNA complex containing correct incoming nucleotide ddCTP, reveals that there is a dNTP-induced subdomain closure originating from the N-subdomain in which there is a 30° rotation of α-helix N toward the nascent base pair along the hinge axis of α-helix M.[78] This movement of
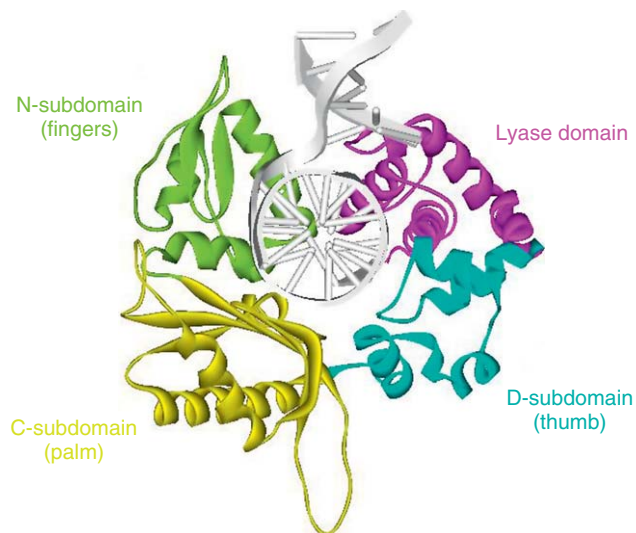
Figure 1   Subdomain organization of Pol β (from ternary complex structure 1BPY as reported in M. R. Sawaya; R. Prasad; S. H. Wilson; J. Kraut; H. Pelletier, *Biochemistry* **1997**, *36*, 11205–11215.[78]) The C-terminal polymerase domain is subdivided into three subdomains: DNA binding D-subdomain (cyan), catalytic C-subdomain (gold), and nascent base pair binding N-subdomain (green). The N-terminal lyase domain is highlighted in pink. DNA is shown in gray.

α-helix N by Pol β, paralleled by movement of α-helix O in T7 and Klentaq[80,91] and $\beta3$–$\beta4$ in HIV-1 RT,[79] serves to create a tight binding pocket for the nascent base pair, bringing about important interactions between polymerase side chains and the minor groove edge of DNA, as well as with the base, sugar, and triphosphate of the incoming dNTP (**Figure 2**).
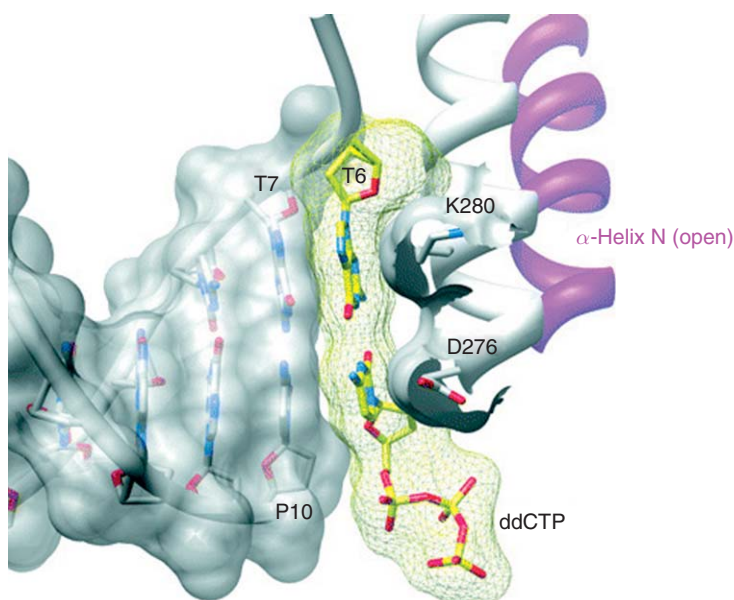
Figure 2   Positioning of α-helix N upon Pol β N-subdomain closure. Pictured are the nascent base pair (incoming and templating nucleotides) (yellow), the semitransparent molecular surface of the upstream DNA duplex (gray), and the molecular surface of the nascent base pair (yellow mesh). In the closed conformation (white), Asp276 and Lys280 of α-helix N stack with the bases of the incoming and templating nucleotides, respectively. These interactions are lost in the open conformation (magenta). Reproduced from W. A. Beard; S. H. Wilson, *Chem. Rev.* **2006**, *106*, 361–382. Copyright 2006 American Chemical Society.

Overall, the crystal structures of many polymerases to date, including Pol β, yield important mechanistic insights in implicating the existence of a large conformational change that occurs upon correct nucleotide binding, from an 'open' E•DNA binary complex conformation to a 'closed' E′•DNA•dNTP ternary complex conformation (for review see Sawaya et al.,[78] Doublie et al.[80] and Beard and Wilson[83]). Formation of the 'closed' ternary complex is a prerequisite to chemistry in order to provide proper active-site alignment. Existence of this conformational change has led to the proposal of a general 'induced-fit' mechanism,[92] in which this conformational change is postulated to be the major contributor to polymerase fidelity.[13,93]

There appears to be a correlation of polymerase fidelity with active-site spaciousness.[94] Crystal structures of low-fidelity Y-family DNA polymerases differ from higher fidelity polymerases in possessing distinctly smaller N- and D-subdomains, which give rise to a largely solvent-accessible active site. This structural feature allows for flexibility in accommodation of bulky lesions (for review see Prakash et al.[63]). Furthermore, it appears that some of the Y-family members do not undergo an open-to-closed conformational transition upon dNTP binding, as evidenced by comparison of the binary and ternary complexes of Dpo4[95–97] and Pol ι.[98]

8.10.3.3   Two Metal-Ion Mechanism

Pol β utilizes a 'two metal-ion' mechanism for nucleotide incorporation chemistry.[99] This mechanism is likely conserved for all DNA and RNA polymerases.[82,100] Crystal structures of Pol β ternary complexes were the first to validate this mechanism on a structural basis.[78,86] Three highly conserved catalytic residues of C-subdomain, Asp 190, Asp192, and Asp256, coordinate two metal ions (**Figure 3**). Metal ion A coordinates dNTP in a tridentate fashion, and presumably enters the active site as a Mg•dNTP complex. Metal ion B coordinates the 3′-oxygen of the primer, as well as the pro-R_p oxygen of the α-phosphate of the incoming nucleotide, and is referred to as the catalytic Mg^{2+} ion in mechanism studies. Both hexacoordinated metal ions serve to stabilize the structure and charge of the pentacovalent transition state formed upon in-line nucleophilic attack of the α-phosphate of an incoming nucleotide by the primer's 3′-oxygen. The nucleotidyl transfer reaction continues through the transition state as the primer increases one nucleotide in length and a pyrophosphate leaving group is formed (**Figure 3**).
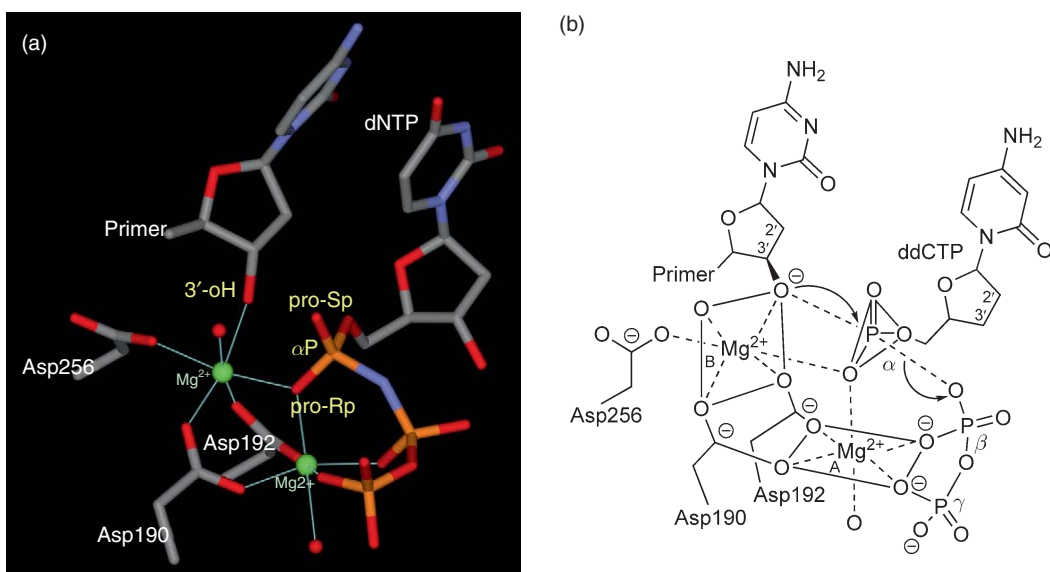
Figure 3   Nucleotidyl transfer mechanism of Pol β. (a) Orientation of incoming dNTP in the DNA Pol β active site depicted from the ternary complex structure 2FMS as reported in V. K. Batra; W. A. Beard; D. D. Shock; J. M. Krahn; L. C. Pedersen; S. H. Wilson, *Structure* **2006**, *14*, 757–766. Pro-S_p and pro-R_p oxygens on the α-phosphate of the incoming dNTP moiety are labeled. Magnesium ions are shown in green. The catalytic Mg^{2+} interacts with pro-R_p non-bridging oxygen.
(b) Schematic representation of nucleophilic attack to form a pentacovalent transition state as modeled from ternary complex structure 1BPY. Part B is reproduced with permission from M. R. Sawaya; R. Prasad; S. H. Wilson; J. Kraut; H. Pelletier, *Biochemistry* **1997**, *36*, 11205–11215. Copyright 1997 American Chemical Society.

8.10.4   General Mechanism of Nucleotide Incorporation Catalyzed by DNA Polymerases

8.10.4.1   Modern Methods Used in DNA Polymerase Mechanism Studies

Consistent with their important roles in living organisms, DNA polymerases are objects of intensive multifaceted studies employing a broad spectrum of biochemical and biophysical methods in conjunction with cellular and molecular biology approaches. In-depth studies of DNA polymerase mechanism require application of modern biophysical techniques and a variety of molecular probes. Great advances in our understanding of the mechanism of dNTP incorporation can be attributed to the utilization of a variety of modern approaches including rapid mixing techniques (rapid chemical quench and stopped-flow), FRET-based kinetics, fluorescence-based assays, and single molecule kinetics. These techniques allow us to monitor the overall progress of single-nucleotide incorporation, as well as to probe specific steps along the reaction pathway, for example, by varying reaction conditions, applying substrate analogues, or altering an enzyme active site through site-directed mutagenesis. This section begins with an overview of the principal methods used in mechanistic studies of DNA polymerases and the main conclusions obtained in these studies. Then we focus on a detailed characterization of DNA polymerase kinetic mechanism using Pol β as a model enzyme. Note that the main purpose of this chapter is to examine the mechanism of dNTP incorporation, therefore it does not include many other important aspects of DNA polymerase function such as DNA substrate recognition and binding, processivity, and self-editing mechanisms.

8.10.4.1.1   Pre-steady-state kinetics using discontinuous assays (rapid chemical quench)

Pre-steady-state (or transient-state) kinetic approaches, allowing the dissection of individual steps and intermediates in an enzymatic reaction, are superior to classical steady-state approaches. Pre-steady-state kinetic methods were first applied to DNA polymerases in late 1980s to early 1990s in classical studies of *E. coli* Pol I (Klenow fragment, KF)[101–104] and bacteriophage T7 DNA polymerase.[105,106] These studies have served as a framework for the kinetic analysis of many polymerases to date. The general polymerase kinetic mechanism derived from these studies and subsequent studies of Pol β is presented in **Figure 4**.

Under pre-steady-state conditions, the enzyme in the reaction is used in stoichiometric amounts, meaning that one of the substrates has a concentration smaller than or comparable to the enzyme concentration. In single-turnover DNA polymerase assays of dNTP incorporation, the enzyme concentration is in excess of the DNA substrate concentration. These conditions allow us to follow the enzyme through one complete catalytic cycle, thus eliminating complications from multiple turnovers. Nucleotide incorporation is a relatively fast process occurring on a millisecond timescale, and rapid chemical quench is the specialized instrument generally
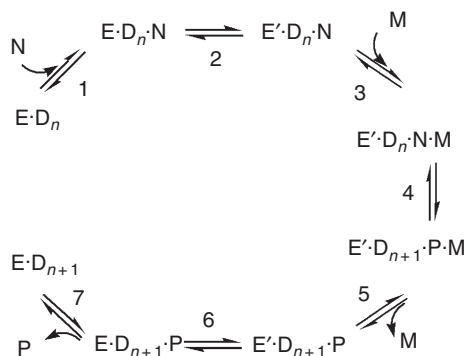
Figure 4   Simplified kinetic scheme of single-nucleotide incorporation by a DNA polymerase. Step 1, Mg•dNTP binding; step 2, N-subdomain closing; step 3, catalytic Mg^{2+} binding; step 4, nucleotidyl transfer (chemistry); step 5, catalytic Mg^{2+} dissociation; step 6, N-subdomain reopening; step 7, pyrophosphate release. E = DNA polymerase in open conformation; E′ = closed conformation; D_n = DNA substrate; D_{n+1} = DNA product elongated by addition of one nucleotide; N = Mg•dNTP; M = catalytic Mg^{2+}; P = Mg•PP_i. Reproduced with permission from M. Bakhtina; S. Lee; Y. Wang; C. Dunlap; B. Lamarche; M. D. Tsai, *Biochemistry* **2005**, *44*, 5177–5187. Copyright 2005 American Chemical Society.
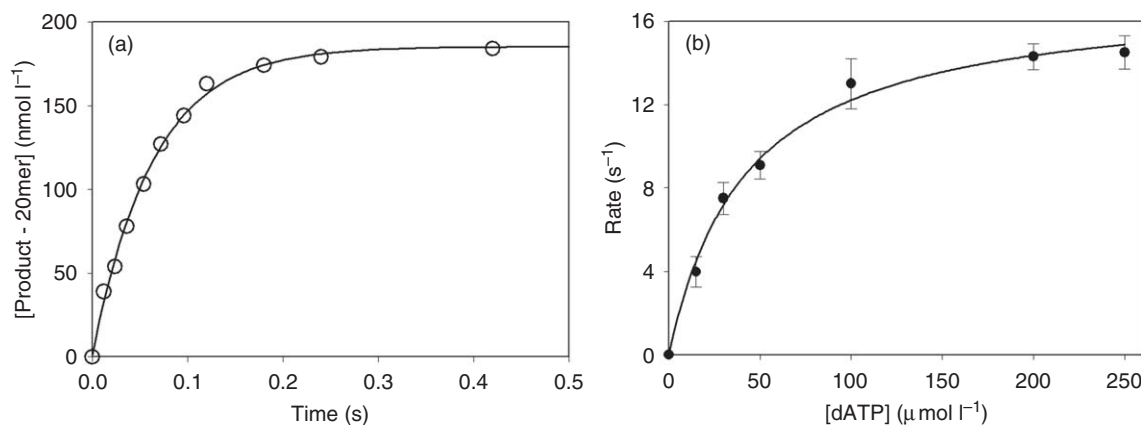
Figure 5 Typical rapid chemical quench assay of single-nucleotide incorporation by Pol β. (a) Single-turnover time course of DNA product formation fitted to a single exponential equation $[DNA_{n+1}] = A(1 - e^{-k_{obs}t})$. (b) Plot of rate versus dATP concentration fitted to a hyperbolic equation $k_{obs} = k_{pol}[dNTP]/(K_{d,app} + [dNTP])$ to obtain values for $K_{d,app}$ and k_{pol}. Adapted with permission from C. A. Dunlap; M. D. Tsai, *Biochemistry* **2002**, *41*, 11226–11235. Copyright 2002 American Chemical Society.

employed to obtain time courses for single-nucleotide incorporation into a DNA substrate. The primer strand of duplex DNA usually carries a ^{32}P-labeled phosphate group or a fluorescent label, so the initial DNA substrate (D_n, where n is the length of the primer) and DNA product (D_{n+1}, where primer is extended in length by one nucleotide) can be resolved and quantified. The observed first-order rate constant of DNA product formation can be obtained from single exponential fit of a plot of the concentration of extended primer versus time (**Figure 5(a)**). Two important kinetic constants, $K_{d,app}$ (the apparent constant of nucleotide dissociation from kinetically active ternary complex) and k_{pol} (the maximum rate constant of single dNTP incorporation), can be obtained from a hyperbolic dependence of the observed rate constant on nucleotide concentration (**Figure 5(b)**). Another important parameter is the ratio of $k_{pol}/K_{d,app}$, which defines the 'catalytic efficiency' constant. The comparison of $k_{pol}/K_{d,app}$ values for correct and incorrect dNTP incorporation serves as a quantitative measure of DNA polymerase fidelity. Fidelity is defined as $[(k_{pol}/K_{d,app})_{cor} + (k_{pol}/K_{d,app})_{inc}]/[(k_{pol}/K_{d,app})_{inc}]$ – where the subscripts 'cor' and 'inc' indicate correct (matched) and incorrect (mismatched) nucleotide incorporation, respectively. In general, fidelity analysis of multiple polymerases reveals that the catalytic efficiencies for matched dNTP incorporation vary widely, whereas the catalytic efficiency for mismatched dNTP incorporation is relatively constant.[107,108]

The interpretation of kinetic parameters in terms of microscopic steps in the reaction pathway depends on the kinetic mechanism of the enzyme. For example, $K_{d,app}$ is often misinterpreted as the thermodynamic dNTP dissociation constant. This would be true if a rapid-equilibrium dNTP binding (step 1 in **Figure 4**) was directly followed by the rate-limiting and virtually irreversible step (this would be step 2 in **Figure 4**). However, as shown later, numerous evidences argue against this scenario, instead pointing to the existence of a fast conformational step (or even multiple steps) prior to the rate-limiting step. It is also important to mention that pre-steady-state kinetic experiments are generally designed with the assumption that DNA polymerases initially form a binary complex with a DNA substrate and then bind a nucleotide substrate. Earlier studies of the kinetic order of substrate binding have concluded that DNA polymerases possess a sequential ordered mechanism where DNA binding always occurs before productive dNTP binding.[109–111] This proposed mechanism is consistent with the template-directed and processive nature of nucleotide incorporation catalyzed by 'classical' high-fidelity DNA polymerases. However, a number of recently discovered DNA polymerases with novel properties,[112] possessing low fidelity, selectivity for certain base pairs, and the ability to use damaged/unnatural substrates, might not necessarily follow the same substrate binding order. For instance, one of the lowest fidelity polymerases, ASFV Pol X, possesses an altered order of substrate binding, with dNTP as the preferred first substrate.[113]

Information extracted from kinetic data collected under 'burst' conditions, in which there is a two- to fourfold excess of DNA substrate over DNA polymerase, illustrates another important application of pre-steady-state experiments. This type of experiment provides useful information about the transient concentration of kinetically active ternary complex. A time course of DNA product formation under these conditions demonstrates a transient exponential phase followed by a steady-state linear phase. By examining the dependence of the burst amplitude on DNA concentration, the enzyme's binding affinity for DNA can be evaluated.

'Burst' conditions have historically been utilized in the pulse-chase/pulse-quench experiment. In the pulse-quench portion of the experiment, a strong acid is used as a chemical quencher of the dNTP incorporation reaction, immediately quenching all enzyme species. In the pulse-chase portion of the experiment, addition of a large excess of nonlabeled dNTP allows incorporation of radiolabeled dNTP trapped in the DNA polymerase active site (**Figure 6**). In KF studies,[101] observation of 20% larger 'burst' amplitude in the pulse-chase compared to the pulse-quench indicates the accumulation of kinetically active E•DNA•dNTP ternary complex in a 'closed' conformation (i.e., in a conformation where dNTP bound in the active site cannot be exchanged with dNTP in solution). Note that the kinetic evidence of the existence of the 'closed' E•DNA•dNTP conformation (which in turn suggests the existence of an 'open-to-closed' conformational change) was obtained before X-ray structural evidence.

It is necessary to mention that the results of pulse-chase/pulse-quench experiments are often interpreted as an evidence for the existence of a rate-limiting conformational step. However, while indicating the existence of a conformational step before the nucleotidyl transfer step (chemistry), the results do not prove that this conformational step is rate limiting. Moreover, observation of transient appearance of the 'closed' conformation would not be possible if the conformational step was much slower than other microscopic steps in the reaction
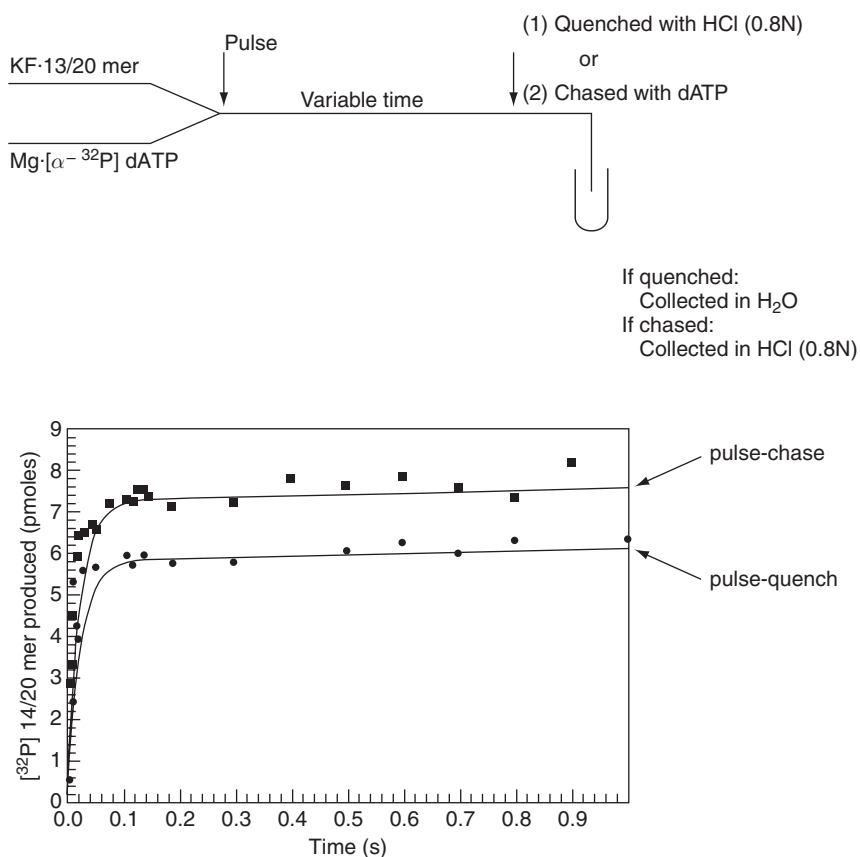
Figure 6 Pulse-chase/pulse-quench experimental design and results obtained for KF. Reproduced with permission from M. E. Dahlberg; S. J. Benkovic, *Biochemistry* **1991**, *30*, 4835–4843. Copyright 1991 American Chemical Society.

pathway. Indeed, there are two conditions that need to be satisfied for the larger 'burst' amplitude to be observed in the pulse-chase compared to the pulse-quench experiment:

1. The rate of the reverse step (k_{-2} in **Figure 4**) cannot be much faster than the rate of forward chemistry step (k_4 in **Figure 4**), so that the majority of the 'closed' ternary complex forms the product.
2. The rate of forward reaction (k_4 in **Figure 4**) cannot be much faster than the rate of 'closing' (k_2 in **Figure 4**), so that an appreciable amount of the 'closed' ternary complex is accumulated.

Pulse-chase/pulse-quench experiments with KF[101] indicated accumulation of the nucleotide bound enzyme species, which would not be possible if the forward reaction was much faster than the rate of conformational 'closing.' To explain this observation, the authors proposed the presence of a kinetic 'road block' – a slow step after the phosphodiester bond formation. However, the results of the pulse-chase/pulse-quench experiments can also be explained by designating chemistry as the slow step, meaning that the chemical step itself plays the role of the 'road block.' The conclusion that chemistry is a fast step in the KF reaction pathway was made based on the observation of a small thio-effect magnitude,[104] which, as elaborated in the following section, should not be used as a solid evidence of the chemical step being nonrate limiting.

8.10.4.1.2   Use of substrate analogues to probe DNA polymerase mechanism

1. *Use of phosphorothioate dNTP analogues.* Nucleotide analogues, in which the nonbridging oxygen at the α-phosphate (αP) is substituted with a sulfur atom, can be used to probe active-site environment. Such substitution makes αP a chiral center, resulting in S_p and R_p stereoisomers of phosphorothioate dNTP analogues. As is the case for many other enzymes which catalyze phosphoryl transfers, DNA polymerases demonstrate stereoselectivity – specifically maintaining a preference toward incorporation of the S_p isomer (**Figure 3**). The most straightforward explanation of the observed stereoselectivity is that the magnesium ion essential for catalysis preferably binds an oxygen ('hard' ligand). If this were the major reason for the observed stereoselectivity, substitution of a 'hard' Mg^{2+} for a 'softer' metal would be expected to reverse the observed stereoselectivity such that the R_p isomer would be preferred. Contrary to that predicted, results of Pol β and KF studies by Burgers and Eckstein[114] and Liu and Tsai[115] have demonstrated that the polymerases select S_p-dNTPαS even in the presence of a 'soft' Mn^{2+}, Cd^{2+}, or Co^{2+} ligand, though with slightly lower selectivity. In light of these results, it is likely that other factors, possibly involving the active-site geometry, play determining roles in selection of the S_p-dNTPαS analogue.

   Thio-substituted dNTP analogues were widely used in early studies of DNA polymerases to address the question of whether or not phosphodiester bond formation is the rate-limiting step in the nucleotide incorporation pathway. The thio-effect is defined as the ratio of phosphoryl transfer rates for a nonsubstituted nucleotide substrate and its corresponding thio-analogue ($k_{pol,dNTP}/k_{pol,dNTP\alpha S}$). Owing to its decreased electronegativity, sulfur is less effective in withdrawing electron density from the central phosphorus atom, making the α-phosphorus less accessible to nucleophilic attack by 3'-OH group of DNA primer terminus. Based on model nonenzymatic reactions of phosphorothioate diesters, the expected range for the thio-effect is 4–11.[116] Application of thio-analogues in DNA polymerase studies revealed that a number of polymerases demonstrate only a three- to fourfold slower rate for sulfur-substituted nucleotide analogue incorporation.[104,105,117] The absence of a full magnitude of thio-effect was interpreted as evidence that chemistry is not the rate-limiting step in the DNA polymerase catalyzed nucleotide incorporation pathway. However, the enzyme active-site environment could potentially affect the magnitude of the 'intrinsic thio-effect' (defined as the observed thio-effect when the chemical step is fully rate limiting), by either increasing or decreasing it. Let us consider a hypothetical nonenzymatic model that perfectly represents the DNA polymerase catalyzed reaction. In this case, the model reaction would occur in an achiral environment where both nonbridging oxygens make an equal contribution to the transition state stabilization. In contrast, the pro-R_p and pro-S_p positions are not equal within the DNA polymerase active site, such that the divalent metal ion coordination occurs exclusively with the pro-R_p oxygen atom (**Figure 3**). This should result in a larger electron-withdrawing role for the pro-R_p oxygen, and therefore a less important contribution of the pro-S_p

oxygen toward transition state stabilization. Consequently, the isomer with the pro-S_p oxygen substituted for sulfur (i.e., the S_p isomer) is expected to demonstrate a smaller reduction in dNTP incorporation rate than the R_p isomer, and since DNA polymerases accept S_p-dNTPαS analogues almost exclusively, this would result in a smaller intrinsic thio-effect than initially predicted by nonenzymatic reactions of phosphorothioate diesters.

2. *Use of halomethylene-modified β–γ bridging dNTP analogues.* A linear free energy relationship (LFER) is a powerful tool for examination of reaction mechanisms. Nucleotide analogues modified at the β–γ bridging oxygen have been used to probe the contribution of leaving group elimination to the overall rate of dNTP incorporation catalyzed by DNA polymerase β.[118,119] A variety of halomethylene-modified dNTP analogues can be used to provide a broad range of leaving group pK_as. Evaluation of LFER for the Pol β catalyzed reaction revealed a strong correlation between the logarithm of the observed rate constant and leaving group pK_a. The observed Brønsted correlation with a steep slope suggests that P–O bond breaking is at least partially rate limiting (**Figure 7**).

The chemical transformation necessary for dNTP insertion by DNA polymerases (step 4 in **Figure 4**) can be considered as a progression of the following events: 3′ OH nucleophile activation (deprotonation), nucleophilic attack on the α-phosphate of incoming dNTP (P–O bond formation), and pyrophosphate leaving group departure (P–O bond cleavage). Evaluation of transition state free energy values associated with each of these microscopic steps is a focus of extensive quantum and molecular mechanical calculation studies, which often result in contradicting conclusions (see Section 8.10.5). Notably, results of Pol β studies using halomethylene-modified dNTP analogues, pH dependence, and the kinetic isotope (deuterium) effect suggest that all three microscopic steps make a significant contribution to the overall rate of single-nucleotide incorporation, or in other words, all three steps have comparable transition state energies.[119]

3. *Analogues with modified nucleobases.* Various dNTP analogues modified at the nucleobase have been synthesized and used to probe mechanism of DNA polymerase selectivity toward correct base pair formation.[120] Factors that probably affect the selectivity of nucleotide incorporation include base pair hydrogen bonding, steric factors such as base pair size and shape, and base stacking interactions. A generally accepted hypothesis suggests that polymerases check the geometry of base pairing through hydrogen bonds and steric interactions with the minor groove of the DNA. Steric and hydrogen bonding factors have been studied for several DNA polymerases utilizing series of systematically altered dNTP analogues.[121,122] The most striking results were obtained with difluorotoluene (F in **Figure 8**) – a nucleobase analogue that has the same shape as thymine, but lacks hydrogen bond-forming capability.
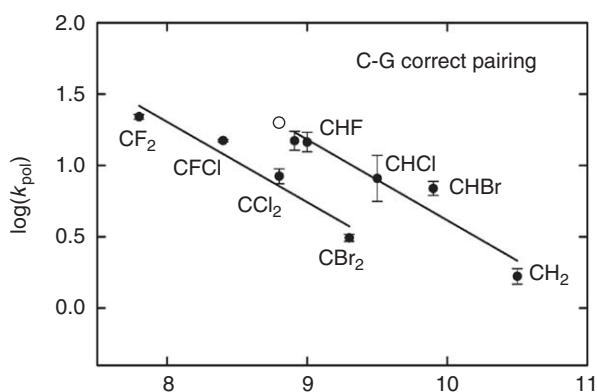
Figure 7 Brønsted correlations of log(k_{pol}) versus leaving-group pK_a. Data for correct dG:dC incorporation catalyzed by Pol β. The data for monohalogenated analogues, the methylene analogue, and native dGTP conform to a linear relationship with $\beta_{lg} = -0.56$. The catalytic sensitivity to leaving group ability in this region is similar to that reported for other enzymatic P–O bond cleavages and suggests leaving-group elimination is rate limiting (or near rate-limiting). The di-halogenated analogues (connected by a separate solid line) show deviation from the other data, suggesting the possibility of a steric effect on k_{pol}. Reproduced with permission from C. A. Sucato; T. G. Upton; B. A. Kashemirov; J. Osuna; K. Oertell; W. A. Beard; S. H. Wilson; J. Florian; A. Warshel; C. E. McKenna; M. F. Goodman, *Biochemistry* **2008**, *47*, 870–879. Copyright 2008 American Chemical Society.
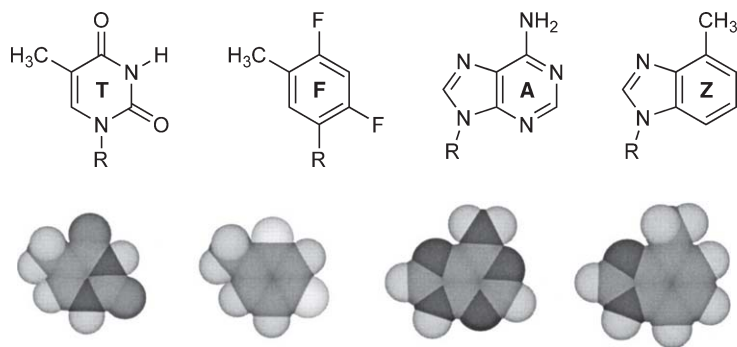
Figure 8   Structure and shape of two natural nucleosides (A and T), alongside two nonpolar analogues (Z and F). Reprinted, with permission, from the *Annual Review of Biophysics and Biomolecular Structure*, Volume 30 ©2001 by Annual Reviews www.annualreviews.org.

If hydrogen bonding makes a major contribution to nucleotide selectivity, the expected dNTP incorporation efficiency opposite templating dF would be greatly reduced. However, several studies on KF, T7, HIV-1 RT, and Taq polymerase have demonstrated that these enzymes are able to incorporate dATP opposite templating dF (forming dA:dF base pair) with efficiency and accuracy comparable to incorporation opposite templating dT.[123–125] Similar results were obtained with 4-methylbenzimidazole (Z in **Figure 8**), which closely mimics the shape of adenine. Use of dFTP and dZTP as nucleotide substrates in polymerization reactions also revealed that they can be efficiently and selectively inserted opposite templating dA and dT, correspondingly. The aforementioned results indicate that hydrogen bonding alone does not account for the origin of DNA polymerase selectivity. Steric effects were further investigated using a series of thymine analogues in which size and shape were gradually changed by use of variably sized atoms (H, F, Cl, Br, and I) to replace the oxygens.[126] Results with KF show that both replication efficiency and fidelity initially increase through the series, reaching a maximum at the chlorinated analogue. Then, as a steric limit is apparently reached, bulkier compounds demonstrate substantially decreased efficiency and selectivity. These results reinforce a tight steric fit within the polymerase active site that plays an essential role in DNA replication fidelity.

On the contrary, application of nucleotide substrate analogues lacking hydrogen bonding potential to mammalian DNA polymerases α, β, and γ suggest that hydrogen bond energy does contribute significantly to substrate selectivity for these polymerases.[126,127] Furthermore, in Pol α studies, application of an array of purine nucleotide analogues with systematically removed or added H-bond forming groups revealed importance of these groups for correct dNTP selection, while the shape of the base pair was found to be essentially irrelevant.[121] It is evident that the relative contribution of steric fit and hydrogen bonding to nucleotide selection differs among polymerases.

8.10.4.1.3   *Site-directed mutagenesis and sequence alignment*

Another very powerful approach to gaining additional information about the contribution of individual residues to polymerase catalysis is to introduce specific alteration of amino acid residues through site-directed mutagenesis. Sequence alignment examining homology between previously studied family members can also be used to speculate on the effect that site-directed mutagenesis of a specific residue may have on an enzyme. For example, the Tyr–Phe motif found in the α-helix M of the N-subdomain of X-family members Pol β (Tyr271–Phe272) and Pol λ (Tyr505–Phe506) provide a preference for strong ribonucleotide discrimination.[86,128] The corresponding motif in Pol μ is replaced with a Gly435–Trp436.[129] Prior to the completion of the crystal structure of Pol μ, sequence alignment analysis coupled with site-directed mutagenesis identified the amino acid residue Gly435 of Pol μ as the residue is critically responsible for this enzyme's lack of sugar discrimination during incorporation into single-nucleotide gapped DNA. This is probably because the small side chain of Gly435 does not act as a steric barrier for the 2′-OH of an incoming rNTP.[130] Single amino acid substitutions have also been used to examine the basis of sugar selectivity for a number of other polymerases.[131–135]

Countless studies on Pol β mutants lend insight into the role of individual side chains in binding and catalysis (reviewed in Beard and Wilson[83]). Specifically, mutagenesis has been used to examine the active-site interactions of Asp276 with incoming dNTP, Lys280 with the templating base, and Tyr271, Asn279, and Arg283 with the minor groove edge of DNA in proximity to the nascent base pair binding pocket. Individual mutagenesis of Pol β catalytic residues Asp190, Asp192, and Asp256 abolishes polymerase activity, and gave first clues as to their active-site location and critical role in catalysis prior to structural information availability.[136,137] Two site-specific Pol β mutants, R258A and I260Q (discussed below), have been used to further dissect the mechanism of Pol β nucleotide incorporation.[138,139]

8.10.4.1.4   Continuous transient-state kinetic methods (stopped-flow assays)

Perhaps the largest advances in elucidating the kinetic mechanism of DNA polymerases have been obtained using stopped-flow technique and various fluorescent probes. For example, use of a fluorophore-modified DNA substrate has facilitated real-time visualization of conformational changes that occur during the course of dNTP incorporation catalyzed by a number of DNA polymerases, including Pol β, KF, Dbh, and T4 DNA polymerase.[138,140–144] Alternatively, a fluorophore can be attached to the polymerase itself and serve as an independent reporter of changes in enzyme conformation.[145] The significant advantage of the stopped-flow fluorescence method is that it allows direct measurement of the rate of intermediate species formation and disappearance. Application of various substrate analogues and alteration of the reaction conditions in fluorescence stopped-flow studies of Pol β result in a most complete understanding of DNA polymerase kinetic mechanism (as described in Section 8.10.4.2).

Development of fluorescence resonance energy transfer (FRET) systems have provided perhaps the most direct way of determining the rate of the 'open-to-closed' conformational change in DNA polymerase mechanism.[146–148] A single fluorescent probe (such as fluorophore-modified DNA or a fluorophore-labeled enzyme) might reflect both global and local structural changes, which makes interpretation of results from stopped-flow fluorescence methods more complex and, in some cases, ambiguous. In contrast, a pair of strategically positioned donor/acceptor fluorescence probes can serve as immediate reporters of DNA polymerase global conformational changes, as changes in distance between the two fluorophores upon the motion of the enzyme subdomains result in changes of FRET efficiency.

8.10.4.1.5   Single molecule kinetics

T7 DNA polymerase was the first polymerase studied at the single molecule level.[149] In these studies DNA substrate was labeled with a Cy3 fluorescent probe that is known to be sensitive to changes in the environment. Addition of correct dNTP triggers a change in $E{\cdot}D_n$ binary complex conformation, which results in an increase in Cy3 fluorescence. Notably, no increase of fluorescence intensity was observed on the single-molecule level in the absence of dNTP substrate. This observation argues against the hypothesis that the binary $E{\cdot}D_n$ complex exists in a dynamic equilibrium between 'open' and 'closed' conformational states and that dNTP binding shifts the equilibrium toward the 'closed' state.[150] Therefore, it is probable that dNTP binding is absolutely required for the initiation of the conformational change. Importantly, the results of the single molecule kinetics correlate perfectly with the results of stopped-flow ensemble-averaged assays using the same Cy3 modified DNA substrate.[149]

8.10.4.2   DNA Polymerase Kinetic Mechanism

Pol β is possibly the best-studied DNA polymerase to date, and we present it here as a model enzyme to illustrate current progress in the field. Our current knowledge of the DNA polymerase kinetic mechanism can be illustrated by **Figure 4** and is summarized as follows. Binding of dNTP to $E{\cdot}D_n$ binary complex (step 1) induces fast N-subdomain closing (step 2). This step remains mostly unperturbed for both matched and mismatched nucleotide incorporation. Active-site residue rearrangements during the course of the subdomain-closing conformational change create a binding pocket for the catalytic Mg^{2+} ion (step 3). There is no evidence for a kinetically distinct rate-limiting conformational step induced by catalytic Mg^{2+} binding. Instead, the chemistry step (step 4) for both matched and mismatched incorporation is rate limiting through phosphodiester bond formation. However, the mismatched $E'{\cdot}D{\cdot}N{\cdot}M$ complex is substantially destabilized in

comparison to the matched E′•D•N•M complex. The free energy difference between these complexes is further enhanced through the chemical transition state. Catalytic Mg^{2+} dissociation (step 5), subdomain reopening (step 6), and pyrophosphate dissociation (step 7) follow phosphodiester bond formation to complete a single turnover of dNTP incorporation. In the following subsections, experimental evidence for this proposed scheme is presented. It must be clarified that for purposes of discussion in this chapter, inclusive to the chemical step are all microscopic steps pertaining to both enzyme and substrate active-site adjustments through nucleotidyl transfer, from E′•D$_n$•N•M to E′•D$_{n+1}$•P•M in **Figure 4**. Computational studies (Section 8.10.5) have been used to further investigate these kinetically unresolvable microscopic steps within the chemistry step.

8.10.4.2.1 *Stopped-flow fluorescence*

Stopped-flow fluorescence analyses of Pol β have been very important in advancing DNA polymerase mechanism.[89,138,140,144,151] Commonly used in DNA polymerase studies is 2-aminopurine (2-AP) – a fluorescent probe, which acts as an excellent analogue of adenine in forming an almost undistorted pair with thymidine (**Figure 9**). It has been suggested that changes in aromatic stacking within DNA dominate the fluorescence emission of 2-AP.[152] Crystal structures of Pol β binary and ternary complexes reveal that the templating position downstream from the nascent base pair experiences dramatic changes in base-stacking interactions during the course of a single turnover.[78] Consequently, a DNA substrate with 2-AP occupying this position (**Figure 9**) yields the greatest signal-to-noise ratio for 2-AP emission in DNA polymerase fluorescence studies.[140] This is probably due to the 90° DNA backbone kink observed in the closed ternary structure which minimizes the base-stacking interactions of 2-AP in this position with the preceding and following bases.

For purposes of the discussion in the following sections, unless otherwise specified, a typical stopped-flow reaction is initiated by rapid mixing of preformed E•D$_n$ binary complex with Mg•dNTP in the presence of excess Mg^{2+}. Monitoring of 2-AP fluorescence (present in the DNA substrate) during correct dNTP incorporation in stopped-flow results in a biphasic trace, in which one phase has a rate identical to that of single dNTP incorporation (as determined by rapid chemical quench experiments) and the other phase has a significantly faster rate (**Figure 10**). Importantly, use of fluorescence from tryptophan, naturally present in the N-subdomain of Pol β, as an independent probe of Pol β conformational changes, also reveals a biphasic fluorescence transition possessing rates similar to those obtained in 2-AP fluorescence stopped-flow assays.[140,151] This observation indicates that both fluorophores report the same steps in Pol β reaction pathway.
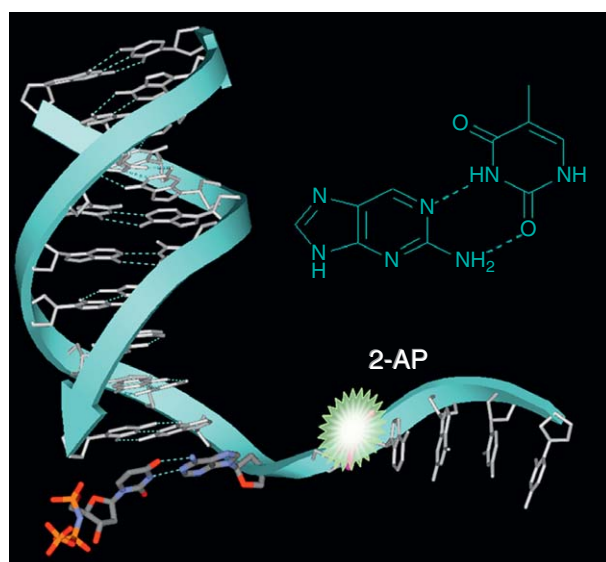
Figure 9 Position of 2-AP fluorescent probe in DNA substrate from Pol β ternary complex structure 2FMS as reported in V. K. Batra; W. A. Beard; D. D. Shock; J. M. Krahn; L. C. Pedersen; S. H. Wilson, *Structure* **2006**, *14*, 757–766. Right corner: Structure of 2-aminopurine base-paired with thymidine.
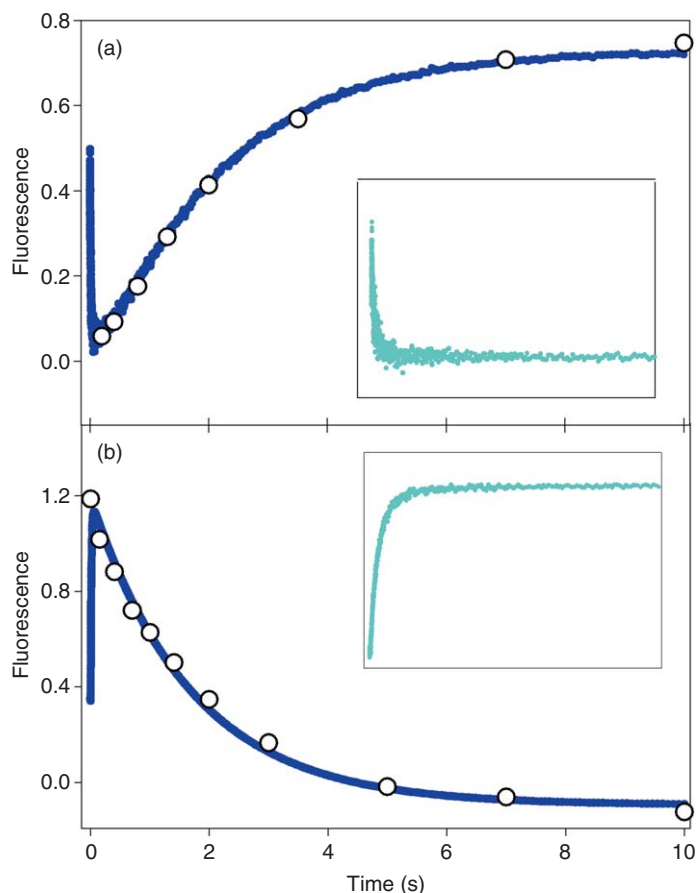
Figure 10   Superimposition of rapid chemical quench (open circles ○) and stopped-flow fluorescence (blue) assays. In (a) tryptophan emission was detected, and in (b) the fluorescence change from 2-AP was monitored. Insets show the dNTP binding-induced conformational change in the presence of dideoxy-terminated DNA substrate. Adapted with permission from A. K. Showalter; B. J. Lamarche; M. Bakhtina; M. I. Su; K. H. Tang; M. D. Tsai, *Chem. Rev.* **2006**, *106*, 340–360. Copyright 2006 American Chemical Society.

8.10.4.2.2   *Structural bases of the fast and slow fluorescence transitions*

In order to delineate the structural bases of the fast and slow fluorescence transitions, further studies involving a variety of chemical probes and altered reaction conditions have been conducted:

1. Traces from stopped-flow experiments utilizing natural dNTP in combination with dideoxy-terminated DNA substrate, in which the 3′-OH of the primer terminus is absent, do not exhibit the slow phase of fluorescence transition (**Figure 10**, inset). Since nucleotidyl transfer is not possible with dideoxy-terminated DNA substrate, the existence of fast fluorescence transition alone indicates that the fast phase of the stopped-flow trace is associated with a conformational step preceding chemistry, while completion of the chemical step is required for the slow phase to appear.[89,140,151] Similarly, the presence of only the fast phase is observed in stopped-flow experiments utilizing the combination of regular nondideoxy-terminated DNA substrate and nonhydrolyzable α,β-methylene-dNTP (dNMPCPP) analogues.[153]

2. It is generally accepted that thio-substituted analogues (S_p-dNTPαS) should perturb mainly the chemical step rather than conformational steps. Use of thio-substituted analogues in stopped-flow assays show that the rate of the fast phase remains unperturbed, while the rate of the slow fluorescence transition is significantly reduced (**Figure 11(a)**).[151] This reinforces the association of the rate of the slow fluorescence transition with the rate of chemistry.
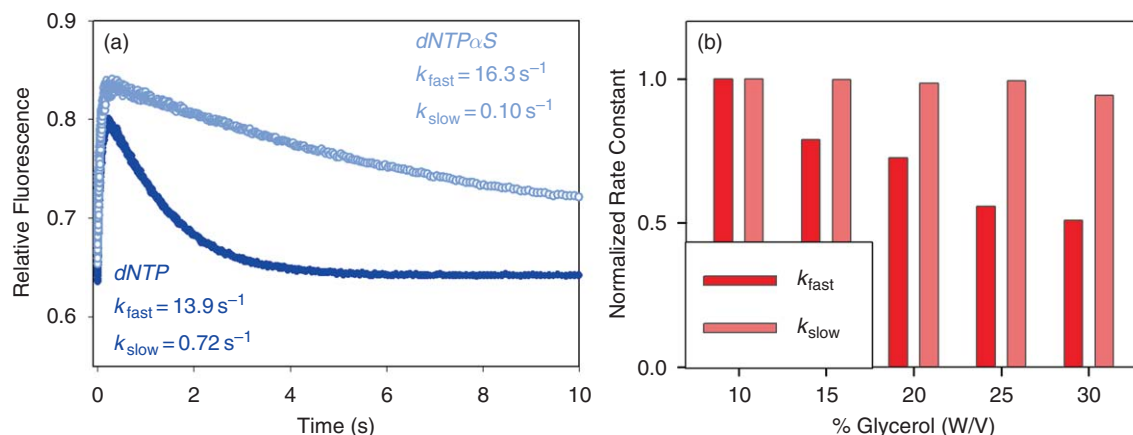
Figure 11 (a) Effect of dNTPαS in 2-AP stopped-flow fluorescence assays. Reproduced with permission from M. Bakhtina; S. Lee; Y. Wang; C. Dunlap; B. Lamarche; M. D. Tsai, *Biochemistry* **2005**, *44*, 5177–5187. Copyright 2005 American Chemical Society. (b) Viscosity effect on the fast and slow fluorescence transitions. Adapted with permission from M. Bakhtina; M. P. Roettger; S. Kumar; M. D. Tsai, *Biochemistry* **2007**, *46*, 5463–5472. Copyright 2007 American Chemical Society.

3. Viscogens, such as glycerol, can be used to perturb conformational steps in an enzymatic pathway which involve large spatial motions.[154–157] Viscosity studies at neutral pH indicate that upon increasing sucrose or glycerol concentrations in the reaction buffer, the fast fluorescence transition significantly slows down, while the rate of the slow fluorescence transition remains virtually unaffected (**Figure 11(b)**).[151] This selective sensitivity to the presence of a viscogen suggests that the fast fluorescence transition reflects a major conformational change associated with progression from the 'open' binary to the 'closed' ternary complex (step 2 in **Figure 4**). Also, the rate of the slow phase does not show sensitivity to altered buffer viscosity, as would be expected of a step reflecting the rate of chemistry.

4. On the basis of negligible structural changes observed between the Pol β pre-chemistry ternary complex and the Pol β post-chemistry ternary complex prior to pyrophosphate release,[89] it is important to point out that the slow fluorescence transition is not caused by the phosphodiester bond formation directly. Instead, the fluorescence change originates from a conformational change rate limited by chemistry. In other words, this conformational step should occur after chemistry with a rate faster than chemistry.

Examination of the effects of pH on Pol β dNTP incorporation demonstrate that as pH increased, the rate of the fast fluorescence phase remains mostly unchanged, while the rate of the slow fluorescence phase increases considerably.[138] Hence, under conditions of high pH (green profile in **Figure 12**), the energy barrier of the chemical step is lowered relative to that at neutral pH (blue profile in **Figure 12**),
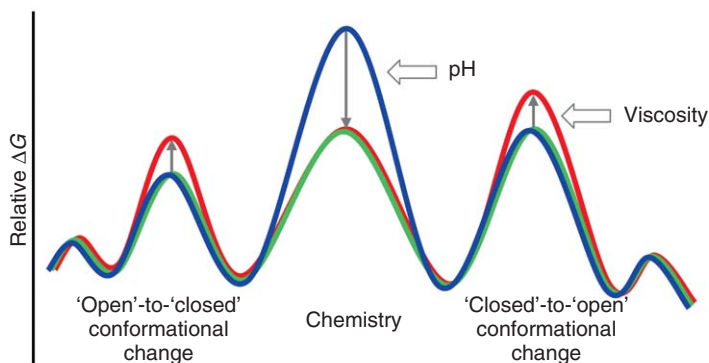

Figure 12 Qualitative free energy profile of correct dNTP incorporation by Pol β. This figure depicts Pol β mechanism at neutral pH (blue), high pH (green), and high pH and high viscosity (red). Adapted with permission from M. Bakhtina; M. P. Roettger; S. Kumar; M. D. Tsai, *Biochemistry* **2007**, *46*, 5463–5472. Copyright 2007 American Chemical Society.

becoming comparable to the energy barriers of the conformational steps. As discussed above, viscosity can selectively affect the rate of conformational steps, so that as reaction buffer viscosity increases, the rate of the fast fluorescence transition decreases. Therefore, selective slowing of the conformational steps (through increased buffer viscosity) in combination with a selective increase in chemistry (through increased pH) should yield conditions in which a conformational change, rather than chemistry, has the highest energy transition step (red profile in **Figure 12**) during the course of single-nucleotide incorporation.

   In agreement with this proposal, results of stopped-flow assays performed at high pH demonstrate that both fast and slow fluorescence transitions are sensitive to increased buffer viscosity, which indicates that chemistry does not limit the rate of the slow phase under these conditions (**Figure 13(a)**). Notably, the rate of dNTP incorporation determined in rapid chemical quench remains unaffected by increases in viscosity (**Figure 13(b)**). Hence, under conditions of high pH and high viscosity, the rate of single dNTP incorporation is faster than the rate of the second fluorescence change. The concomitant effects of pH and viscosity used to dissect the rate of chemistry and the rate of the slow fluorescence change observed in stopped-flow during dNTP incorporation lend support to the hypothesis that the slow fluorescence transition actually originates from N-subdomain reopening after chemistry (step 6 in **Figure 4**).[138] Similar conclusions have been obtained from kinetic analyses of KF and Dpo4 DNA polymerases;[138,158] however, for these enzymes the rate of the reopening from the closed conformation is slower than the chemical step under neutral pH conditions.

5. Site-directed mutagenesis has been successfully utilized to further support the assignment of the slow fluorescence transition to the N-subdomain reopening step.[138] Comparison of crystal structures of Pol β binary, ternary, and product complexes demonstrates that dNTP-induced conformational closing and reopening after dNTP incorporation are accompanied by Arg258 side-chain reorientation.[78] Recent computational studies have suggested that Arg258 side-chain reorientation is probably the rate-limiting microscopic event during the course of Pol β's subdomain closing before chemistry and reopening after chemistry.[159,160] It has also been proposed that the R258A mutant has a lower energy barrier for the subdomain-closing and reopening steps compared to wild type (WT). Stopped-flow fluorescence analyses of R258A indicate that the mutant enzyme has a facilitated rate for the conformational step responsible for the origin of the slow fluorescence transition, which further supports that the slow fluorescence transition results from Pol β's N-subdomain-reopening step.[138]
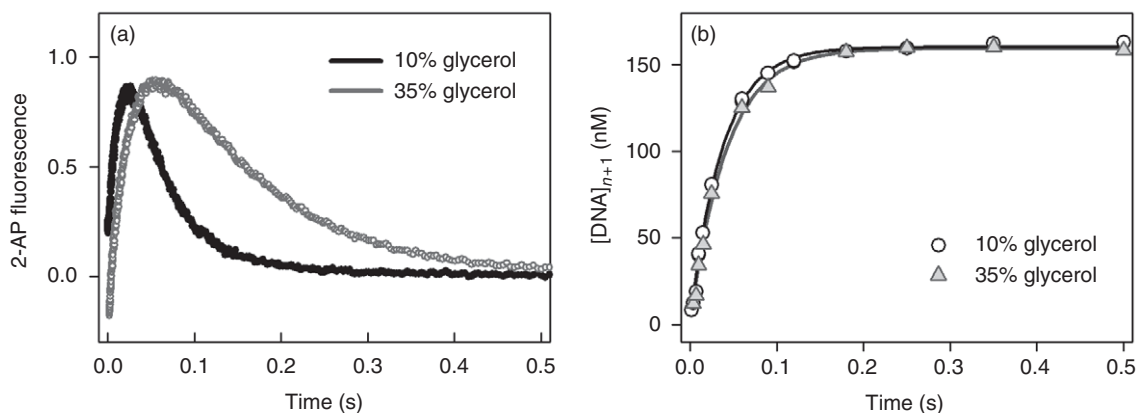
Figure 13   Comparison of Pol β catalyzed single-nucleotide incorporation at 10% and 35% glycerol, pH 8.3. (a) 2-AP fluorescence stopped-flow assays show that both phases of the fluorescence change are slowed down at increased glycerol concentration. (b) In contrast, rapid chemical quench assays demonstrate that the rate of nucleotide incorporation remains unaffected by the altered glycerol concentration. Adapted with permission from M. Bakhtina; M. P. Roettger; S. Kumar; M. D. Tsai, *Biochemistry* **2007**, *46*, 5463–5472. Copyright 2007 American Chemical Society.

8.10.4.2.3   Kinetic analysis of 2-AP fluorescence stopped-flow data

Having the two fluorescence transitions assigned to certain steps in a kinetic scheme, it is possible to extract kinetic constants from the dependence of the stopped-flow traces on dNTP concentration. Typically, the amplitude of fluorescence change and rates of fast and slow phases increase as the concentration of dNTP increases (**Figure 14**). The closed $E' \cdot D_n \cdot N$ and $E' \cdot D_{n+1} \cdot P$ complexes have a higher intensity of 2-AP fluorescence emission compared to binary $E \cdot D_n$ and $E \cdot D_{n+1}$ complexes (as well as open $E \cdot D_n \cdot N$ and $E \cdot D_{n+1} \cdot P$); therefore, upon increasing dNTP substrate concentration the increasing observed fluorescence amplitude signifies an increasing concentration of closed ternary complex, a pattern entirely consistent with the general polymerase kinetic mechanism depicted in **Figure 4**.

In agreement with the kinetic scheme, rates for both fast and slow fluorescence transitions show a hyperbolic dependence on dNTP concentration. For fast-phase analysis, observed rate constants plotted as a function of dNTP concentration should fit to a hyperbolic equation with a nonzero intercept.[139,144] According to the Pol β kinetic mechanism (**Figure 4**), fitting reveals values for the microscopic rate constant k_2 (conformational closing step) and dNTP thermodynamic dissociation constant $K_d = k_{-1}/k_1$. For slow-phase analysis, dNTP concentration dependence should fit to a simple rectangular hyperbola, and kinetic parameters obtained from this analysis correspond to k_{pol} and $K_{d,app}$ from single-turnover rapid chemical quench assays. Note again, that interpretation of the parameters obtained from kinetic analyses depends on the particular proposed kinetic scheme and could be inappropriate for certain DNA polymerases. One such case is Pol X (mentioned in Section 8.10.4.1.1), which binds dNTP prior to DNA substrate. The fast 2-AP fluorescence transition in Pol X stopped-flow studies corresponds to conformational rearrangements upon DNA binding to Pol X·dNTP binary complex.[113]

8.10.4.2.4   Measurement of the reverse rate of the conformational step

Conformational closing of the N-subdomain requires binding of the nucleotide substrate in complex with a divalent metal ion. Sequestering of Mg^{2+} by addition of EDTA eliminates the Mg·dNTP substrate from the reaction and thus shifts the equilibrium of the conformational step toward the open conformation (step 2 in **Figure 4**), which allows direct measuring of the reverse rate k_{-2}.[153] When the stopped-flow reaction is initiated by rapid mixing of preformed $E' \cdot D_n \cdot N$ ternary complex with excess EDTA, the observed 2-AP fluorescence change (**Figure 15**, pink trace) has the direction opposite to that elicited by the forward conformational closing initiated by mixing preformed $E \cdot D_n$ binary complex and Mg·dNTP (**Figure 15**, dark blue trace). Importantly, in the case of Pol β, the rate of reverse opening (k_{-2}) is faster than or comparable to the rate of the chemical step
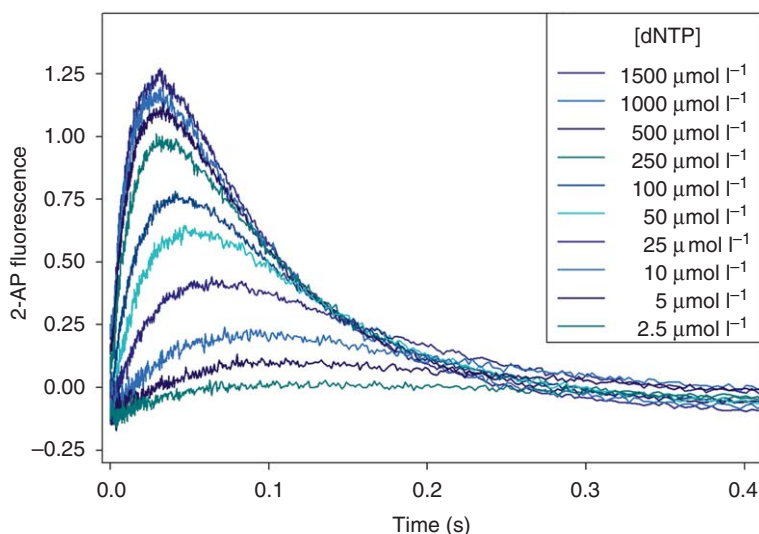
Figure 14   Concentration dependence of correct dNTP incorporation as monitored by stopped-flow 2-AP fluorescence assays.
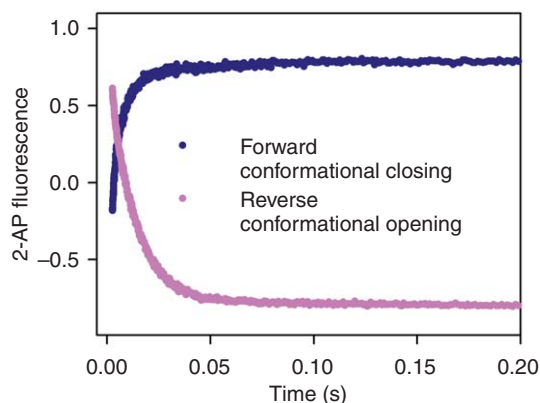
Figure 15 Forward conformational closing (dark blue) and reverse conformational opening (pink) as monitored by stopped-flow 2-AP fluorescence assays. Adapted with permission from M. Bakhtina; M. P. Roettger; M. D. Tsai, *Biochemistry* **2009**, *Submitted.* Copyright 2009 American Chemical Society.

(k_3). This result reinforces that the highest energy transition state of Pol β's reaction pathway corresponds to the chemical transition state and highlights the key role of chemistry in the fidelity of Pol β. Interestingly, the reverse opening rate of T7 DNA polymerase is quite slow relative to chemistry, suggesting that the conformational step plays a determining role in fidelity of this polymerase.[145,161]

8.10.4.2.5   *Pol β kinetic mechanism summary and comparison with other DNA polymerases*

The results of the aforementioned stopped-flow fluorescence studies of Pol β support the proposed model in **Figure 4**. It is worth noting that the relative energy barriers for chemical and conformational steps might be influenced by a number of factors such as pH, viscosity, DNA sequence, and interactions with other proteins. For instance, the stopped-flow results for Pol β suggest that the subdomain-closing step is faster than the rate of single-nucleotide incorporation by a factor of 100 at pH 6.1 and by less than a factor of 3 at pH 8.5. Obviously, the relative rates could differ for different DNA polymerases. However, there is evidence for a fast conformational change during correct dNTP incorporation among other DNA polymerases. Direct monitoring of N-subdomain motions in Klentaq1, using FRET, indicate that the subdomain closing is substantially faster than the rate of single-nucleotide incorporation.[147] Stopped-flow fluorescence and FRET studies of KF revealed fast dNTP binding-induced conformational changes prior to phosphodiester bond formation.[138,142,146] Besides the above empirical studies, recent computational analysis of the free energy landscape for correct and mismatched nucleotide incorporation by T7 DNA polymerase also indicate that the chemical step is rate limiting.[162] Nevertheless, even though current experimental and computational data are consistent with the model that chemistry is the rate-limiting step through phosphodiester bond formation for several DNA polymerases, it would be too hasty to conclude a common rate-limiting step for all members of this class of enzymes.

8.10.4.3   Dissection of the Role of Two Metal Ions

8.10.4.3.1   *Use of exchange-inert metals in stopped-flow analysis*

In addition to characterization of the two phases observed in stopped-flow fluorescence assays during Pol β correct nucleotide incorporation, studies using exchange-inert chromium(III) and rhodium(III) dNTP complexes have been able to further examine the order of magnesium ion binding in the active site during catalysis. Unlike Mg^{2+} nucleotide complexes, which are in rapid equilibrium in aqueous solution (with exchange rates of $5000\,s^{-1}$),[163] exchange-inert nucleotide complexes have ligand exchange rates measured in days under non-basic conditions.[164] Enzyme binding of two magnesium ions, namely consisting of a nucleotide-binding Mg^{2+} and a catalytic Mg^{2+} (as described in Section 8.10.3.3 and **Figure 3**), is a requirement for catalysis. Binding of

exchange-inert Cr(III)•dNTP to Pol β•DNA binary complex in the absence of Mg^{2+} induces the fast phase of fluorescence alone. Upon subsequent addition of Mg^{2+} the slow phase of fluorescence is restored.[165] Similarly, mixing of preformed Pol β•DNA•Rh(III)dNTP ternary complex with Mg^{2+} results in the slow fluorescence phase only. Upon mixing of Mg^{2+} with preformed Pol β•DNA•Rh(III)dNTP ternary complex in which the DNA primer is dideoxy-terminated, no stopped-flow signal is observed at all.[151] All of these results suggest that a fast conformational change (step 3, **Figure 4**) occurs upon metal•dNTP binding to the E•D_n binary complex, and support that binding of the catalytic Mg^{2+} (step 4, **Figure 4**) occurs after the formation of the E′•D_n•N ternary complex.

To address the possibility of obtaining unnatural results through use of the exchange-inert metal•dNTP complexes mentioned above, the respective difference in binding affinity for each of the two magnesium ions ($K_{d,app}^{Mg^{2+}} = 1.0\,\mathrm{mmol\,l^{-1}}$ and $K_{d,app}^{Mg\bullet dNTP} = 46\,\mathrm{\mu mol\,l^{-1}}$)[140] was employed to further test the Mg^{2+} binding order.[151] Stopped-flow results showed that upon rapid mixing of Pol β•DNA with dNTP under limiting magnesium concentrations which allow Mg•dNTP binding site saturation, yet do not provide sufficient Mg^{2+} to support catalysis, only the fast phase of fluorescence transition was observed. Paralleling the rhodium(III) experiments mentioned above, fluorescence monitoring of Pol β•DNA•MgdNTP ternary complex, preformed under limited magnesium concentration so as to prevent chemistry and then mixed with excess Mg^{2+}, demonstrates only the slow fluorescence transition as well. Such results with the use of Mg^{2+} instead of its analogues further corroborate conclusions of the mechanism studies using exchange-inert metal•dNTP complexes.

8.10.4.3.2 *Structural evidence for order of metal binding*

Further validation regarding the metal binding order proposed for the Pol β mechanism is found in the crystal structure of pathway intermediate, Pol β•DNA•Cr(III)dTMPPCP.[89] This ternary structure represents a fully functional pre-chemistry intermediate, as the primer retains the 3′OH and the product complex was observed upon soaking of the crystals in a solution containing divalent metal ion. Notably, in the absence of a bound catalytic metal ion, Pol β is found in a fully closed conformation. Importantly, this finding provides structural support to the conclusions from the aforementioned stopped-flow studies confirming that M•dNTP binding is sufficient to induce N-subdomain closure with no requirement for catalytic Mg^{2+} binding.

8.10.4.4 **Mismatched dNTP Incorporation**

Polymerase fidelity, defined by the ratio of the catalytic efficiencies for correct and incorrect dNTP incorporations, is governed by the free energy difference between the highest energy barriers along the correct and incorrect dNTP incorporation pathways. Since fidelity requires knowledge of both pathways, the disproportion of structural and mechanistic information available for matched dNTP incorporation compared with that for mismatched dNTP incorporation, gives rise to an ambiguous understanding of the molecular mechanism of DNA polymerase fidelity.[166,167] It must be noted that the means by which mismatch discrimination is accomplished is probably both polymerase and mispair specific. A variety of deviant kinetic pathways have been suggested for misincorporation among polymerases,[142,145,162,168–170] in which either: (1) mismatched dNTP incorporation occurs through a distinctly different pathway than matched dNTP, or (2) mismatched dNTP and matched dNTP incorporations proceed through analogous kinetic pathways, in which the energetics of the transition states and intermediate complexes differ. In either case, the presence of a mismatch within the active site probably generates a perturbed ternary complex conformation.

Gaining structural insights into the mismatched dNTP mechanism is a difficult task since true mismatched ternary complexes (E′•D_n•N•M, **Figure 4**) are not readily crystallized due to inherent thermodynamic instability. However, first insights into mismatch incorporation for various polymerases were resolved through complexes depicting mispairs within the active site[169,171] and mismatch extension.[172,173] For example, a post-chemistry structure of Pol β complexed with a nicked DNA substrate containing a terminal mispair shows only partial closing of the α-helix N, and a staggered arrangement of the terminal bases rather than normal Watson–Crick base pairing.[169] On the basis of these structures, it was proposed that mismatched dNTP incorporation may occur with a partially open N-subdomain conformation. Recently, a high-resolution functional pre-chemistry closed ternary complex for Pol β with an incoming mismatched nonhydrolyzable dAMPCPP demonstrates that the enzyme remains in a fully closed conformation comparable to that observed for matched

dNTP.[174] In the active site of this structure, however, the template DNA is shifted in order to accommodate the mismatched dNTP, such that no coding template base is in the nascent base pair binding pocket. The aberrant positioning of the DNA provides structural evidence for the mechanism of Pol β's inefficient misincorporation, as the 3′-OH of the primer, Pα of the incoming dNTP, and catalytic metal ion, do not demonstrate proper active-site geometry. It has been suggested that minor adjustments of active-site residues are required to bring about proper active-site geometry for catalysis, and that these local adjustments may play a critical role in defining Pol β's fidelity.[175,176]

In general, as discussed above, in the presence of a mismatch the polymerase active site does not fit properly as in Watson–Crick pairing, and catalysis is unfavorable due to less than ideal active-site geometry. As a result of improper alignment, polymerases generally incorporate mismatches with slower rates (smaller k_{pol}) and bind mismatched dNTP loosely (larger $K_{d,app}$ values).[107] Kinetic data for mismatched dNTP incorporations by many DNA polymerases have been reported using rapid chemical quench,[11,102,106,177–180] yet this commonly used kinetic approach does not allow characterization of individual steps in the mismatched dNTP incorporation pathway. To circumvent this, transient kinetic methods (as have been discussed in Sections 8.10.4.1 and 8.10.4.2 for matched dNTP incorporation), including stopped-flow and FRET, have been applied to examine mismatch incorporation, and they reveal important information about conformational motions of polymerases occurring during mismatch catalysis as highlighted in the following sections.

8.10.4.4.1   *Mismatched and matched dNTP incorporation occur through analogous kinetic pathways*

1. *Stopped-flow fluorescence reveals a fast conformational change for mismatched dNTP incorporation.* Recently, as an extension of previous methodology used to delineate Pol β's correct dNTP incorporation mechanism,[138,140,151] stopped-flow fluorescence and steady-state fluorescence spectroscopy have been employed to examine the mechanism of mismatched dNTP incorporation by Pol β.[139] These studies led to the conclusion that there is probably a conformational closing event that occurs for mismatched dNTP incorporation, as evidenced by the existence of a fast fluorescence phase preceding chemistry. In addition, the rate of the conformational change induced by mismatched dNTP was comparable to that of N-subdomain closing induced by correct dNTP. Furthermore, steady-state fluorescence studies demonstrated that both matched and mismatched dNTP elicit the same direction of fluorescence change (**Figure 16(a)**).
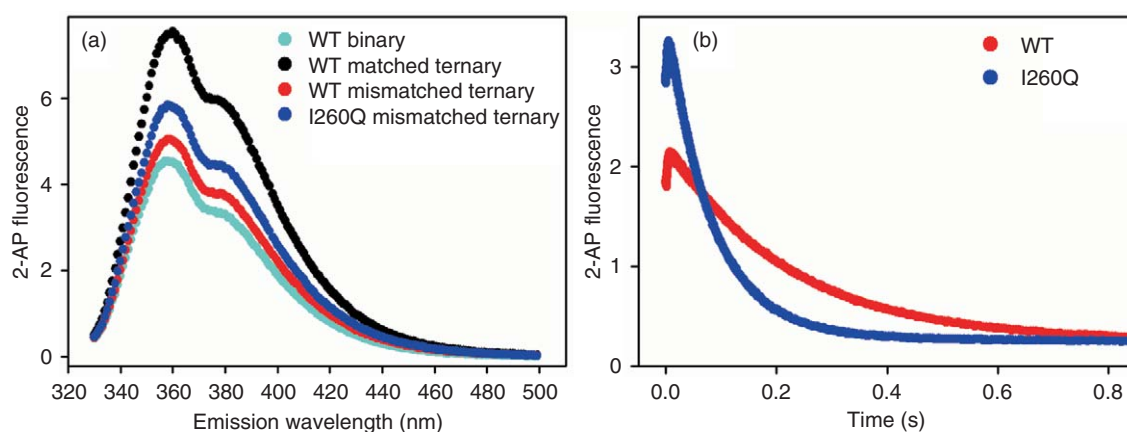
Figure 16   (a) Steady-state fluorescence spectra of matched and mismatched ternary complexes for WT Pol β and I260Q. Emission spectra shown include WT binary complex (cyan), WT T:A matched ternary complex (black), WT T:G mismatched ternary complex (red), I260Q T:G mismatched ternary complex (blue). Omitted for simplification are the I260Q binary and T:A matched ternary complexes, which overlay accurately with the corresponding WT traces. (b) Comparison of WT Pol β and I260Q T:G mismatched incorporation as monitored by 2-AP fluorescence assays. Reproduced with permission from M. P. Roettger; M. Bakhtina; M. D. Tsai, *Biochemistry* **2008**, *47*, 9718–9727. Copyright 2008 American Chemical Society.

Previous stopped-flow fluorescence assays investigating matched dNTP incorporation showed that both the fast and the slow fluorescence transitions demonstrated a hyperbolic dependence on dNTP concentration.[140,144] Similarly, the dNTP dependence of both the fast and the slow fluorescence phases during mismatched dNTP incorporation in stopped-flow has been examined. The observed rate constants for the fast and the slow phases, individually plotted as a function of dNTP concentration, reveal that both phases demonstrate a hyperbolic dependence on dNTP concentration (parameters obtained for k_2, K_d, k_{pol}, and $K_{d,app}$ as described in Section 8.10.4.2.3 and reported in **Table 1**). The observed hyperbolic dependence of the fast phase on mismatched dNTP largely indicates that this phase originates from a conformational change induced by mismatched dNTP binding.

Overall, all the aforementioned results support that, analogous to the prior assignments for correct dNTP incorporation, the fast and slow fluorescence changes observed for mismatched incorporation (**Figure 16(b)**) can be assigned to the dNTP-induced subdomain-closing conformational change and the chemical step (which probably limits the reopening step), respectively. An important observation is that the forward rate of conformational closing (k_2, **Figure 4**) for mismatched dNTP incorporation is comparable with that for correct dNTP incorporation (though with significant increase in K_d), while the maximum rate of nucleotide incorporation (k_{pol}) is substantially slower (also with significant increase in $K_{d,app}$) (**Table 1**). This suggests that overall mismatched incorporation follows a similar pathway, though both the K_d and $K_{d,app}$ values are higher and the rate of the chemical step is slower. These results are further supported by the ternary mismatch structure of Pol β which exists in the closed form.[174]

The observation that the conformational change occurs with a similar rate for both correct and mismatched incorporation by Pol β differs from the conclusion of the recent single molecule kinetic analysis of T7 DNA polymerase,[149] which reported a significantly reduced rate of conformational closing induced by mismatched dNTP binding. However, another study on T7 reported little difference in the forward rates of conformational closing between matches and mismatches, while noting a large difference in the reverse rates of conformational closing.[145] Additional studies on KF,[142] Dbh,[141] and T4[181] utilizing 2-AP fluorescence conclude on the basis of differences in fluorescence signals between matched and mismatched dNTP, so that there may exist distinctly different misincorporation pathways. It remains to be established whether such discrepancy reflects differences in the mismatched discrimination mechanism employed by various polymerases, or whether it results from different experimental systems and conditions.

Table 1 Kinetic comparison of rate and binding constants for WT versus I260Q for matched and mismatched dNTP incorporation

	dT:dA (matched)		dT:dG (mismatched)	
	Wild Type	I260Q	Wild Type	I260Q
K_2 (s^{-1})	116 ± 5	108 ± 7	256 ± 8	485 ± 25
K_d (µmol l^{-1})	29.5 ± 3.6	6.57 ± 2.12	488 ± 103	232 ± 37
k_{pol} (s^{-1})	42.9 ± 0.6	43.6 ± 1.4	5.70 ± 0.11	13.6 ± 0.2
$K_{d,app}$ (µmol l^{-1})	6.79 ± 0.48	7.07 ± 1.20	489 ± 26	48.8 ± 2.5
$k_{pol}/K_{d,app}$[a]	6.32	6.17	0.0117	0.279
Fidelity[b]	–	–	541	23.1

[a] Catalytic efficiency measured in units s^{-1} µmol^{-1} l.

[b] Fidelity defined as $[(k_{pol}/K_{d,app})_{cor} + (k_{pol}/K_{d,app})_{inc}]/(k_{pol}/K_{d,app})_{inc}$ – where the subscripts 'cor' and 'inc' indicate the correct (matched) and incorrect (mismatched) nucleotide incorporation, respectively.

The k_2, K_d, k_{pol}, and $K_{d,app}$ values were obtained from hyperbolic fit of the dNTP concentration dependence of the observed rates of the fast and slow fluorescence phases as described in Section 8.10.4.2.3. The k_2 value represents the rate constant of forward conformational closing, while K_d reflects the stability of the ternary complex before closing. Therefore, WT and I260Q show little difference during initial dNTP binding before the conformational change. The k_{pol} value represents the maximum rate of dNTP incorporation, while the $K_{d,app}$ value possesses a contribution from all steps up to the rate-limiting step and can be thought as the dissociation constant of the closed ternary complex.

2. *Rate-limiting chemistry for matched and mismatched dNTP incorporation.* Comparative kinetic analyses of Pol β strongly support that the chemistry step is rate limiting for both matched and mismatched incorporation pathways.[139] Similarly, results of linear free energy relationship (LFER) studies utilizing dNTP analogues, in which the β,γ-bridging oxygen is substituted with various halomethylene moieties, suggests that for both matched and mismatched incorporation P–O bond breaking makes a significant contribution to the rate-limiting step.[118,119] Based on the observation that the Brønsted correlation between $\log k_{pol}$ and the leaving group pK_a for monohalogenated analogues is very similar between correct and incorrect dNTP incorporations, it is concluded that the corresponding transition states have similar positions on the free energy surface. However, notably different Brønsted correlations observed with bulkier dihalogenated analogues suggest the existence of structural differences at the chemical transition states of correct and mismatched incorporations. In addition, analysis of pH dependence and solvent deuterium isotope effects revealed that a proton transfer step (steps) might be at least partially rate limiting for both matched and mismatched dNTP insertion.[119] Overall, these results support the proposal that the rate-limiting chemical step is a major contributor to Pol β fidelity.[167]

8.10.4.4.2 *Correlation between fidelity and mismatched transition state destabilization*

Site-directed mutagenesis allows us to investigate fidelity variances consequent of single residue mutations of a single polymerase. The Ile260 residue of Pol β is located in the hydrophobic hinge region between the C- and the N-subdomains.[77,87] The mutator activity of site-directed mutant I260Q was first identified by a genetic screen,[182] and subsequent pre-steady-state kinetic characterization showed that it possessed a low fidelity due to loose binding discrimination of mismatched dNTP substrates.[183] These characteristics make the I260Q mutant of specific interest in fidelity studies of Pol β, in order to further understand how one mutation can alter the mismatch discrimination profile of this enzyme. Mechanistic studies using stopped-flow fluorescence demonstrated both enzymes possess similar correct dNTP incorporation profiles (**Figure 17**, black profile), and that the main difference between I260Q and WT lies in the ability of I260Q to more efficiently stabilize the mismatched ternary complex, as suggested by the 10-fold decrease in $K_{d,app}$ of I260Q ($48.8 \pm 2.5 \, \mu mol \, l^{-1}$) relative to WT ($489 \pm 26 \, \mu mol \, l^{-1}$) for dT:dG mismatch incorporation (**Table 1**). This was further corroborated by an observed higher amplitude for I260Q in stopped-flow fluorescence traces and steady-state emission spectra for mismatched dNTP (**Figure 16**).[139] Overall, comparison studies between WT Pol β and the lower fidelity I260Q variant illustrate the correlation between fidelity and mismatch destabilization, as the infidelity of I260Q originates from enhanced stabilization of the mismatched ternary complex and the chemical transition state (**Figure 17**, blue trace).[139] These results strongly support that both matches and mismatches are incorporated through analogous mechanisms, and that the fidelity of Pol β is controlled, at least partly, by
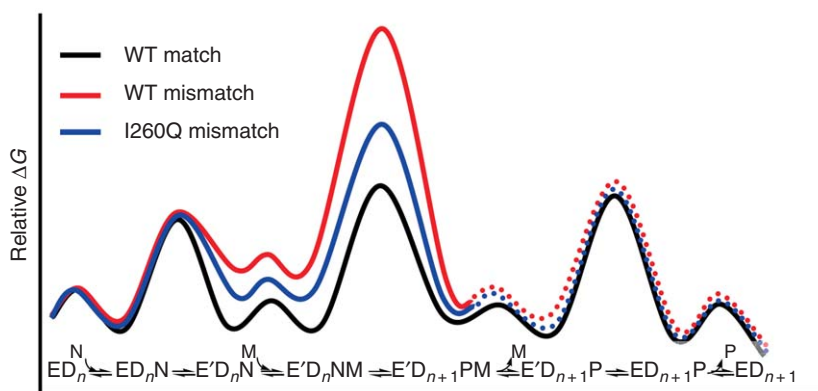
Figure 17   Qualitative free energy profile of matched and mismatched dNTP incorporation by WT Pol β versus I260Q. E = DNA polymerase in open conformation; E′ = closed conformation; D_n = DNA; N = Mg·dNTP; M = catalytic Mg^{2+}; P = Mg·PP$_i$. Reproduced with permission from M. P. Roettger; M. Bakhtina; M. D. Tsai, *Biochemistry* **2008**, *47*, 9718–9727. Copyright 2008 American Chemical Society.

destabilization of the mismatched ternary complex and the chemical transition state in the same reaction pathway (**Figure 17**, red profile).

The observation of an incorrect dNTP-induced conformational change by Pol β may appear to contradict the findings of two FRET-based studies monitoring the conformational motions of Klentaq[147] and KF.[148] Both studies report an increase in FRET signal, upon addition of correct dNTP, yet do not observe any noticeable change upon addition of incorrect dNTP. In contrast to moderate and low-fidelity polymerases, including Pol β, since Klentaq and KF are higher fidelity enzymes, it is probable that they more effectively destabilize the mismatched ternary complex to the extent that no fluorescence change is observable for mismatch binding. This is comparable to the aforementioned case where Pol β demonstrates significantly reduced fluorescence change amplitude in stopped-flow and steady-state experiments compared to the lower fidelity I260Q variant. Paralleling the differences between WT Pol β and I260Q, the differences between high-fidelity Klentaq and KF and lower fidelity Pol β further emphasize the correlation between fidelity and mismatch destabilization.

8.10.5   Computational Studies

8.10.5.1   Molecular Dynamics Simulations

Kinetic data provide mechanistic details of catalysis and lend insight into structural changes occurring for an enzyme during catalysis. Complementarily, crystal and solution structures provide reliable, yet static, pictures of reaction systems during different stages of catalysis. Importantly, modeling and simulation methods bridge the remaining gaps in functional and structural experimental data by providing further information required for a detailed understanding of local and global motions involved along a reaction pathway. In addition to complementing information already obtained from experimental data, modeling studies incite potentially testable experiments regarding enzyme mechanism, and also provide insight into questions that may be experimentally nontestable. In spite of their usefulness in approaching a varied range of problems, molecular mechanics (MM)-based methods have inherent shortcomings because of the use of 'crude' approximations and imperfect force fields, as well as dependence on initial models. As a result, MD simulations of a single DNA polymerase, which employ initial models derived from structures with only minor active-site differences, can produce ambiguous or even conflicting results. However, with the availability of more refined polymerase structures, including structures with well-defined water molecules and catalytic metal ions, the information extracted by MD simulations is continually improving.

MD simulations of the Pol β•DNA system with and without dNTP substrates have been used to delineate the microscopic motions involved in the formation of the catalytic ternary complex. The conformational closing of the N-subdomain upon dNTP binding is accompanied by several functionally relevant movements in key Pol β active-site residues, including the rotation of Asp192 to coordinate with the catalytic Mg^{2+}, rotation of Arg258 to interact with Tyr296, and base-flipping of Phe272 to prevent interaction between Arg258 and Asp192 (**Figure 18**).[184] The order of events and the intermediate states in the conformational closing of Pol β have been further investigated using transition path sampling[159] and stochastic path approaches.[185] These studies show that the sequenced order of events during the N-subdomain closing event is as follows: partial N-subdomain closing, flip of Asp192, partial rotation of Arg258, and completion of N-subdomain closure, flip of Phe272, followed by rearrangement of catalytic region and stabilization of Arg258 in the fully rotated state. The computed conformational landscape shows that the cascade of events along this pathway is highly cooperative. For example, the open ternary complex undergoes partial N-subdomain closing, with a concomitant change in the puckering of the sugar of the incoming dNTP which facilitates base pairing with the templating base. This motion is also coupled with the additional motions involving the incoming dNTP, templating base, and Tyr271 residue to achieve Watson–Crick base pairing. Similarly, the flip of Asp192 is accompanied by the breakage of the salt-bridge between Asp192 and Arg258 residues (**Figure 18**). In addition to the cooperative motion of residues and substrates, the position and the coordination of the catalytic Mg^{2+} also undergo subtle but systematic transformations.[159] The computed reaction profile and associated free energy barriers suggest that the partial rotation of the Arg258 residue is rate limiting within the conformational closing step.[159] Further computational analysis of the R258A mutant predicted facilitated N-subdomain reopening after chemistry,[186] which was later corroborated experimentally in stopped-flow fluorescence analyses.[138]
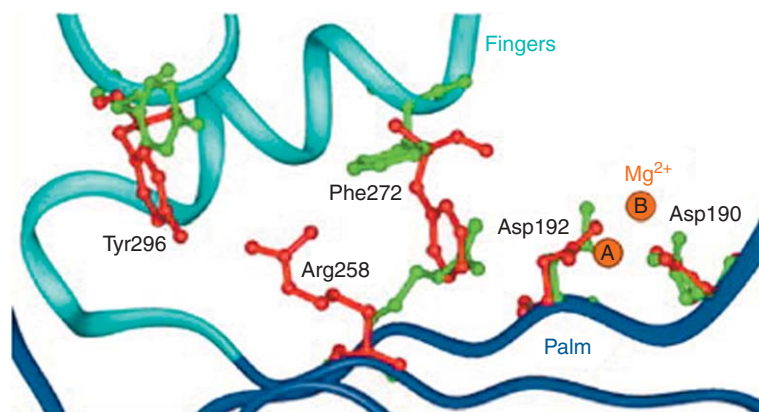
Figure 18 Motion of Arg258 upon N-subdomain (fingers subdomain) closure. Green residues denote the position assumed in the Pol β open state, whereas red residues denote the position assumed in the closed catalytic state. Adapted from L. Yang; W. A. Beard; S. H. Wilson; S. Broyde; T. Schlick, *Biophys. J.* **2004**, *86*, 3392–3408. Permission kindly granted by Tamar Schlick. Copyright 2004 Biophysical Society.

In addition to modeling of the correct dNTP incorporation by Pol β, the influence of the mismatches in the active site of the polymerase has also been investigated using MD simulations. MD simulations indicate that mismatched base pairs at the primer terminus of DNA influence the closing motion of the α-helix N of the N-subdomain, which in turn hampers conformational closing prior to chemistry.[187] Additionally, in these simulations the terminal mismatch base pairs do not adopt a planar conformation, and the presence of a mismatch is shown to influence the ligand arrangement around the Mg^{2+} ions. Transition path sampling simulations comparing G:C matched incorporation versus G:A mismatched incorporation suggest that the cascade of transition states experienced in Pol β conformational closing during mismatched dNTP incorporation is different from the cascade incurred during matched incorporation, and that the mismatched incorporation reaction exhibits a more transient closed state overall.[188] These simulations also suggest that the rate-limiting step for both matched and mismatched nucleotide incorporation pathways occurs after dNTP-induced conformational closing, but prior to the actual phosphoryl transfer. The nature of this step is thought to entail subtle active-site adjustments, including slow adjustments of critical metal/phosphoryl coordinations in the active site. This concept led to the emergence of the term 'pre-chemistry avenue,' used to denote the scenario following N-subdomain conformational closing in which the aforementioned active-site adjustments are required to facilitate an active-site geometry that is poised to support nucleotidyl transfer.[175] As mentioned earlier, it should be noted in terms of general DNA polymerase mechanism discussion, that these local active-site rearrangements should not mechanistically be considered as kinetically distinct steps, but rather considered as a part of the overall chemical step.

In an effort to estimate the contributions of the nucleotide binding, conformational change, and the chemical step to the overall fidelity of T7 DNA polymerase, empirical valence bond (EVB) and linear response approximation (LRA) approaches have been used to calculate the free energy landscape for correct and incorrect dNTP incorporations.[162] **Figure 19** shows the free energy profile as a function of two coordinates: one corresponds to the dNTP-induced conformational closing, and the other corresponds to the chemical reaction. The energy barrier for the conformational closing, for both correct and incorrect dNTP incorporations, is much lower than that for the chemical step, suggesting that the N-subdomain closing is not the rate-determining step. Furthermore, the energetically preferred pathway for mismatched dNTP incorporation is suggested to occur through a partially open enzyme conformation. EVP-LRA methods were also used to explore the role of conformational changes in the fidelity of Pol β.[189] These studies suggest that, similar to T7 polymerase, Pol β also incorporates correct dNTP through a transition state in the closed conformation, whereas incorrect incorporations are realized through transition states in partially open conformations. Consistent with this proposal, recent studies using small-angle X-ray scattering (SAXS) also indicate that incorrect dNTP incorporations catalyzed by Pol β may proceed through a partially open (or partially closed) ternary complex.[190]
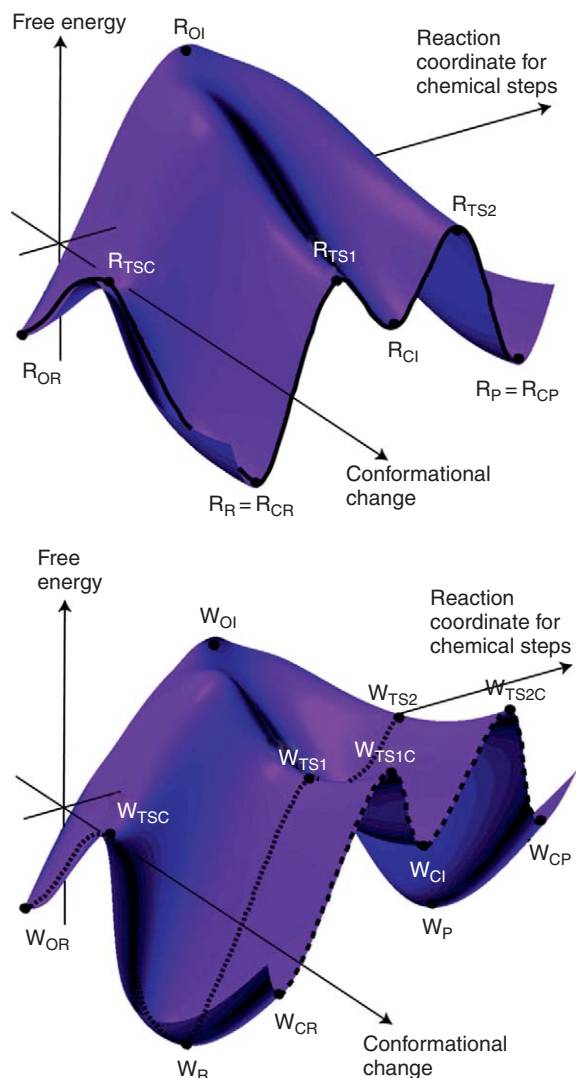
Figure 19   Free energy surfaces for correct (R, top) and incorrect (W, bottom) incorporations by T7 DNA Polymerase. TSC, TS1, and TS2 subscripts refer to transition states for conformational change, O3′-Pα bond formation, and pyrophosphate departure, respectively. W$_{TS1C}$ and W$_{TS2C}$ refer to transition states for mismatched incorporation in a fully closed conformation. Adapted with permission from J. Florian; M. F. Goodman; A. Warshel, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6819–6824. Copyright 2005 National Academy of Sciences.

8.10.5.2   Quantum Mechanical (QM) Studies of the Chemical Step

Although it is important to study the polymerase mechanism in its entirety, it is also interesting to investigate the unfolding of events within the chemical step itself, especially in relation to achievement of the transition state. MM methods do not adequately treat the electronic changes associated with the bond formation/cleavage processes and hence, more sophisticated approaches are required to understand the chemical transformations during catalysis. QM computational studies provide mechanistic details of the catalytic steps at the atomic level and provide useful insights into the origin of mismatch discrimination or fidelity. However, care should be taken while interpreting the results of such analyses because the results show a strong dependence on the initial molecular model.[191]

    The mechanism of nucleotidyl transfer during correct dNTP incorporation by DNA polymerases has been investigated by QM methods. The postulated pathway of nucleotidyl transfer includes three critical

microscopic steps. The first step corresponds to the deprotonation of the 3′-OH of the primer. The next step includes nucleophilic attack of the O3′ of the primer on Pα (α-phosphate of dNTP), and is followed by final elimination of the pyrophosphate group.

Several different hypotheses have been proposed for the initial 3′-OH deprotonation event including, direct transfer of the proton to O2α(Pα), transfer of this proton to an active-site aspartate residue, or initial proton transfer to an active-site water molecule, followed by proton migration to an active-site aspartate residue, and finally to pyrophosphate. Starting with an initial model of the active site derived from the Pol β ternary complex structure containing both the catalytic Mg^{2+} ions and the 3′-OH terminal of the primer,[90] QM studies of the chemical step exploring all three hypotheses conclude that the favored catalytic route involves direct proton transfer from O3′ of the primer to the O2α(Pα) of the dNTP through an associative mechanism (**Figure 20**). This leads to the formation of a pentacovalent trigonal bipyramidal Pα center, and is followed by the cleavage of the triphosphate unit and subsequent elimination of the pyrophosphate group.[191] In contrast, computational studies using EVB and QM/MM methods, reveal alternative routes of proton transfer, to either an adjacent Asp residue[192–194] or an active-site water molecule.[195]

Despite several studies focusing on the proton transfer step, there is discrepancy regarding the energetics of individual microscopic steps during nucleotidyl transfer, as QM methods may provide inconsistent results depending on the starting geometry and the computational approach. In the following subsection, we describe the coupling of QM treatment of the active site with empirical simulation of the complete enzyme system.
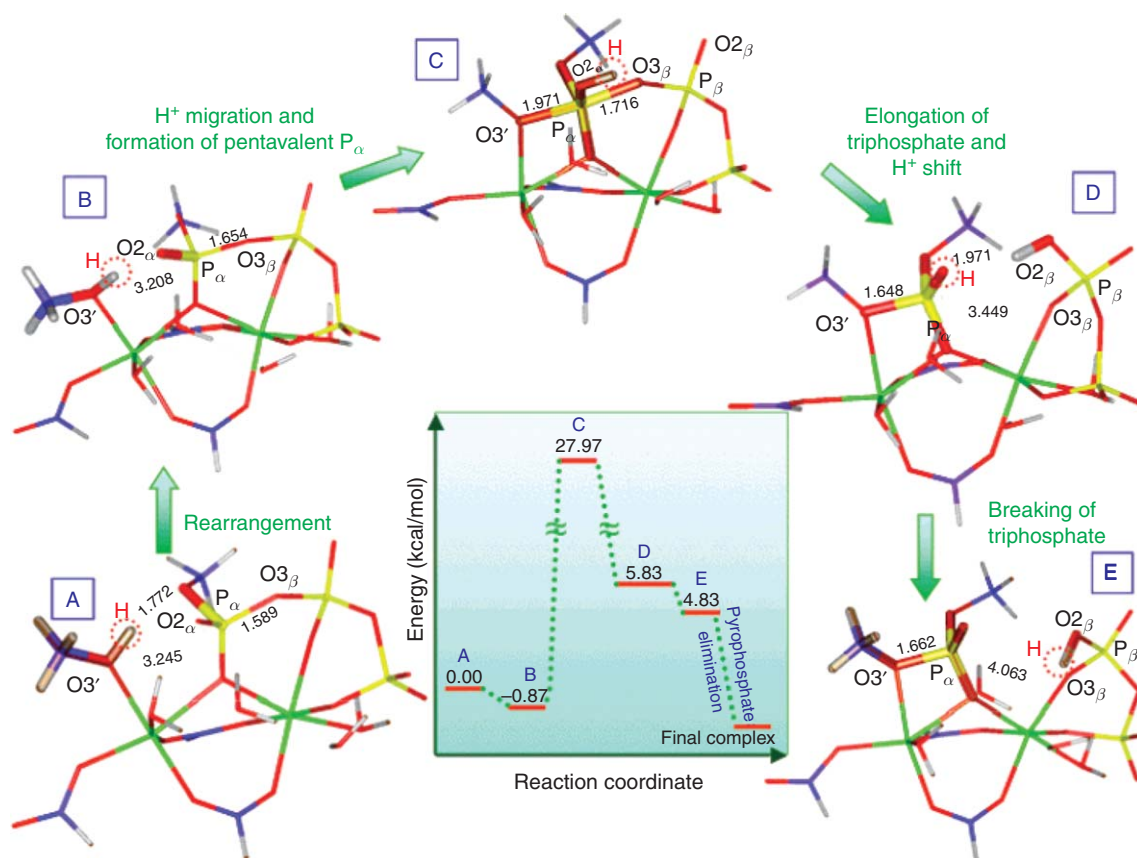
Figure 20 Reaction coordinates for the chemical step of Pol β catalyzed correct nucleotide incorporation and the corresponding structures for each step. Only one of three distinct possible pathways of proton transfer (direct transfer from O3′ to O2α(Pα)) is shown. The atoms are shown as color-coded sticks: Mg (green), P (yellow), O (red), C (blue), and H (gray). Energies are in $kcal\,mol^{-1}$. Reproduced with permission from M. D. Bojin; T. Schlick, *J. Phys. Chem. B* **2007**, *111*, 11244–11252. Copyright 2007 American Chemical Society.

8.10.5.3   Mixed QM/MM Studies of the Catalytic Mechanism

Stand-alone QM approaches, by necessity, ignore protein residues and DNA bases not directly involved in catalysis, and additionally do not account for solvent effects. In an attempt to overcome these limitations, several new studies have adopted a mixed quantum and molecular mechanical (QM/MM) approach to study the fully solvated enzyme system. The QM/MM ONIOM approach, has been used to examine the energetic landscape for correct dNTP incorporation by Pol β.[193] This analysis shows that nucleotidyl transfer, following the formation of the ternary complex, is a two-step process in which proton transfer occurs prior to the nucleophilic displacement. The favored route of proton transfer from O3′ to the adjacent Asp256 residue has a lower energy barrier than does the nucleophilic attack (**Figure 21**). The latter is associated with an energy barrier of $14.6\,\mathrm{kcal\,mol^{-1}}$, comparable to the experimentally derived value of $\sim16\,\mathrm{kcal\,mol^{-1}}$. Interestingly, in this study, no stable pentacovalent intermediate was identified.[193] Another study, employing QM/MM dynamics in conjunction with umbrella sampling to estimate free energy of intermediates during the phosphoryl transfer reaction, showed that the dominant pathway of the initial step involves proton transfer from the O3′ to water molecules, followed by proton migration to the Asp256 residue in a series of Grotthuss hopping steps.[195] This pathway of initial proton transfer was also corroborated in another study exploring several different potential mechanisms of the phosphoryl transfer reaction using QM/MM methods.[196] However, this study suggests that the initial deprotonation occurs in conjunction with the nucleophilic attack by O3′ on Pα, with an associated energy barrier of $15.4\,\mathrm{kcal\,mol^{-1}}$.

Until recently, a detailed understanding of mismatch discrimination by DNA polymerases has been hampered by the lack of structural information on mismatch incorporation by a polymerase. The recent structure of a Pol β mismatch ternary complex[174] has opened the gates for computational analyses of molecular basis of fidelity discrimination. Starting with this structure detailed molecular dynamics (MD) and mixed
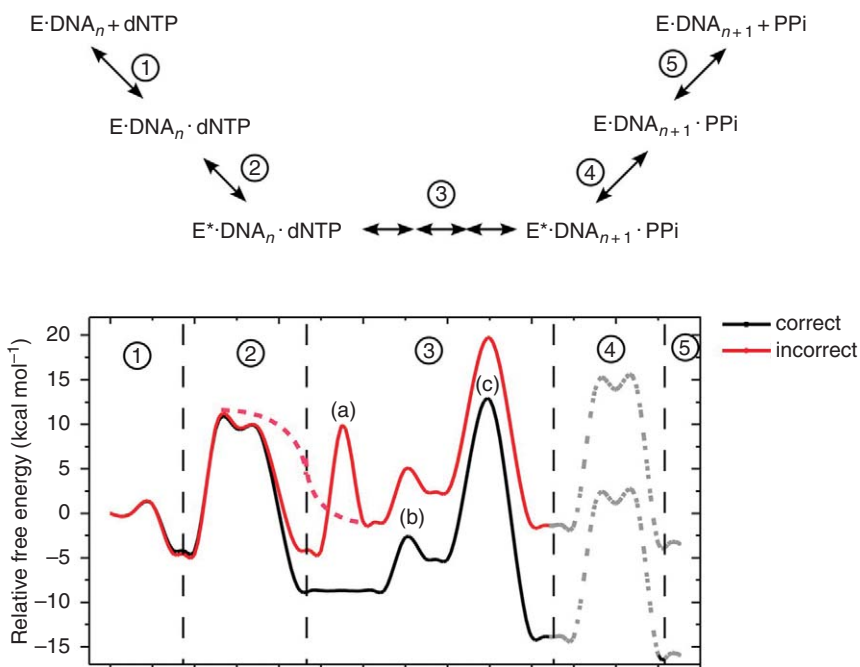
Figure 21   Schematic representation of the free energy profiles for correct (black) and incorrect (red) nucleotide incorporations by Pol β. Reactions follow: initial binding of dNTP (1), open-to-closed conformational change (2), O3′ replacing water in Mg_{cat}^{2+} coordination (3a, for misincorporation only), proton transfer from O3′ to Asp256 (3b), phosphodiester bond formation (3c), closed-to-open conformational change (4), and release of pyrophosphate and Mg ions (5). The dashed lines represent possible transition pathways and barriers. Adapted with permission from P. Lin; V. K. Batra; L. C. Pedersen; W. A. Beard; S. H. Wilson; L. G. Pedersen, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5670–5674. Copyright 2008 by the National Academy of Sciences.

QM/MM computational analyses of the mechanism of G:A mismatch incorporation by Pol β have been performed.[176] The results suggest that for G:A incorporation the most stable conformation of the mismatched ternary complex undergoes additional local structural changes in the active site. These local changes include replacement of an active-site water molecule by O3′ as a ligand for the catalytic Mg^{2+}, resulting in catalytically competent active-site geometry similar to that observed for match G:C incorporation.[193] After this reorganization, the enzyme proceeds to follow the catalytic pathway similar to that for the correct insertion, which involves proton transfer followed by nucleophilic displacement. Similarities in the transition state geometries for matched and mismatched incorporations emphasize the importance of the reaction scaffold formed by triphosphate moiety of dNTP, active site Asp residues, and catalytic Mg^{2+} ions, required for phosphodiester bond formation. The analysis of the energetics of the mismatched versus matched incorporation pathways suggests that the free energy required for destabilization of the ground state to attain a catalytically competent active-site geometry is the major factor contributing to fidelity discrimination (**Figure 21**).[176] This conclusion also correlates well with kinetic evaluations which suggest that the destabilization of the mismatched ternary complex plays an important role in DNA polymerase fidelity.[139]

Over the years, computational methods have elucidated various aspects of the nucleotidyl transfer reaction that complement our knowledge from experimental methods, and provide enhanced insights into the overall mechanism of DNA polymerases. Dynamics simulations have shed light on the subtle and cooperative molecular motions involved during the dNTP-induced conformational closing in Pol β. Calculated energy profiles suggest that this conformational change is not rate limiting in the overall polymerase mechanism. Although there is some ambiguity in literature regarding whether the initial deprotonation of O3′H group is followed by nucleophilic attack by O3′ or whether these processes occur in conjunction, it is generally agreed that pentacovalent transition state formation corresponds to the highest energy barrier within the chemical step and is therefore rate limiting in the overall dNTP incorporation pathway. Quantum mechanical analyses of the catalytic mechanism have shown that the active-site rearrangements prior to nucleotidyl transfer are critical to the overall polymerase mechanism. Both matched and mismatched dNTP incorporation probably follow similar pathways in ultimately achieving catalytically competent active-site geometry prior to nucleotidyl transfer. However, to achieve this geometry, mismatched incorporation may require additional local active-site rearrangements.

Even though this chapter has largely focused on the mechanism of correct dNTP incorporation and the molecular origin of fidelity, the computational approaches described here can also be utilized to study the effects of modified substrates in the active site of polymerases,[197] as well as the effect of interacting proteins on catalysis.[198] Overall, computational methods will continue to be a valuable tool for DNA polymerase mechanism studies. A complete understanding of polymerase mechanism and fidelity will surely be facilitated by future advancements in computational methodology, as well as the availability of new complex structures.

8.10.6 Final Thoughts

Over the past several decades, the collaborative efforts of multiple approaches have brought us to a more comprehensive understanding of the DNA polymerase catalytic mechanism. In attempts to understand the origin of polymerase fidelity, efforts over the past decade encompass extensive pursuits to identify the rate-limiting step during dNTP incorporation, and to provide a structural basis for DNA polymerase selectivity. Most notable progress includes the demonstration that the dNTP-induced subdomain conformational closing is a fast step relative to chemistry. Moreover, large advances in our understanding of mismatch incorporation have been obtained from the crystal structure of a mismatched ternary complex,[174] and results of extensive stopped-flow fluorescence analysis of incorrect dNTP incorporation.[139] The main focus of future research will probably be on detailed structural and mechanistic comparison of mismatched dNTP incorporation with matched dNTP incorporation, in which recent insights into mismatched dNTP incorporation is further investigated with other methods. Additionally, differences among various mispairs also warrant further investigation. Most importantly, in order to fully understand how DNA polymerases control fidelity, the methodological depth of structural and kinetic analyses must be applied to multiple DNA polymerases exhibiting the entire spectrum of fidelity.

Abbreviations

2-AP	2-aminopurine
AP	apurinic/apyridinic
BER	base excision repair
dAMPCPP	2′-deoxyadenosine 5′-α,β-methylenetriphosphate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dRP	deoxyribosephosphate
DSB	double-strand break
dTMPPCP	2′-deoxythymidine 5′-β,γ-methylenetriphosphate
EVB	empirical valence bond
FRET	fluorescence resonance energy transfer
ICL	interstrand cross-link repair
KF	Klenow fragment
LRA	linear response approximation
MD	molecular dynamics
MM	molecular mechanics
MMR	mismatch repair
NER	nucleotide excision repair
PAD	polymerase-associated domain
Pol	polymerase
QM	quantum mechanics
RNA	ribonucleic acid
SHM	somatic hypermutation
TLS	translesion synthesis
V(D)J	mechanism of genetic recombination
WT	wild type

Nomenclature

Units:

kDa	kilodalton; mass
kcal mol^{-1}	kilocalories per mole; energy
s^{-1}	per second; reaction rate
μmol l^{-1}	micromolar; concentration

Terms:

$k_{-1}, k_1, k_{-2}, k_2, k_4$	kinetic parameters in polymerase mechanism
K_d	dissociation constant
$K_{d,app}$	apparent dissociation constant
k_{pol}	rate of polymerization

References

1. P. V. Shcherbakova; K. Bebenek; T. A. Kunkel, *Sci. Aging Knowledge Environ.* **2003**, *2003*, RE3.
2. U. Hubscher; G. Maga; S. Spadari, *Annu. Rev. Biochem.* **2002**, *71*, 133–163.
3. Y. I. Pavlov; P. V. Shcherbakova; I. B. Rogozin, *Int. Rev. Cytol.* **2006**, *255*, 41–132.
4. S. D. McCulloch; T. A. Kunkel, *Cell Res.* **2008**, *18*, 148–161.
5. T. A. Kunkel, *Cancer Cell* **2003**, *3*, 105–110.
6. D. Starcevic; S. Dalal; J. B. Sweasy, *Cell Cycle* **2004**, *3*, 998–1001.
7. M. Raszka; N. O. Kaplan, *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 2025–2029.
8. F. Aboul-ela; D. Koh; I. Tinoco; F. H. Martin, *Nucleic Acids Res.* **1985**, *13*, 4811.
9. L. A. Loeb; T. A. Kunkel, *Annu. Rev. Biochem.* **1982**, *51*, 429–457.
10. J. Ahn; V. S. Kraynov; X. Zhong; B. G. Werneburg; M. D. Tsai, *Biochem. J.* **1998**, *831*, 79–87.
11. J. Ahn; B. G. Werneburg; M. D. Tsai, *Biochemistry* **1997**, *36*, 1100–1107.
12. T. A. Kunkel, *J. Biol. Chem.* **2004**, *279*, 16895–16898.
13. T. A. Kunkel; K. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497–529.
14. M. J. Bessman; A. Kornberg; I. R. Lehman; E. S. Simms, *Biochim. Biophys. Acta* **1956**, *21*, 197–198.
15. I. R. Lehman; M. J. Bessman; E. S. Simms; A. Kornberg, *J. Biol. Chem.* **1958**, *233*, 163–170.
16. M. Delarue; O. Poch; N. Tordo; D. Moras; P. Argos, *Protein Eng.* **1990**, *3*, 461–467.
17. D. K. Braithwaite; J. Ito, *Nucleic Acids Res.* **1993**, *21*, 787–802.
18. J. Ito; D. K. Braithwaite, *Nucleic Acids Res.* **1991**, *19*, 4045–4057.
19. H. Ohmori; E. C. Friedberg; R. P. Fuchs; M. F. Goodman; F. Hanaoka; D. Hinkle; T. A. Kunkel; C. W. Lawrence; Z. Livneh; T. Nohmi; L. Prakash; S. Prakash; T. Todo; G. C. Walker; Z. Wang; R. Woodgate, *Mol. Cell* **2001**, *8*, 7–8.
20. J. B. Sweasy; J. M. Lauper; K. A. Eckert, *Radiat. Res.* **2006**, *166*, 693–714.
21. A. Kornberg, Ed., *DNA Replication*; W. H. Freeman: San Francisco, 1980.
22. A. Kornberg; T. A. Baker, Eds., *DNA Replication*; W. H. Freeman: New York, 1992.
23. L. S. Kaguni, *Annu. Rev. Biochem.* **2004**, *73*, 293–320.
24. M. A. Graziewicz; M. J. Longley; W. C. Copeland, *Chem. Rev.* **2006**, *106*, 383–405.
25. M. J. Longley; R. Prasad; D. K. Srivastava; S. H. Wilson; W. C. Copeland, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12244–12248.
26. S. W. Graves; A. A. Johnson; K. A. Johnson, *Biochemistry* **1998**, *37*, 6050–6058.
27. K. Takata; T. Shimizu; S. Iwai; R. D. Wood, *J. Biol. Chem.* **2006**, *281*, 23445–23455.
28. M. E. Arana; K. Takata; M. Garcia-Diaz; R. D. Wood; T. A. Kunkel, *DNA Repair* **2007**, *6*, 213–223.
29. M. Seki; R. D. Wood, *DNA Repair* **2008**, *7*, 119–127.
30. M. Seki; F. Marini; R. D. Wood, *Nucleic Acids Res.* **2003**, *31*, 6117–6126.
31. M. Seki; C. Masutani; L. W. Yang; A. Schuffert; S. Iwai; I. Bahar; R. D. Wood, *EMBO J.* **2004**, *23*, 4484–4494.
32. K. Masuda; R. Ouchida; A. Takeuchi; T. Saito; H. Koseki; K. Kawamura; M. Tagawa; T. Tokuhisa; T. Azuma; J. O-Wang, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13986–13991.
33. H. Zan; N. Shima; Z. Xu; A. Al-Qahtani; A. J. Evinger Iii; Y. Zhong; J. C. Schimenti; P. Casali, *EMBO J.* **2005**, *24*, 3757–3769.
34. K. Masuda; R. Ouchida; M. Hikida; M. Nakayama; O. Ohara; T. Kurosaki; J. O-Wang, *DNA Repair* **2006**, *5*, 1384–1391.
35. H. Pospiech; J. E. Syvaoja, *Sci. World J.* **2003**, *3*, 87–104.
36. R. Hindges; U. Hubscher, *Biol. Chem.* **1997**, *378*, 345–362.
37. I. R. Lehman; L. S. Kaguni, *J. Biol. Chem.* **1989**, *264*, 4265–4268.
38. A. Johnson; M. O'Donnell, *Annu. Rev. Biochem.* **2005**, *74*, 283–315.
39. C. W. Lawrence, *Adv. Protein Chem.* **2004**, *69*, 167–203.
40. G. N. Gan; J. P. Wittschieben; B. O. Wittschieben; R. D. Wood, *Cell Res.* **2008**, *18*, 174–183.
41. Z. Kelman; M. O'Donnell, *Annu. Rev. Biochem.* **1995**, *64*, 171–200.
42. M. O'Donnell; D. Jeruzalmi; J. Kuriyan, *Curr. Biol.* **2001**, *11*, R935–R946.
43. I. K. Cann; Y. Ishino, *Genetics* **1999**, *152*, 1249–1267.
44. Y. Kubota; R. A. Nash; A. Klungland; P. Schar; D. E. Barnes; T. Lindahl, *EMBO J.* **1996**, *15*, 6662–6670.
45. I. D. Nicholl; K. Nealon; M. K. Kenny, *Biochemistry* **1997**, *36*, 7557–7566.
46. R. W. Sobol; J. K. Horton; R. Kuhn; H. Gu; R. K. Singhal; R. Prasad; K. Rajewsky; S. H. Wilson, *Nature* **1996**, *379*, 183–186.
47. S. S. Parikh; C. D. Mol; J. A. Tainer, *Structure* **1997**, *5*, 1543–1550.
48. S. H. Wilson, *Mutat. Res.* **1998**, *407*, 203–215.
49. T. Lindahl; R. D. Wood, *Science* **1999**, *286*, 1897–1905.
50. C. E. Piersen; R. Prasad; S. H. Wilson; R. S. Lloyd, *J. Biol. Chem.* **1996**, *271*, 17811–17815.
51. R. Prasad; R. K. Singhal; D. K. Srivastava; J. T. Molina; A. E. Tomkinson; S. H. Wilson, *J. Biol. Chem.* **1996**, *271*, 16000–16007.
52. R. A. Nash; K. W. Caldecott; D. E. Barnes; T. Lindahl, *Biochemistry* **1997**, *36*, 5207–5211.
53. S. Gilfillan; C. Benoist; D. Mathis, *Immunol. Rev.* **1995**, *148*, 201–219.
54. T. Komori; A. Okada; V. Stewart; F. W. Alt, *Science* **1993**, *261*, 1171–1175.
55. S. Gilfillan; A. Dierich; M. Lemeur; C. Benoist; D. Mathis, *Science* **1993**, *261*, 1175–1178.
56. J. D. Fowler; Z. Suo, *Chem. Rev.* **2006**, *106*, 2092–2110.
57. S. A. Nick McElhinny; D. A. Ramsden, *Immunol. Rev.* **2004**, *200*, 156–164.
58. A. F. Moon; M. Garcia-Diaz; V. K. Batra; W. A. Beard; K. Bebenek; T. A. Kunkel; S. H. Wilson; L. C. Pedersen, *DNA Repair* **2007**, *6*, 1709–1725.
59. D. R. Carson; M. F. Christman, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8270–8275.
60. Z. Wang; I. B. Castano; A. De Las Penas; C. Adams; M. F. Christman, *Science* **2000**, *289*, 774–779.
61. M. Oliveros; R. J. Yanez; M. L. Salas; J. Salas; E. Vinuela; L. Blanco, *J. Biol. Chem.* **1997**, *272*, 30899–30910.
62. A. J. Rattray; J. N. Strathern, *Annu. Rev. Genet.* **2003**, *37*, 31–66.
63. S. Prakash; R. E. Johnson; L. Prakash, *Annu. Rev. Biochem.* **2005**, *74*, 317–353.

64. A. R. Lehmann; A. Niimi; T. Ogi; S. Brown; S. Sabbioneda; J. F. Wing; P. L. Kannouche; C. M. Green, *DNA Repair* **2007**, *6*, 891–899.
65. A. R. Lehmann, *Exp. Cell Res.* **2006**, *312*, 2673–2676.
66. A. R. Lehmann, *Mol. Cell* **2006**, *24*, 493–495.
67. P. Kannouche; A. Lehmann, *Meth. Enzymol.* **2006**, *408*, 407–415.
68. S. Prakash; L. Prakash, *Genes Dev.* **2002**, *16*, 1872–1883.
69. T. A. Kunkel; Y. I. Pavlov; K. Bebenek, *DNA Repair* **2003**, *2*, 135–149.
70. A. Vaisman; A. R. Lehmann; R. Woodgate, *Adv. Protein Chem.* **2004**, *69*, 205–228.
71. B. Oberg, *Antiviral Res.* **2006**, *71*, 90–95.
72. H. C. Castro; N. I. Loureiro; M. Pujol-Luz; A. M. Souza; M. G. Albuquerque; D. O. Santos; L. M. Cabral; I. C. Frugulhetti; C. R. Rodrigues, *Curr. Med. Chem.* **2006**, *13*, 313–324.
73. M. Gotte, *Curr. Pharm. Des.* **2006**, *12*, 1867–1877.
74. R. Prasad; W. A. Beard; S. H. Wilson, *J. Biol. Chem.* **1994**, *269*, 18096–18101.
75. Y. Matsumoto; K. Kim, *Science* **1995**, *269*, 699–702.
76. R. Prasad; W. A. Beard; P. R. Strauss; S. H. Wilson, *J. Biol. Chem.* **1998**, *273*, 15263–15270.
77. M. R. Sawaya; H. Pelletier; A. Kumar; S. H. Wilson; J. Kraut, *Science* **1994**, *264*, 1930–1935.
78. M. R. Sawaya; R. Prasad; S. H. Wilson; J. Kraut; H. Pelletier, *Biochemistry* **1997**, *36*, 11205–11215.
79. H. Huang; R. Chopra; G. L. Verdine; S. C. Harrison, *Science* **1998**, *282*, 1669–1675.
80. S. Doublie; S. Tabor; A. M. Long; C. C. Richardson; T. Ellenberger, *Nature* **1998**, *391*, 251–258.
81. D. L. Ollis; P. Brick; R. Hamlin; N. G. Xuong; T. A. Steitz, *Nature* **1985**, *313*, 762–766.
82. T. A. Steitz; S. J. Smerdon; J. Jager; C. M. Joyce, *Science* **1994**, *266*, 2022–2025.
83. W. A. Beard; S. H. Wilson, *Chem. Rev.* **2006**, *106*, 361–382.
84. F. Boudsocq; R. J. Kokoska; B. S. Plosky; A. Vaisman; H. Ling; T. A. Kunkel; W. Yang; R. Woodgate, *J. Biol. Chem.* **2004**, *279*, 32932–32940.
85. J. F. Davies 2nd; R. J. Almassy; Z. Hostomska; R. A. Ferre; Z. Hostomsky, *Cell* **1994**, *76*, 1123–1133.
86. H. Pelletier; M. R. Sawaya; A. Kumar; S. H. Wilson; J. Kraut, *Science* **1994**, *264*, 1891–1903.
87. H. Pelletier; M. R. Sawaya; W. Wolfle; S. H. Wilson; J. Kraut, *Biochemistry* **1996**, *35*, 12742–12761.
88. H. Pelletier; M. R. Sawaya; W. Wolfle; S. H. Wilson; J. Kraut, *Biochemistry* **1996**, *35*, 12762–12777.
89. J. W. Arndt; W. Gong; X. Zhong; A. K. Showalter; J. Liu; C. A. Dunlap; Z. Lin; C. Paxson; M.-D. Tsai; M. K. Chan, *Biochemistry* **2001**, *40*, 5368–5375.
90. V. K. Batra; W. A. Beard; D. D. Shock; J. M. Krahn; L. C. Pedersen; S. H. Wilson, *Structure* **2006**, *14*, 757–766.
91. Y. Li; S. Korolev; G. Waksman, *EMBO J.* **1998**, *17*, 7514–7525.
92. K. A. Johnson, *Annu. Rev. Biochem.* **1993**, *62*, 685–713.
93. S. Doublie; M. R. Sawaya; T. Ellenberger, *Structure* **1999**, *7*, R31–R35.
94. E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191–219.
95. H. Ling; F. Boudsocq; R. Woodgate; W. Yang, *Cell* **2001**, *107*, 91–102.
96. J. H. Wong; K. A. Fiala; Z. Suo; H. Ling, *J. Mol. Biol.* **2008**, *379*, 317–330.
97. O. Rechkoblit; L. Malinina; Y. Cheng; V. Kuryavyi; S. Broyde; N. E. Geacintov; D. J. Patel, *PLoS Biol.* **2006**, *4*, e11.
98. D. T. Nair; R. E. Johnson; L. Prakash; S. Prakash; A. K. Aggarwal, *Nat. Struct. Mol. Biol.* **2006**, *13*, 619–625.
99. T. A. Steitz, *Curr. Opin. Struct. Biol.* **1993**, *3*, 31–38.
100. T. A. Steitz, *J. Biol. Chem.* **1999**, *274*, 17395–17398.
101. M. E. Dahlberg; S. J. Benkovic, *Biochemistry* **1991**, *30*, 4835–4843.
102. B. T. Eger; S. J. Benkovic, *Biochemistry* **1992**, *31*, 9227–9236.
103. R. D. Kuchta; P. Benkovic; S. J. Benkovic, *Biochemistry* **1988**, *27*, 6716–6725.
104. R. D. Kuchta; V. Mizrahi; P. A. Benkovic; K. A. Johnson; S. J. Benkovic, *Biochemistry* **1987**, *26*, 8410–8417.
105. S. S. Patel; I. Wong; K. A. Johnson, *Biochemistry* **1991**, *30*, 511–525.
106. I. Wong; S. S. Patel; K. A. Johnson, *Biochemistry* **1991**, *30*, 526–537.
107. W. A. Beard; D. D. Shock; B. J. Vande Berg; S. H. Wilson, *J. Biol. Chem.* **2002**, *277*, 47393–47398.
108. W. A. Beard; S. H. Wilson, *Structure* **2003**, *11*, 489–496.
109. W. R. McClure; T. M. Jovin, *J. Biol. Chem.* **1975**, *250*, 4073–4080.
110. K. Tanabe; E. W. Bohn; S. H. Wilson, *Biochemistry* **1979**, *18*, 3401–3406.
111. T. S. F. Wang; D. Korn, *Biochemistry* **1982**, *21*, 1597–1608.
112. W. Yang; R. Woodgate, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15591–15598.
113. S. Kumar; M. Bakhtina; M. D. Tsai, *Biochemistry* **2008**, *47*, 7875–7887.
114. P. M. Burgers; F. Eckstein, *J. Biol. Chem.* **1979**, *254*, 6889–6893.
115. J. Liu; M. D. Tsai, *Biochemistry* **2001**, *40*, 9014–9022.
116. D. Herschlag; J. A. Piccirilli; T. R. Cech, *Biochemistry* **1991**, *30*, 4844–4854.
117. B. G. Werneburg; J. Ahn; X. Zhong; R. J. Hondal; V. S. Kraynov; M. D. Tsai, *Biochemistry* **1996**, *35*, 7041–7050.
118. C. A. Sucato; T. G. Upton; B. A. Kashemirov; V. K. Batra; V. Martinek; Y. Xiang; W. A. Beard; L. C. Pedersen; S. H. Wilson; C. E. McKenna; J. Florian; A. Warshel; M. F. Goodman, *Biochemistry* **2007**, *46*, 461–471.
119. C. A. Sucato; T. G. Upton; B. A. Kashemirov; J. Osuna; K. Oertell; W. A. Beard; S. H. Wilson; J. Florian; A. Warshel; C. E. McKenna; M. F. Goodman, *Biochemistry* **2008**, *47*, 870–879.
120. K. H. Jung; A. Marx, *Cell. Mol. Life Sci.* **2005**, *62*, 2080–2091.
121. J. Beckman; K. Kincaid; M. Hocek; T. Spratt; J. Engels; R. Cosstick; R. D. Kuchta, *Biochemistry* **2007**, *46*, 448–460.
122. A. T. Krueger; E. T. Kool, *Curr. Opin. Chem. Biol.* **2007**, *11*, 588–594.
123. J. C. Morales; E. T. Kool, *J. Am. Chem. Soc.* **2000**, *122*, 1001–1007.
124. S. Moran; R. X. F. Ren; S. Rumney; E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056–2057.
125. O. Potapova; C. Chan; A. M. DeLucia; S. A. Helquist; E. T. Kool; N. D. Grindley; C. M. Joyce, *Biochemistry* **2006**, *45*, 890–898.
126. T. W. Kim; J. C. Delaney; J. M. Essigmann; E. T. Kool, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15803–15808.

127. H. R. Lee; S. A. Helquist; E. T. Kool; K. A. Johnson, *J. Biol. Chem.* **2008**, *283*, 14402–14410.
128. M. Garcia-Diaz; K. Bebenek; J. M. Krahn; T. A. Kunkel; L. C. Pedersen, *Nat. Struct. Mol. Biol.* **2005**, *12*, 97–98.
129. A. F. Moon; M. Garcia-Diaz; K. Bebenek; B. J. Davis; X. Zhong; D. A. Ramsden; T. A. Kunkel; L. C. Pedersen, *Nat. Struct. Mol. Biol.* **2007**, *14*, 45–53.
130. J. F. Ruiz; R. Juarez; M. Garcia-Diaz; G. Terrados; A. J. Picher; S. Gonzalez-Barrera; A. R. Fernandez de Henestrosa; L. Blanco, *Nucleic Acids Res.* **2003**, *31*, 4441–4449.
131. M. Astatke; K. Ng; N. D. Grindley; C. M. Joyce, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3402–3407.
132. A. Bonnin; J. M. Lazaro; L. Blanco; M. Salas, *J. Mol. Biol.* **1999**, *290*, 241–251.
133. G. Gao; M. Orlova; M. M. Georgiadis; W. A. Hendrickson; S. P. Goff, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 407–411.
134. C. M. Joyce, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 1619–1622.
135. P. H. Patel; L. A. Loeb, *J. Biol. Chem.* **2000**, *275*, 40266–40272.
136. T. Date; S. Yamamoto; K. Tanihara; Y. Nishimoto; A. Matsukage, *Biochemistry* **1991**, *30*, 5286–5292.
137. K. L. Menge; Z. Hostomsky; B. R. Nodes; G. O. Hudson; S. Rahmati; E. W. Moomaw; R. J. Almassy; Z. Hostomska, *Biochemistry* **1995**, *34*, 15934–15942.
138. M. Bakhtina; M. P. Roettger; S. Kumar; M. D. Tsai, *Biochemistry* **2007**, *46*, 5463–5472.
139. M. P. Roettger; M. Bakhtina; M. D. Tsai, *Biochemistry* **2008**, *47*, 9718–9727.
140. C. A. Dunlap; M. D. Tsai, *Biochemistry* **2002**, *41*, 11226–11235.
141. A. M. DeLucia; N. D. Grindley; C. M. Joyce, *Biochemistry* **2007**, *46*, 10790–10803.
142. V. Purohit; N. D. Grindley; C. M. Joyce, *Biochemistry* **2003**, *42*, 10200–10211.
143. C. Hariharan; L. B. Bloom; S. A. Helquist; E. T. Kool; L. J. Reha-Krantz, *Biochemistry* **2006**, *45*, 2836–2844.
144. X. Zhong; S. S. Patel; B. G. Werneburg; M. D. Tsai, *Biochemistry* **1997**, *36*, 11891–11900.
145. Y. C. Tsai; K. A. Johnson, *Biochemistry* **2006**, *45*, 9675–9687.
146. C. M. Joyce; O. Potapova; A. M. Delucia; X. Huang; V. P. Basu; N. D. Grindley, *Biochemistry* **2008**, *47*, 6103–6116.
147. P. J. Rothwell; V. Mitaksov; G. Waksman, *Mol. Cell* **2005**, *19*, 345–355.
148. G. Stengel; J. P. Gill; P. Sandin; L. M. Wilhelmsson; B. Albinsson; B. Norden; D. Millar, *Biochemistry* **2007**, *46*, 12289–12297.
149. G. Luo; M. Wang; W. H. Konigsberg; X. S. Xie, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12610–12615.
150. P. J. Rothwell; G. Waksman, *J. Biol. Chem.* **2007**, *282*, 28884–28892.
151. M. Bakhtina; S. Lee; Y. Wang; C. Dunlap; B. Lamarche; M. D. Tsai, *Biochemistry* **2005**, *44*, 5177–5187.
152. E. L. Rachofsky; R. Osman; J. B. Ross, *Biochemistry* **2001**, *40*, 946–956.
153. M. Bakhtina; M. P. Roettger; M. D. Tsai, *Biochemistry* **2009**, *48*, 3197–3208.
154. A. C. Brouwer; J. F. Kirsch, *Biochemistry* **1982**, *21*, 1302–1307.
155. B. Gavish; M. M. Werber, *Biochemistry* **1979**, *18*, 1269–1275.
156. L. C. Kurz; E. Weitkamp; C. Frieden, *Biochemistry* **1987**, *26*, 3027–3032.
157. J. Lew; S. S. Taylor; J. A. Adams, *Biochemistry* **1997**, *36*, 6717–6724.
158. J. W. Beckman; Q. Wang; F. P. Guengerich, *J. Biol. Chem.* **2008**, *283*, 36711–36723.
159. R. Radhakrishnan; T. Schlick, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5970–5975.
160. L. Yang; W. A. Beard; S. H. Wilson; S. Broyde; T. Schlick, *J. Mol. Biol.* **2002**, *317*, 651–671.
161. K. A. Johnson, *J. Biol. Chem.* **2008**, *283*, 26297–26301.
162. J. Florian; M. F. Goodman; A. Warshel, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6819–6824.
163. J. A. Cowan, *Chem. Rev.* **1998**, *98*, 1067–1087.
164. W. W. Cleland, *Meth. Enzymol.* **1982**, *87*, 159–179.
165. X. Zhong; S. S. Patel; M.-D. Tsai, *J. Am. Chem. Soc.* **1998**, *120*, 235–236.
166. A. K. Showalter; B. J. Lamarche; M. Bakhtina; M. I. Su; K. H. Tang; M. D. Tsai, *Chem. Rev.* **2006**, *106*, 340–360.
167. A. K. Showalter; M. D. Tsai, *Biochemistry* **2002**, *41*, 10571–10576.
168. C. M. Joyce; S. J. Benkovic, *Biochemistry* **2004**, *43*, 14317–14324.
169. J. M. Krahn; W. A. Beard; S. H. Wilson, *Structure* **2004**, *12*, 1823–1832.
170. Y. Xiang; P. Oelschlaeger; J. Florian; M. F. Goodman; A. Warshel, *Biochemistry* **2006**, *45*, 7036–7048.
171. S. J. Johnson; L. S. Beese, *Cell* **2004**, *116*, 803–816.
172. J. Trincao; R. E. Johnson; W. T. Wolfle; C. R. Escalante; S. Prakash; L. Prakash; A. K. Aggarwal, *Nat. Struct. Mol. Biol.* **2004**, *11*, 457–462.
173. V. K. Batra; W. A. Beard; D. D. Shock; L. C. Pedersen; S. H. Wilson, *Structure* **2005**, *13*, 1225–1233.
174. V. K. Batra; W. A. Beard; D. D. Shock; L. C. Pedersen; S. H. Wilson, *Mol. Cell* **2008**, *30*, 315–324.
175. R. Radhakrishnan; K. Arora; Y. Wang; W. A. Beard; S. H. Wilson; T. Schlick, *Biochemistry* **2006**, *45*, 15142–15156.
176. P. Lin; V. K. Batra; L. C. Pedersen; W. A. Beard; S. H. Wilson; L. G. Pedersen, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5670–5674.
177. A. K. Showalter; M. D. Tsai, *J. Am. Chem. Soc.* **2001**, *123*, 1776–1777.
178. M. P. Roettger; K. A. Fiala; S. Sompalli; Y. Dong; Z. Suo, *Biochemistry* **2004**, *43*, 13827–13838.
179. W. M. Kati; K. A. Johnson; L. F. Jerva; K. S. Anderson, *J. Biol. Chem.* **1992**, *267*, 25988–25997.
180. K. A. Fiala; W. Abdel-Gawad; Z. Suo, *Biochemistry* **2004**, *43*, 6751–6762.
181. E. Fidalgo da Silva; S. S. Mandal; L. J. Reha-Krantz, *J. Biol. Chem.* **2002**, *277*, 40640–40649.
182. D. Starcevic; S. Dalal; J. Sweasy, *Biochemistry* **2005**, *44*, 3775–3784.
183. D. Starcevic; S. Dalal; J. Jaeger; J. B. Sweasy, *J. Biol. Chem.* **2005**, *280*, 28388–28393.
184. K. Arora; T. Schlick, *Biophys. J.* **2004**, *87*, 3088–3099.
185. K. Arora; T. Schlick, *J. Phys. Chem. Ref. Data* **2005**, *109*, 5358–5367.
186. L. Yang; W. A. Beard; S. H. Wilson; S. Broyde; T. Schlick, *Biophys. J.* **2004**, *86*, 3392–3408.
187. K. Arora; W. A. Beard; S. H. Wilson; T. Schlick, *Biochemistry* **2005**, *44*, 13328–13341.
188. R. Radhakrishnan; T. Schlick, *J. Am. Chem. Soc.* **2005**, *127*, 13245–13252.
189. Y. Xiang; M. F. Goodman; W. A. Beard; S. H. Wilson; A. Warshel, *Proteins* **2008**, *70*, 231–247.
190. K.-H. Tang; M. Niebuhr; C.-S. Tung; H.-c. Chan; C.-C. Chou; M.-D. Tsai, *Nucleic Acids Res.* **2008**, *36*, 2948–2957.

191. M. D. Bojin; T. Schlick, *J. Phys. Chem. B* **2007**, *111*, 11244–11252.
192. J. Florian; M. F. Goodman; A. Warshel, *J. Am. Chem. Soc.* **2003**, *125*, 8163–8177.
193. P. Lin; L. C. Pedersen; V. K. Batra; W. A. Beard; S. H. Wilson; L. G. Pedersen, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13294–13299.
194. R. C. Rittenhouse; W. K. Apostoluk; J. H. Miller; T. P. Straatsma, *Proteins* **2003**, *53*, 667–682.
195. R. Radhakrishnan; T. Schlick, *Biochem. Biophys. Res. Commun.* **2006**, *350*, 521–529.
196. I. L. Alberts; Y. Wang; T. Schlick, *J. Am. Chem. Soc.* **2007**, *129*, 11100–11110.
197. R. Venkatramani; R. Radhakrishnan, *Proteins* **2008**, *71*, 1360–1372.
198. A. Abyzov; A. Uzun; P. R. Strauss; V. A. Ilyin, *PLoS Comput. Biol.* **2008**, *4*, e1000066.

Biographical Sketches

Michelle P. Roettger received her B.S. in biochemistry from Otterbein College, Westerville, Ohio, and her M.S. in biochemistry from The Ohio State University (OSU), Columbus, Ohio. In 2008, she received her Ph.D. in biochemistry from OSU, where she was an NIH Chemistry–Biology Interface Program Fellow. Under the supervision of Ming-Daw Tsai, her dissertation research focused on the catalytic mechanism of DNA polymerase β. Currently, Michelle is working in the pharmaceutical industry under Research & Development.

Marina Bakhtina received a B.S. in molecular biology from Novosibirsk State University, Russia, in 1993. She worked as a research associate for 4 years in the Institute of Molecular Biology (SRC VB 'Vector'), and subsequently earned an M.S. in philosophy from Novosibirsk State University in 1999. Working with Ming-Daw Tsai, she received a Ph.D. in chemistry from OSU in 2006. She is currently a postdoctoral researcher in the Tsai group, where her focus is on the kinetic mechanism of DNA polymerase fidelity.

Sandeep Kumar received his M.S. in chemistry from the Indian Institute of Technology, Kanpur, India, in 2003. He received his Ph.D. from OSU in 2008, where he worked on the catalytic mechanism of DNA polymerases under the supervision of Professor Ming-Daw Tsai. He is currently a post-doctoral researcher with Professors Karin Musier-Forsyth and Christopher Hadad at OSU. His areas of interest include molecular modeling and quantum mechanical investigations of the enzyme mechanisms.

Ming-Daw Tsai received a B.S. degree in chemistry from the National Taiwan University in 1972, a Ph.D. in medicinal chemistry from Purdue University in 1978, and joined the faculty of OSU in 1981. He established the Chemistry–Biology Interface Training Program of OSU in 1996 and served as its director through 2003. He has also directed OSU's Office of Research Campus Chemical Instrument Center for 14 years (1995–2007). In 2008 Tsai led the Institute of Biological Chemistry of Academia Sinica, Taiwan as a director. Tsai has published over 220 articles in chemical and biological journals. His honors include an Alfred P. Sloan Fellowship (1983–85), the Camille and Henry Dreyfus Teacher–Scholar Award (1985–90), the Distinguished Scholar Award of OSU (1992), an Elected Fellow of the American Association for the Advancement of Science (1992), the Kimberly Professor of Chemistry at OSU (2003–07), and Distinguished Alumnus Award (Purdue College of Pharmacy). He also serves on the editorial advisory board of *Biochemistry*. His research interests include mechanistic enzymology and structure–function relationship of proteins in DNA damage response signaling and cancer, particularly ankyrin repeat proteins and FHA domains.

8.11   Mechanisms of Enzymatic Glycosyl Transfer

Ran Zhang, Vivian L. Y. Yip, and Stephen G. Withers, University of British Columbia, Vancouver, BC, Canada

8.11.1   Introduction

8.11.1.1   Definitions and Categorizations

Carbohydrates and glycoconjugates are among the most abundant biological molecules occurring in nature and play crucial roles such as the storage of metabolic energy, maintenance of structural integrity, and participation in a range of important biological recognition processes.[1] The controlled biosynthesis and degradation of these structures is therefore vital to the function of all organisms. This task is rendered challenging by the considerable stability of the glycosidic bonds involved. It has been estimated that the half lives for the spontaneous hydrolysis of cellulose and starch are in the range of five million years![2,3] Fortunately, nature provides a solution to this problem in the form of highly proficient enzymes known as glycosidases and glycosyltransferases.

Glycosyltransferases can be broadly defined as enzymes that catalyze the transfer of glycosyl residues from their specific donor to an acceptor molecule.[4] This is essentially a nucleophilic displacement reaction at a saturated carbon, the anomeric center. Most of these glycosyl transfer reactions occur between two oxygen
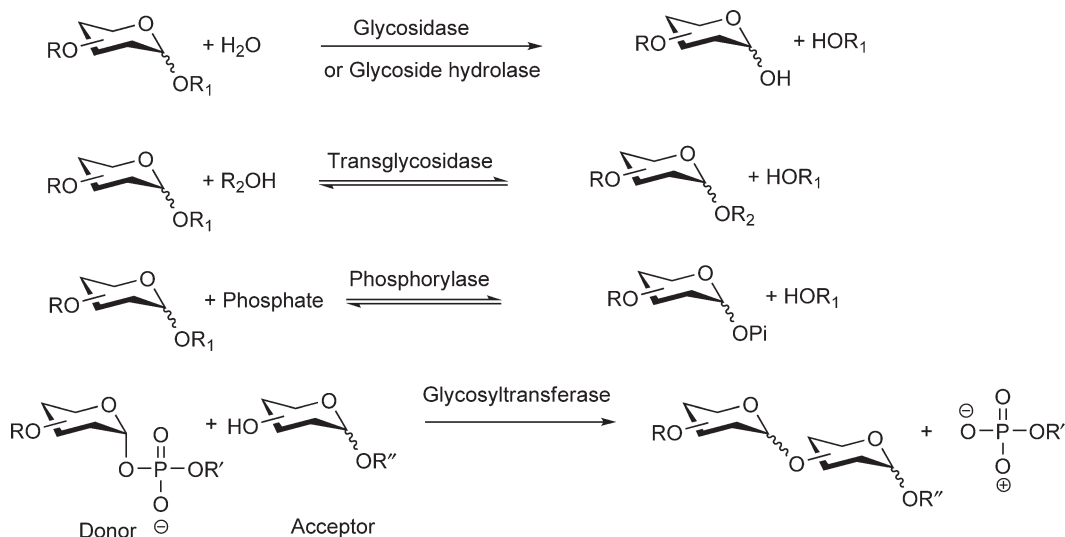
Scheme 1   Reactions catalyzed by different types of glycosyltransferases.

nucleophiles. However, transfer reactions between oxygen and other nucleophiles such as nitrogen, sulfur, and carbon are also well known. From the perspective of metabolism, glycosyltransferases include those enzymes that catalyze glycosyl transfer in both the physiologically catabolic direction and anabolic direction.[5] However, as explained below, the term 'glycosyltransferases' tends to be reserved for those using nucleotide phosphosugars in the synthesis of glycosides.

Enzymes belonging to the former category are divided into three types, depending on their specific acceptors (**Scheme 1**). Glycosidases or glycoside hydrolases (GH) utilize water as the acceptor with the net reaction being hydrolysis. Considering the concentration of water under physiological conditions, this reaction is typically regarded as irreversible though it is not strictly correct thermodynamically.[4] If the acceptor is an alcohol functionality from another molecule such as a sugar or a lipid, the term 'transglycosidase' is used to describe the enzyme involved. Although the net reaction leads to the formation of a new glycosidic bond, transglycosidases are still grouped together with glycosidases since these two types of enzymes share strong similarity in terms of both structures and catalytic machinery. Equilibrium constants for such reactions typically lie close to unity given the chemical similarity of the substrate and product. Well-known examples of transglycosidases are the cyclodextrin glycosyltransferases, which catalyze the formation of cyclic dextrins from starch.[6] The third type of enzymes are the phosphorylases,[7] which also typically have equilibrium constants close to unity. This can easily be driven in either direction by an excess of one reagent. Depending on the direction of the catalyzed reaction, they can either degrade a polysaccharide by transferring the glycosyl moiety to an inorganic phosphate acceptor or synthesize a new glycosidic bond by utilizing a glycosyl phosphate as the donor. For example, glycogen phosphorylase, working in conjunction with glycogen debranching enzyme, breaks glycogen down to glucose-1-phosphate, which is subsequently metabolized.[8] As pointed out in a recent review,[7] there are two types of phosphorylases, which differ in being related structurally and mechanistically to either glycosidases or glycosyltransferases (below). Therefore, only the mechanisms of glycosidases and glycosyltransferases are discussed in the subsequent sections.

Enzymes catalyzing the transfer of glycosyl residue to various acceptors by utilizing activated donor sugars are called glycosyltransferases[9] (**Scheme 1**). In most of the cases, nucleoside diphosphate sugars, such as uridine diphosphate (UDP)-α-D-glucose (**1**), UDP-α-D-galactose, UDP-α-D-N-acetyl glucosamine and GDP-α-D-mannose are the donor substrates (**Figure 1**). This class of molecules was first discovered by Nobel Laureate Luis F. Leloir in 1950[10] and soon found to be the donor substrates for a wide spectrum of glycosyltransferases. Although nucleoside diphosphate sugars are the most common donors, some glycosyltransferases, such as sialyltransferases, use nucleoside monophosphate sugars CMP Neu5Ac (**2**) as the donors[11] (**Figure 1**). Whereas others use lipid phosphate sugars such as dolichol phosphosugars (**3**, **Figure 1**), which are themselves synthesized from nucleoside diphosphate sugars.[12] Presumably, the lipid moieties in these donors allow them
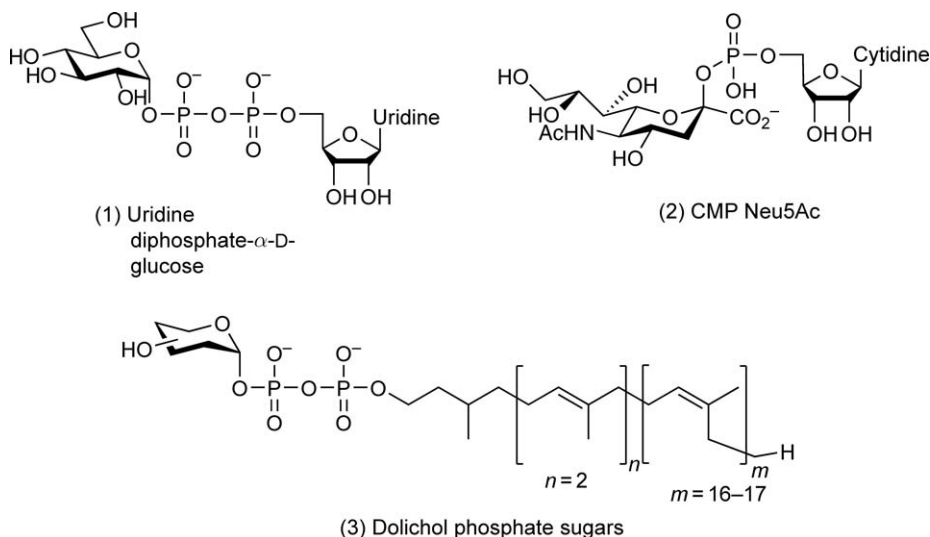
Figure 1 Structures of common donor sugars of glycosyltransferases.

to associate with the membrane lipid bilayer, facilitating their use by membrane-bound glycosyltransferases. The acceptor scope for glycosyltransferases is very broad (though individual enzymes are specific) and can include sugars, lipids, proteins, nucleic acids, and other small molecules such as antibiotics.[13]

All the enzymatic reactions involving glycosyl transfer consist of a nucleophilic substitution at a chiral acetal or ketal center and there are only two possible stereochemical outcomes for this type of reaction, namely inversion or retention of configuration at the anomeric center. Since different chemical mechanisms are associated with each scenario, determination of the stereochemical outcome is of fundamental importance for discussion of catalytic mechanisms for any enzymatic glycosyl transfer. The mechanisms of glycosidases or transglycosidases are well studied and many reviews have been published.[4,5,14–21] However, less mechanistic information is available for glycosyltransferases and phosphorylases.[7,22] Sometimes, 'α-glycosidase' or 'β-glycosidase' is used in the literature to describe a specific glycosidase and this indicates that the leaving group at the anomeric center of the donor substrate has α or β configuration, respectively.

8.11.1.2 Classification of Glycoside Hydrolases and Glycosyltransferases

A vast diversity of glycan structures exists in nature, with each glycosidic bond requiring a specific glycosyl-transferase or glycosidase for its synthesis and breakdown, as discussed in the next section. A systematic classification is essential for this large number of enzymes. The simplest classification method is based upon the substrate specificity of a specific enzyme, as recommended by the International Union of Biochemistry and Molecular Biology (IUBMB). For glycosidases, EC 3.2.1.x is used to describe them. The first three digits 3.2.1 indicate that this enzyme hydrolyzes an O-glycosidic linkage and the last digit can be variable, depending on the specific structure of the glycone. This system has its advantages as it avoids the ambiguities of the many trivial names used in the literature and provides a unified nomenclature. However, there are shortcomings. Many glycosidases are multifunctional enzymes and act on different substrates, thereby leading to difficulties for the IUBMB system. In addition, it has been found that many structurally unrelated glycosidases have identical EC numbers simply because they have similar substrate specificity while many other glycosidases, which are structurally and mechanistically related, have different EC numbers. The IUBMB classification is therefore a useful substrate-focused classification, but fails to reflect the structural and mechanistic features of glycosidases.

A new classification of glycosidases was introduced by Bernard Henrissat in 1991, based on the premise that the primary sequences of proteins dictate their three-dimensional structures and hence their catalytic

mechanisms.[23,24] Glycosidases that share significant similarities in their primary sequences were grouped together and a family number was assigned for each such group. At the time of writing, there are 112 families and the number of enzymes entered into the CAZy database is still increasing rapidly. Valuable information regarding these carbohydrate-active enzyme families, including links to the primary sequences, three-dimensional structures as well as mechanistic information, is available at the following website: http://www.cazy.org/index.html.

The classification of glycosidases based upon sequence similarity is of great value. First, it has greatly simplified research in this field since enzymes within a single family share structural and mechanistic similarities allowing generalizations to be made based on structural and mechanistic studies on representative members of a family. Withers and co-workers were the first to test this concept by examining the stereo-chemical outcome for ten β-1, 4-glucanases, and xylanases, which were sampled from five different families. Indeed, enzymes from the same family were found to exhibit the same stereoselectivity, which strongly suggests that the same catalytic mechanism is employed.[25] Over the past decade with the availability of three-dimensional structures of many glycosidases, it was found that the general fold and active site topology are shared by members from the same family, validating the generality of this classification.[26,27] Therefore, once the catalytic residues of one glycosidase have been identified, the corresponding residues for other members of the same family can be easily and accurately predicted by sequence alignment.[28,29] Further, when structural information is available on one family member, homology models can be constructed between members from the same family and useful insights can be gleaned.[30] Second, it has been found that enzymes from the same family can often act on different substrates. For example, glycosidases from GH1 can hydrolyze β-glucosides, β-galactosides, β-mannosides, β-glucosinolates, and 6-phospho-β-glucosides. Similarly, GH 13, the largest glycosidase family, also possesses the feature of 'multi-specificity'. Members from this family can not only hydrolyze a range of maltooligosaccharides, but also transfer glycosyl residues to other sugar acceptors to make new glycosidic linkages. Remarkably, given the wide range of substrates on which each family can act, they share similar catalytic machinery and three-dimensional structures, since these features are more conserved during evolution. This phenomenon suggests the existence of a possible common ancestor and that acquisition of new substrate specificities is a common evolutionary trend for each family of glycosidases, probably due to the pressure of the emergence of new glycan structures.

As the number of solved crystal structures of glycosidases has grown, it has been found that some families having little sequence similarity are nonetheless structurally related. As three-dimensional structures are more conserved than primary sequences in the evolutionary process, the concept of 'clans' was created to accom-modate different families having similar structures.[27] Currently, there are 14 clans, each of which contains at least two families. The largest clan is GH A, which currently includes families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86. All these families have a $(\beta/\alpha)_8$ barrel fold with the catalytic general acid/base residue and nucleophile residue located on strands 4 and 7, respectively. They are therefore sometimes called the 4/7 superfamily.[31] A wide variety of folds are found in the other clans, as illustrated in **Figure 2**.

A new concept of the sequence-based classification was introduced very recently by Henrissat in the form of a 'subfamily' categorization,[32] which provides a useful tool for further correlating primary sequences of glycosidases with their substrate specificities. This is especially important for large glycosidase families where different enzymatic activities and substrate preferences within the same family are found. GH 13 (α-amylase family), being the largest sequence-based family (with 4582 members at the time of writing), is well known to have a wide spectrum of substrate specificities and enzymatic activities including α-(1,4)-glycoside hydrolases, α-(1,6)-glycoside hydrolases, transglycosidases, and isomerases. Therefore, this family was selected for further division into subfamilies by applying sensitive computational methods incorporating phylogenetic information. The majority of the GH 13 sequences, as a result, could reliably be divided into 35 subfamilies. Interestingly, all the subfamilies are found to be either monospecific or composed of enzymes with highly related specificities. In fact, 26 out of the 35 subfamilies are shown to possess a single EC number, clearly demonstrating the power of this 'subfamily' categorization in predicting substrate specificity within GH 13 enzymes. This methodology will be extended to other large glycosidase families in the future, which should greatly improve the predictive ability of the CAZy classification.

Similar to the sequence-based classification for GH, Campbell *et al.*[9] proposed the grouping of glycosyl-transferases into different families based on their sequence similarity.[33] To date, there are 91 such families and
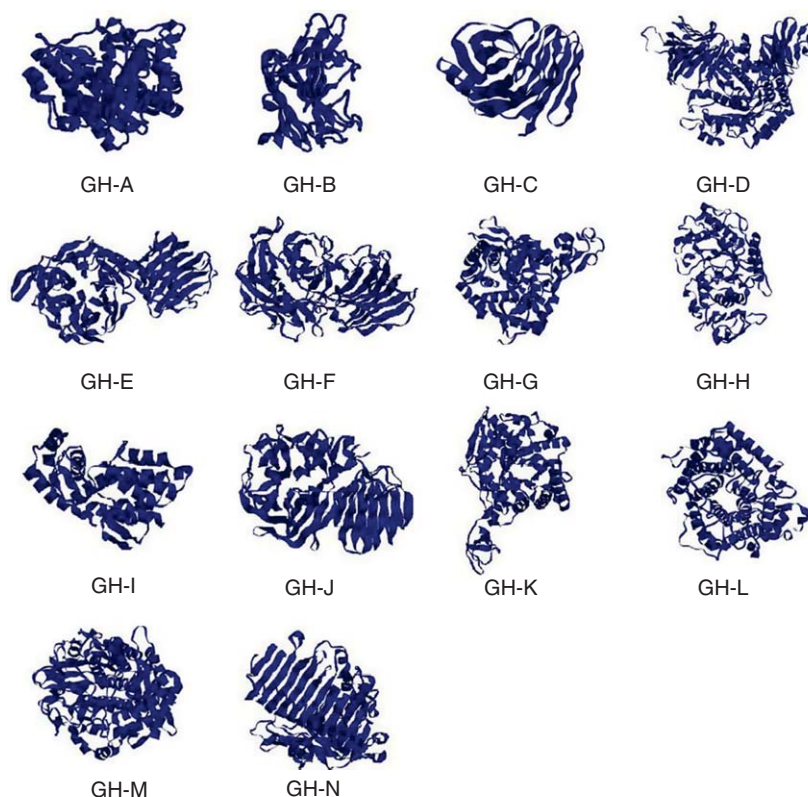
Figure 2 Representative structures for all the clans of glycoside hydrolases (GH). GH-A: Family 1, β-glucosidase B from *Paenibacillus polymyxa*(PDB: 2O9P); GH-B: Family 16, endo-β-1,3-glucanase from alkaliphilic *Nocardiopsis* sp. *F96*(PDB: 2HYK); GH-C: Family 11, xylanase A (PDB: XlnA) from *Bacillus circulans* (PDB: 1BCX); GH-D: Family 31, α-xylosidase (YicI) from *E. coli K12* (PDB: 1XSI); GH-E: Family 33, *trans*-sialidase from *T. cruzi* (PDB: 1MR5); GH-F: Family 43, β-1,4-xylosidase from *B. halodurans C-125* (PDB: 1YRZ); GH-G: Family 37, trehalase from *Escherichia coli* (PDB: 2JF4); GH-H: Family 13, α-amylase from human pancreas (PDB: 1CPU); GH-I: Family 46, chitosanase from *Streptomyces* sp. *N174* (PDB: 1CHK); GH-J: Family 32, β-fructosidase from *Thermotoga maritima* (PDB: 1UYP); GH-K: Family 18, chitinase B from *Serratia marcescens* (PDB: 1E15); GH-L: Family 15, glucoamylase from *Saccharomycopsis fibuligera* (PDB: 1AYX); GH-M: Family 48, cellulase CEL48F from *Clostridium cellulolyticum* (PDB: 1F9D); GH-N: Family 28, exopolygalacturonase from *Yersinia enterocolitica* (PDB: 2UVE).

the relevant information is listed in the CAZy website mentioned earlier. A detailed comprehensive review of mechanistic and structural studies of glycosyltransferases was recently published by Lairson *et al.*[22]

Clans, families, and subfamilies are therefore different concepts in the sequence-based classification of glycosidases/glycosyltransferases and reflect different levels of relatedness between proteins. Clans reflect distant evolutionary relationships but lack functional predictive power while subfamilies have maximum predictive ability such as general fold, catalytic mechanism, and substrate specificity. Therefore, the selection and application of these terms depend on the intended purpose, either functional prediction or evolutionary studies. This classification system, with its hierarchy of levels, has already proven extremely important in research on carbohydrate-active enzymes and will become increasingly important in the post-genome era by providing a reliable tool for the annotation of open reading frames during genome sequencing.

8.11.1.3 Structures of Glycosidases and Glycosyltransferases

Structural characterization of glycosidases has a long history, with the very first crystal structure of an enzyme being that of hen's egg white lysozyme. Of the current 112 families of glycosidases, more than 60 families already have at least one representative structure determined. An important feature emerging from these structural studies is the occurrence of widely diversified structural scaffolds, despite the fact that all these

enzymes catalyze essentially the same chemical reaction, that is, hydrolysis of an acetal.[34,35] Common folds for these enzymes include $(\beta/\alpha)_8$, $(\alpha/\alpha)_6$, fivefold-β-propeller, sixfold-β-propeller, and β-jelly roll but a large number of others exist. **Figure 2** shows representatives of each of the current clan members by way of illustration. This diversity of structures reflects the various evolutionary solutions to the chemical problem of creating a functional enzyme active site that is specific for a particular substrate. In spite of the wide array of folds of glycosidases, however, it has been found that their active site topologies fall into only three categories, which strongly correlate with the action pattern of the corresponding enzyme.[15] Enzymes having pocket-shaped active sites usually recognize and cleave the nonreducing ends of saccharide structures in an 'exo' mode of action. Well-known examples include monosaccharidases and exopolysaccharidases, such as glucoamylases, β-amylases, and sialidases, as apparent from the following sections. The second type contains a cleft-shaped active site. These 'open' structures allow access to their polymeric substrate in a relatively random manner, thus their action mode is 'endo' and the internal linkages of the polysaccharide chain are cleaved. Enzymes that degrade polymeric substrates such as α-amylases, endocellulases, and chitinases belong to this category. The last type of active site topology is tunnel-like. This topology probably evolved from the 'clefts' by 'closing' them with long loops. The advantage of a 'tunnel-like' active site probably lies in the increase of processivity since polymeric substrates thread through the tunnel and the reaction product remains bound close to the enzyme active site, ready for the next cleavage reaction. Cellobiohydrolases were the first group of enzymes that were shown to have this topology of active sites.[36]

Compared to glycosidases, structural analysis of glycosyltransferases has proceeded relatively slowly during the same period of time. The first crystal structure of this important class of enzymes, which appeared in 1994, was that of a DNA-modifying β-glucosyltransferase (BGT) from bacteriophage T4.[37] *Bacillus subtilis* glycosyl transferase SpsA was the second crystal structure reported five years later.[38] Ever since, the pace of glycosyltransferase structural elucidation has greatly accelerated and currently there are 30 families out of the 91 families with at least one representative structure determined. In contrast to the widely diversified structures of glycosidases, only two general folds have been found for the nucleoside phosphosugar-dependent glycosyltransferases characterized to date, namely, GT-A and GT-B[39] (**Figure 3**). Both contain $\alpha/\beta/\alpha$ Rossmann folds that are typical of nucleotide-binding proteins. The GT-A fold was first described for *B. subtilis* glycosyl transferase SpsA and consists of two closely associated $\beta/\alpha/\beta$ domains joined together. Therefore, GT-A folds are regarded as a single domain structure, though distinct donor and acceptor binding sites are present. Another signature of the GT-A fold is the existence of a DxD structural motif, in which the two carboxylic acid residues interact with the phosphate groups of the nucleotide sugar via a divalent metal cation. However, glycosyltransferases having a GT-A fold but lacking a DxD motif are also known. In fact, DxD sequences are extremely common, thus they have little useful predictive power. The first reported glycosyltransferase having a GT-B fold was a DNA-modifying BGT from bacteriophage T4. GT-B folds are composed of two separate Rossmann fold domains, which are loosely connected by a linker region. The catalytic sites are usually located at the interface of these two domains. The existence of such a limited number of folds for nucleotide sugar-dependent glycosyltransferases merits consideration. One possibility for such a limited set of folds is evolutionary. Glycosyltransferases of both GT-A and GT-B folds have been found within the primitive Archae. Therefore it is possible that all glycosyltransferases from higher organisms evolved from their counterparts in the lower organisms. Alternatively, this structural conservation could be simply a reflection of the limitations imposed by the need to bind a nucleotide phosphosugar within a Rossmann fold.

GT-A   GT-B
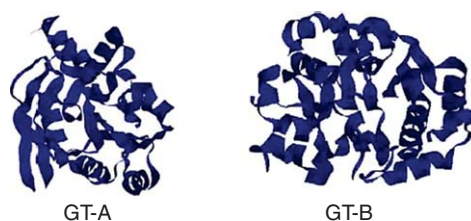
Figure 3   Representative structures for GT-A and GT-B folds. GT-A: Family 2, SpsA enzyme from *Bacillus subtilis* (PDB: 1QG8); GT-B: Family 8, α-galactosyltransferase LgtC from *Neisseria meningitidis 126E* (PDB: 1G9R).

A new glycosyltransferase fold, GT-C, was recently proposed from detailed sequence and structure comparison of the glycosyltransferases listed in the CAZy website.[40] Proteins having this general fold were predicted to have multiple hydrophobic transmembrane domains inserted into the ER or the plasma membrane. Interestingly, the majority of the glycosyltransferases classified into GT-C superfamily use lipid phosphate sugars as the donor substrate, consistent with the existence of hydrophobic domains in these enzymes. Importantly, the first crystal structure of a member of the proposed superfamily, GT-C, appeared in 2008, this being the soluble C-terminal domain protein of the oligosaccharyltransferase STT3 from *Pyrococcus furiosus*.[41] It is indeed a novel structure unrelated to GT-A or GT-B, with one α-helical domain flanked by three β-sheet domains. However, the failure to detect activity from this truncated protein undermines its value and has led to the proposal that the deleted transmembrane region is necessary for the enzymatic activity. Further, this structure revealed that the common predicted structural element in the GT-C fold was the membrane-bound protein and not the catalytic domain – greatly diminishing the predictive value of the 'GT-C' classification. Recently, the structure of another lipid phosphate sugar-dependent glycosyltransferase was solved, this being the long-sought peptidoglycan glycosyltransferase from GT 51.[42,43] It resembled neither GT-A, GT-B, nor the 'GT-C' structure, but rather showed closest structural similarity to lysozyme, the enzyme involved in degrading peptidoglycan. Although it is still too early to make generalizations on the structures of lipid phosphate sugar-utilizing glycosyltransferases, it is evident that a Rossmann fold is not a necessary structural building block of this group of enzymes. Given the fact that no nucleotide is present within the donor substrates, this is not too surprising. Indeed there may well prove to be a wide range of structures adopted by this group of enzymes that has been freed from the tyranny of the Rossmann fold.

8.11.2 Reaction Mechanisms of Glycoside Hydrolases

The two canonical mechanisms of enzymatic glycoside hydrolysis were first proposed by Koshland in 1953.[44] Since then, considerable amounts of structural and mechanistic data have been accumulated in support of the proposed mechanisms, both of which involve nucleophilic displacement steps. The reactions result in either net (1) inversion, or (2) retention of the substrate anomeric configuration.[4,17,18,44] The following section provides an account of these two basic mechanistic groups of enzymes.

8.11.2.1 Retaining Glycosidases

The majority of the enzymes that catalyze the hydrolysis of glycosidic linkages with net retention of substrate anomeric configuration possess two key amino acid residues: (1) a catalytic nucleophile, and (2) a general acid/base catalyst.[4,17–19] Although there are exceptions (discussed in Section 8.11.3), both catalytic residues are usually carboxylic acids, either Asp or Glu.[4,18,19] The two carboxylic acid residues are separated by a distance of approximately 5 Å.[4,17–21,45] The catalytic nucleophile is suitably located for in-line attack on the anomeric center. The second amino acid is in close proximity to the glycosidic oxygen and donates a proton to this oxygen during the nucleophilic displacement step, thereby providing general acid catalysis to leaving group departure. In this glycosylation step, a covalent glycosyl–enzyme intermediate is formed via an oxocarbenium ion-like transition state. The general acid catalyst is now deprotonated and acts as a general base in the second step, by activating a water molecule for nucleophilic attack at the anomeric center of the glycosyl–enzyme intermediate. This second step is called the deglycosylation step and also proceeds via an oxocarbenium ion-like transition state. Thus, two nucleophilic displacement steps are required for substrate hydrolysis, commonly known as the double-displacement mechanism (**Scheme 2**).[4,17–21]

8.11.2.2 Inverting Glycosidases

The inverting glycosidases operate slightly differently from the retaining glycosidases. The key catalytic residues are usually a pair of carboxylic acids, but they are typically separated by a distance of 6–12 Å,
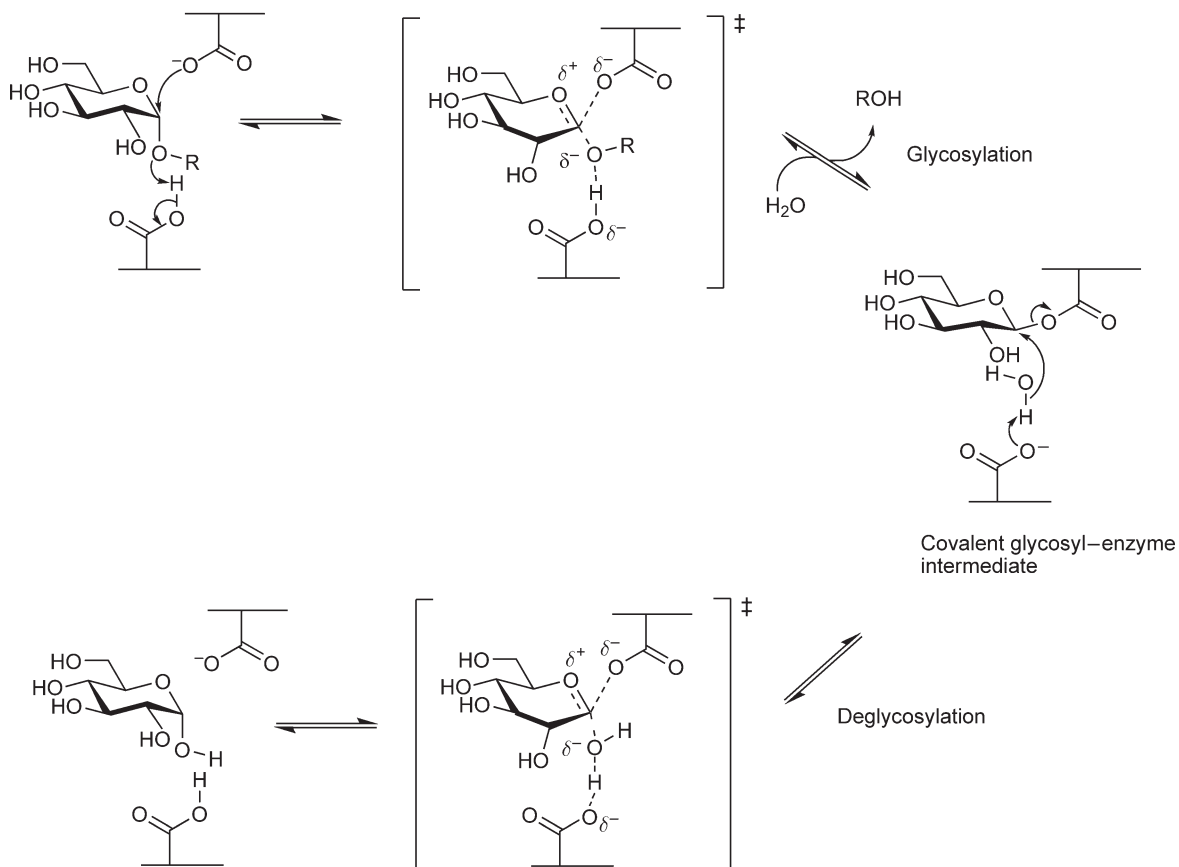
Scheme 2 General mechanism utilized by retaining glycosidases.

somewhat larger than that of the retaining glycosidases.[4,18–21] This large distance is necessary for the accommodation of a water molecule as well as the substrate at the active site and is one of the fundamental differences between the two general glycosidase mechanisms. One carboxylic acid acts as the general base and activates a water molecule for nucleophilic attack at the substrate anomeric center (**Scheme 3**). At the same time, the second carboxylic acid functionality facilitates the departure of the leaving group via general acid catalysis. The inverting glycosidases catalyze the hydrolysis reaction in a single step via direct displacement of the aglycone, and the transition state has oxocarbenium ion-like characteristics similar to that described for the retaining glycosidase mechanism.[4,18–21]
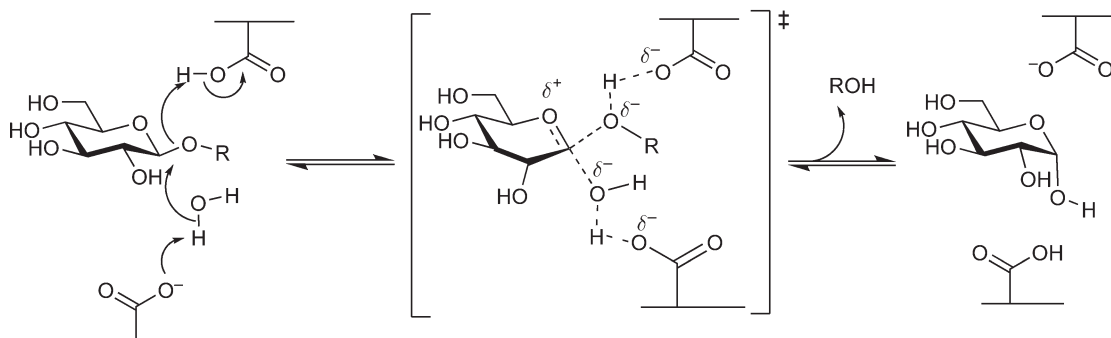
Scheme 3 General mechanism utilized by inverting glycosidases.

8.11.2.3   Mechanistic Anomalies

The inverting and retaining glycosidase mechanisms are used by the majority of glycosidases. Until the mechanistic elucidation of GH4 enzymes in 2004,[46] the major exceptions to the classical nucleophilic displacement mechanisms were those of the chitinases and *N*-acetyl-*β*-hexosaminidases belonging to GH18, 20, 84, and 85, and hyaluronidases in GH 56, which utilize the substrate acetamido carbonyl oxygen as an intramolecular nucleophile rather than a carboxylic amino acid residue (discussed in Section 8.11.3.3).[47–51] The other mechanistic variation is observed in the sialidases and *trans*-sialidases, which utilize a Tyr residue rather than the usual Asp or Glu as the catalytic nucleophile (Section 8.11.3.4).[52–54] One further known variation is that of the GH1 myrosinases, which catalyze the hydrolysis of thioglycoside substrates with an activated anionic leaving group.[55–59] Cleavage of the C1–S1 linkage does not require general acid catalysis and the catalytic acid/base residue for the myrosinases is instead replaced by a neutral glutamine residue.[55,56,60] Exogenous ascorbic acid functions as the catalytic base, activating a water molecule for nucleophilic attack at the anomeric center of the glycosyl–enzyme intermediate.[55,56,60] Common to all these glycosidases is the nucleophilic displacement step via an oxocarbenium ion-like transition state.

8.11.2.4   Strategies for the Identification of Catalytic Residues

The catalytic residues of a large number of glycoside hydrolase families have been identified by various means, as described below. These identities have been summarized in a recent review.[61]

8.11.2.4.1   *Identification of the catalytic nucleophile in retaining glycosidases*

The most reliable method for identifying the nucleophile involves the trapping of a covalent glycosyl–enzyme intermediate. If the relative rate constants for the glycosylation and deglycosylation oblige, nucleophile identification can sometimes be achieved using radioactively-labeled substrates.[62] In these cases, the enzyme reaction is rapidly quenched and subjected to proteolysis. The radioactive peptide can then be isolated and sequenced for catalytic nucleophile identification. A more reliable method has involved the use of a class of specifically fluorinated substrate derivatives[63–68] that form stable glycosyl–enzyme intermediates. The introduction of an electronegative fluorine atom close to the anomeric center, where positive charge is generated at the oxocarbenium ion-like transition state, slows down both the formation and the hydrolysis of the intermediate through inductive effects and via the removal of key transition state-stabilizing interactions that are normally present at the corresponding position. Meanwhile, incorporation of a reactive leaving group speeds up the glycosylation step, allowing the glycosyl–enzyme intermediate to accumulate and be trapped. The first of these reagents were the 2,4-dinitrophenyl 2-deoxy-2-fluoro-glycosides used to trap *Alcaligenes faecalis* *β*-glucosidase.[64] This strategy is now used extensively, and has found most use for *β*-retaining glycosidases.[63,65,69–76] Unfortunately, these 2-fluoro sugars have not proved useful for labeling *α*-retaining glycosidases, but serve as slow substrates instead. Two other classes of reagents have therefore been developed for *α*-retaining glycosidases: 2-deoxy-2,2-difluoro-glycosides[77] and 5-fluoroglycosyl fluorides.[78] Examples of use of 2-deoxy-2,2-difluoro-glycosides and 5-fluoroglycosyl fluorides to trap covalent intermediates include enzymes from GH13,[78,79] GH27,[80] GH29,[81] GH31,[82] and GH38.[83,84] Extension of this strategy to sialidases and *trans*-sialidases, but in those cases using 3-fluoro sugars, allowed the first identification of a tyrosine as the catalytic nucleophile in a glycosidase.[54] Thus, this strategy has been shown to be versatile among different types of glycosidases. Indeed, it has recently been adapted for proteomics studies. Identification of the set of retaining *β*-glycanases in a proteome was achieved using a biotin-tagged aryl 2-deoxy-2-fluoro xylobioside inactivator to selectively target these enzymes in the *Cellulomonas fimi* proteome and permit their identification and quantification.[85,86]

8.11.2.4.2   *General acid/base catalyst*

Unlike the catalytic nucleophile, the general acid/base catalyst cannot be identified completely unambiguously by labeling strategies. Current approaches generally rely on careful inspection of the active site of available crystal structures, primary sequence alignment, site-directed mutagenesis of conserved Asp and Glu residues

and detailed mechanistic evaluation of the resultant mutants. These methods are applicable to both the inverting and retaining glycosidases. Some successful examples are highlighted below.

8.11.2.5   More Mechanistic Characterizations of Glycosidases

8.11.2.5.1   Stereochemical outcome determination

Determination of the stereochemical outcome is a key starting point in the mechanistic elucidation of any glycosidase. The stereochemistry of the hydrolysis product was originally determined by optical rotation, but this method has not proved reliable and new approaches have been adopted.[87,88] The resolution of different anomers by HPLC purification has been reported in some cases,[89,90] but the method involving ^{1}H NMR is more general, thus more widely used.[46,91–94] Different anomers can be distinguished by the differences in ^{1}H-NMR chemical shifts and vicinal proton coupling patterns.

8.11.2.5.2   Hammett relationship

Aryl glycosides have been used extensively in the mechanistic characterization of glycosidases.[95–99] The popularity of aryl glycosides lies in the fact that the rate of hydrolysis is readily measured by directly monitoring changes in UV–vis absorbance upon release of the chromophoric phenol leaving group. Furthermore, since favorable π-stacking interactions with aromatic residues are often found in the -1 and $+1$ subsites of glycosidases, these synthetic substrates bind to the active site quite well. By using different substituents on the aromatic ring, the ability of the leaving group of the substrates can be modified, so that the rate of the glycosylation step can be manipulated. Linear free energy relationships can then be measured. The values of log k_{cat} or log (k_{cat}/K_M) are plotted against the pK_a values of the phenol leaving groups. The k_{cat}/K_M parameter reflects the first irreversible step of the reaction mechanism, typically the glycosylation step, and the k_{cat} parameter provides information on the rate-limiting step. A negative slope is expected if cleavage of the C1–O1 linkage is rate limiting and if substantial negative charge development occurs at O1. The value of the slope reflects the sensitivity of the rate constant to the changes in the ability of the leaving group. A large negative value would indicate substantial charge development at the glycosidic oxygen, thus also likely at the anomeric center. However, as the value of the slope approaches zero, there is either little bond cleavage at the transition state or substantial general acid catalysis at the anomeric oxygen. A flat linear free energy relationship would correspond to either a lack of charge development at O1 or a nonrate limiting cleavage step.

Biphasic plots of log k_{cat} versus pK_a have been determined.[95,100–106] This biphasic plot is characteristic of retaining glycosidases.[95,100–106] The concave downward plots signify a change in the rate-limiting steps, indicating that glycosylation is rate limiting in some cases, while deglycosylation is rate determining in others. In such cases, for substrates with good leaving groups, the reaction rate is independent of the aglycone structure, suggesting that deglycosylation is rate limiting. The leaving group-dependent portion of the plot corresponds to substrates with poor leaving group ability, with the glycosylation step being rate limiting for these substrates. In many cases, the plots of log (k_{cat}/K_M) vs pK_a also display concave downward correlations.[95,100–103] Because the k_{cat}/K_M parameter is the second-order rate constant and typically reflects the glycosylation step (first irreversible step for retaining glycosidases), such biphasic behavior is not well-understood. This phenomenon has been attributed either to substrate 'on' rates becoming rate limiting for the 'best' substrates or to low substrate 'off' rates for tightly bound substrates, making binding the first irreversible step, or possibly due to the higher affinity for phenol leaving groups with low pK_a values, thus rendering the glycosylation step reversible.[100–102,106]

Inverting glycosidases have only one chemical step, thus cleavage of the C1–O1 linkage is more likely to be rate limiting, and alteration of the leaving group ability is more likely to affect the reaction rate, except in those cases where there is extensive proton donation to the glycosidic oxygen. This was indeed shown to be the case for the inverting GH43 β-xylosidase (XynB3) from *Geobacillus stearothermophilus* T-6.[107] The wild-type enzyme showed relatively minor dependence on the ability of the leaving group, while the XynB3 mutant, which cannot provide general acid catalysis to the anomeric center, showed significant rate dependence upon substrate reactivity.[108] However, it is still quite possible that binding steps or even conformation changes could be rate limiting.

8.11.2.5.3   *Kinetic isotope effects*

Kinetic isotope effects (KIEs) for glycosidases can be easily measured using aryl glycoside substrates. These measurements have been used to identify rate-limiting steps and to provide information regarding transition state structures. Primary KIEs are observed if the bond with the isotopic substitution is cleaved during the rate-limiting step. For glycosidases, since the carbon–oxygen bond is cleaved, this requires accurate measurement of heavy-atom isotope effects. This has been demonstrated in a few studies with a β-galactosidase, β-glucosidase, and lysozyme.[109–119] Recently, a new method was developed, which allows for measurement of multiple ^{13}C KIE via NMR based on natural isotope abundance, and was used to study two retaining glycosidases: a yeast α-glucosidase and an almond β-glucosidase.[120] The results indicated that the α- and β-glucosidase employ a D_N*A_N and a A_ND_N mechanism, respectively. Unfortunately, this approach requires huge quantities of substrates. The measurement of secondary α-deuterium KIEs for substrates bearing a deuterium substitution at the anomeric center is also routinely used.[100–102,121] In cases where the anomeric carbon undergoes a change from sp^3 to sp^2 hybridization from the ground state to the oxocarbenium ion-like transition state, a secondary α-deuterium kinetic isotope effect can be observed and these are generally within the range of $k_H/k_D = 1.1$–1.3. A large value is indicative of substantial sp^2 hybridization at the transition state and thus substantial oxocarbenium ion-like character at the transition state. These α-secondary KIEs at the anomeric center have been used broadly in numerous mechanistic studies, including that of *Agrobacterium* sp. β-glucosidase[95] and *Cellulomonas fimi* β-(1,4) exoglycanase.[105] Inverting glycosidases have also been probed via KIEs, an example being an α-secondary KIE that was measured for the GH43 inverting xylosidase from *Bacillus pumilus*. This result suggests that the direct displacement mechanism also proceed via an oxocarbenium ion-like transition state.[122]

Solvent deuterium KIEs have also been measured with yeast α-glucosidase.[123] However, the interpretation of solvent KIEs is often not straightforward for enzymatic systems and thus is not consistently investigated in mechanistic studies.

8.11.3   Examples of Mechanistic Studies of Some Representative Glycosidases

The nucleophilic displacement mechanisms utilized by retaining and inverting glycosidases have been examined in detail. Thus, the interested reader is referred to a number of excellent reviews on these two mechanisms.[17–20,124] The remainder of this section focuses on new mechanisms discovered for retaining glycosidases within the last five years plus studies on a representative group of inverting glycosidases.

8.11.3.1   Family 4 Glycosidases

Family 4 glycosidases were found to display several unusual properties relative to other glycosidase families.[46,90,125–146] First, all enzymes in this family require an NAD$^+$ cofactor and a divalent metal ion (such as Mn^{2+} or Ni^{2+}) for catalytic activity.[46,90,125–146] Second, this family includes both α- and β-glucosidases, and some enzymes are able to hydrolyze both α- and β-glycosides.[130–139] Furthermore, the substrate specificities of these enzymes also vary widely. Some GH4 members require their hexose substrates to contain a phosphate group at the C6 position, while others prefer nonphosphorylated substrates.[46,90,129–139,147] In many cases, reducing conditions are also necessary for enzyme activity. Based on these facts, it seems unlikely that these enzymes would use either the retaining or inverting glycosidase mechanism, and indeed mechanistic and structural studies revealed a completely new mechanism not involving nucleophilic displacement steps.

While a number of X-ray crystal structures of GH4 enzymes have been solved, only two of these enzymes have been subjected to detailed mechanistic investigation.[90,125,135] The two enzymes are BglT, a 6-phospho-β-glucosidase from *Thermotoga maritima*, and GlvA, a 6-phospho-α-glucosidase from *B. subtilis*.[46,90,93,125–127] X-ray crystallographic data revealed that these enzymes are structurally similar to dehydrogenases and dehydratases,[90,135,148] with the NAD$^+$ cofactor and the divalent metal located at the enzyme active site.[46,90,125,135] C4 of the nicotinamide ring is located approximately 4 Å away from C3 of the substrate, in an ideal position for redox chemistry at that position.[46,90,125,135] The metal ion is chelated to the C2 and C3 hydroxyl groups.[46,90,125] These data along with the observation of solvent deuterium incorporation into the substrate C2 position during catalysis supported the proposed elimination mechanism (**Scheme 4**), which involves: (1) C3 hydride
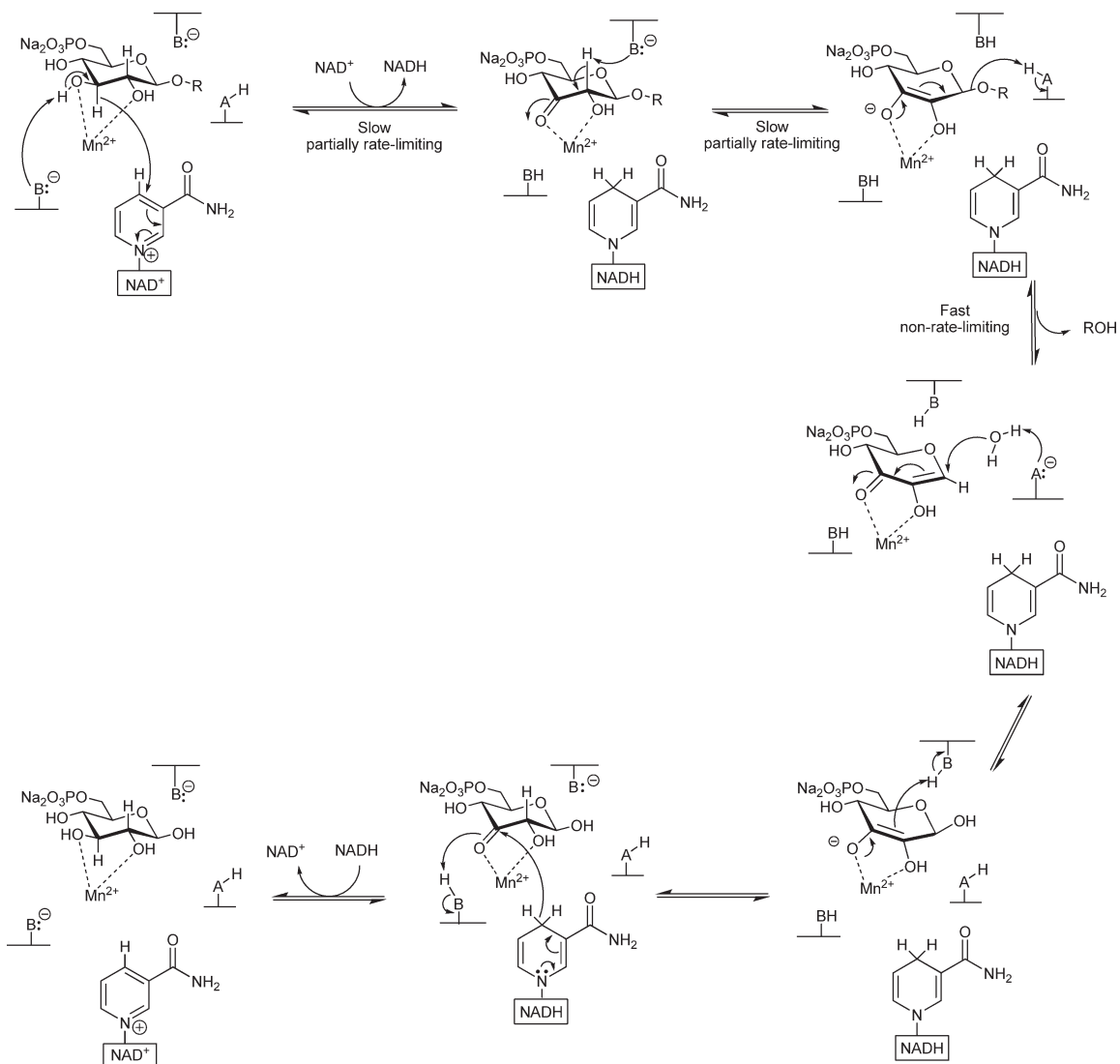
Scheme 4 Proposed E1$_{cb}$ mechanism of the GH4 enzyme BglT.[46]

abstraction by the NAD$^+$ cofactor, thus oxidation of the C3 hydroxyl; (2) abstraction of the C2 proton via general base catalysis; (3) α,β-elimination of the aglycone; (4) 1,4-Michael-like addition of water to the α,β-unsaturated intermediate; (5) reprotonation at C2; and finally (6) reduction of the C3 carbonyl via oxidation of the 'on-board' NADH cofactor.[46,90,125]

The mechanistic probes utilized in the elucidation of the mechanisms of BglT and GlvA include KIE measurements, linear free energy relationships, and the investigation of thioglycosides as potential substrates, as follows.[46,126,127] The mechanistic data accumulated for GlvA and BglT are very similar and thus can be considered as representative of other GH4 enzymes. Both enzymes were found to be retaining glycosidases by NMR analysis of the reaction products.[46,90,93] It was also determined that the reduction of the nicotinamide cofactor, to NADH, inactivates the enzyme, but that addition of exogenous NAD$^+$ rapidly reactivates it.[93,127] This is consistent with the proposed mechanism, in which the NAD$^+$ cofactor must oxidize the C3 hydroxyl in the first step. Small primary KIEs ranging from 1.63 to 2.03 for $(k_{cat})_H/(k_{cat})_D$ and $(k_{cat}/K_M)_H/(k_{cat}/K_M)_D$, were measured for substrates with deuterium substitutions individually at the C2 and C3 positions.[46,93] These KIEs indicate that cleavage of the C2–H2 and C3–H3 linkages, leading to the oxidation of the C3 hydroxyl and C2 deprotonation, respectively, are rate limiting. The primary KIEs are substantially smaller than the maximum theoretical value of 6–7, consistent with the fact that both steps are partially rate limiting. No secondary KIE was measured for the 1-deutero substrates, indicating that the elimination step is fast and nonrate limiting.[93,127] This conclusion was supported by a study of the leaving group dependency of kinetic parameters.[93,127] The plots of the logarithms of k_{cat} and k_{cat}/K_M were essentially flat for the aryl glucopyranosides investigated, indicating that either cleavage of the C1–O1 linkage is not rate limiting, or, if the bond cleavage is rate limiting, there is no significant negative charge development on the glycosidic oxygen at the transition state.[93,127] Since no secondary KIE was measured for the 1-deutero compound, the latter case is most likely.[93,127] Additionally, thioglycosides, which are known competitive inhibitors of glycosidases that utilize nucleophilic displacement mechanisms, were found to be substrates for BglT.[126] It has been suggested[149–155] that thioglycosidic linkages are not cleaved by glycosidases, because there is little general acid catalysis to the departing oxygen atom due to the fact that sulfur has a lower proton affinity than oxygen. (The exceptions are the myrosinases of plant origin. These S-glycosidases have evolved to specifically hydrolyze glucosinolate substrates, anionic 1-thio-β-glucosides. By sequence alignment, the myrosinases belong to GH1, which contain enzymes that catalyze the hydrolysis of β-O-glycosides with retention of the substrate anomeric configuration. It has been proposed that the myrosinases are able to catalyze the cleavage of glucosinolates, because the substrates contain inherently good leaving groups thus do not require general acid assistance for the departure of the leaving group.) BglT, using an elimination mechanism, was found to catalyze the hydrolysis of the thioglycosidic linkage just as efficiently as the O-glycosidic linkage.[126] Although a GH1 β-glucosidase,[156] a GH84 O-GlcNAcase,[152] BtGH84 (the human symbiont B. *thetaiotaomicron* VPI-5482),[157] a sialidase from *Micromonospora viridifaciens*[154] and sweet almond β-glucosidase[155] have been shown to be capable of the hydrolysis of thioglycosides as well, these enzymes only react with thioglycosides containing highly activated leaving groups. BglT is currently the only glycosidase capable of hydrolyzing thiodisaccharides at rates comparable to their oxygen counterparts.[126] The ability to cleave thioglycosides can be easily rationalized for the anionic mechanism proposed, but not for hydrolysis via oxocarbenium ion-like transition states.[126] Altogether, the accumulated KIE data suggested that BglT utilizes an E1$_{cb}$ mechanism in the hydrolysis of these substrates.[46,126,127]

The catalytic residues of GH4 enzymes have not been definitively identified, although there have been some kinetic studies of mutant GH4 enzymes. However, most of these studies were performed prior to the availability of X-ray crystal structures and before the unusual mechanism was unraveled. Therefore, the amino acids of interest were those expected for a classical glycosidase mechanism, namely the conserved Glu or Asp residues, as well as residues that are characteristically found in the NAD$^+$-binding Rossmann fold. Although mutation of these residues resulted in a loss of activity, they were subsequently shown to be located outside the enzyme active site. Instead of the usual Asp or Glu residues utilized by other glycosidase families, a key Cys residue was identified at the enzyme active site, chelated to the divalent metal ion.[46,90,135] In fact, AglA, an α-glucosidase from *T. maritima*, was shown to be inactive when this thiol functionality was oxidized to a sulfinic acid.[135] Subsequently, through kinetic analysis of a mutant enzyme, an Asp residue adjacent to this active site Cys was also shown to be essential for catalytic activity.[131] Other important catalytic residues were suggested based on the structural analysis.[46,90,125,127] The most useful information was garnered from BglT and

GlvA.[46,90,125,127] Not only is detailed mechanistic information available, but the X-ray crystal structures of both enzymes in their catalytically active forms along with all the necessary cofactors have been solved. A Tyr residue has been proposed to play the role of the catalytic base that deprotonates C2 in GlvA and BglT, both 6-phospho-glycosidases.[46,90,125,127] By contrast, occupying the same space and most likely carrying out the same role in AglA is an Asp residue.[46,90,125,127,135] The utilization of two different amino acid residues by the same family of enzymes may seem surprising, but it has been suggested to be necessary to minimize any electrostatic repulsion between the substrate C6 phosphate and the carboxylic acid functionality of an Asp.[46,90,125,127] A similar situation holds for the sialidases and *trans*-sialidases, which have also evolved to employ a Tyr as the key catalytic residue in preference to either the Asp or Glu used by most other 'classical' glycosidases, presumably to better accommodate the substrate carboxylic acid group at C1.[54,158] For GH4, the identity of the general acid that protonates the glycosidic oxygen is unclear and probably variable.[46,90,125,127] This role was suggested to be filled by an Asp in GlvA[90] and a His in AglA,[135] while no amino acid residue was found to occupy the equivalent space in the BglT structure.[46,125,127] Since BglT utilizes a *syn*-elimination mechanism, the Tyr base responsible for C2 deprotonation is only 3.4 Å away from the glycosidic oxygen.[46,125,127] Therefore, the Tyr residue could play a dual role: deprotonating C2 as well as assisting in C1–O1 cleavage by protonation of the glycosidic oxygen.[46,125,127]

The recently discovered GH109 has been found to display some of the unusual properties found in GH4 enzymes.[159] The GH109 α-*N*-acetylgalactosaminidase from *Elizabethkingia meningoseptica* also requires NAD^+ for catalytic activity and can hydrolyze α- and β-glycosidic linkages as well as thioglycosidic linkages. Although the enzyme has low overall structural similarity with GH4 enzymes, the active sites of GH4 and GH109 enzymes are nearly identical. The NAD^+ cofactor, substrate, and the proposed Tyr base are essentially found in the same spatial arrangement. Furthermore, the GH109 enzyme also has structural resemblances to oxidoreductases. These findings along with the fact that it could cleave thioglycosides and catalyze C2 proton exchange in the substrate suggested that it utilizes a mechanism similar to that of GH4.

Although GH4 was the first glycosidase family reported to formally operate via an elimination mechanism, other glycosidases and glycosyl transferases have been known to 'leak' minor elimination products. These enzymes include the nucleoside 2-deoxyribosyltransferase from *Lactobacillus leishmanii*,[160] which releases D-ribal along with the normal transfer product, and the GH31 mammalian glucosidase II,[161] which generates 1,5-anhydrofructose as a minor side product. Indeed, the α-1,4-glucan lyase[162–164] discussed below is also grouped under GH31,[23] so minor lyase activity in glycosidases in this family is quite plausible. However, these enzymes all operate via oxocarbenium ion-like transition states.

8.11.3.2 Glycoside Hydrolase Family 31 and α-Glucan Lyases

Considering primary sequence similarity, the α-1,4-glucan lyases are grouped into the GH31 retaining α-glycosidase family.[23] Mechanistic analysis on the α-glucosidases AglA from *Aspergillus niger* and the α-xylosidase YicI from *Escherichia coli* in GH31 demonstrated that the glycosidases of this family employ a double-displacement mechanism.[82,165] GH31 α-1,4-glucan lyases, on the other hand, cleave the glycosidic linkages in starch and glycogen via an elimination mechanism, releasing a 1,5-anhydrofructose product (**Scheme 5**).[162–164] The GH31 *Gracilariopsis* α-1,4-glucan lyase was found to be inhibited by compounds such as acarbose,[162,166] and 1-deoxynojirimycin (DNJ),[162,166] and inactivated by carbodiimides[167] and 5-fluoro-β-L-idopyranosyl fluoride.[162–164] These inhibitors either react selectively with Asp and Glu residues or inhibit the enzyme by mimicking the oxocarbenium ion-like transition states of nucleophilic displacement mechanisms.[78,168–170] Evidently, this lyase shares mechanistic similarities with GH31 glycosidases, most likely proceeding via a similar transition state, and having one or more Asp/Glu residue(s) play catalytic roles.[162–164] This assertion was supported by the fact that a glycosyl–enzyme intermediate was trapped using the reagent 5-fluoro-β-L-idopyranosyl fluoride, which led to the identification of Asp553 as the catalytic nucleophile, the same residue as that identified in GH31 glycosidases and completely conserved within this family.[82,162–165]

The proposed α-1,4-glucan lyase mechanism consists of two steps: formation of a glycosyl–enzyme intermediate and eliminative cleavage of the C1–O1 linkage.[162–164] In order to probe this, a linear free energy plot was generated using substrates of varying ability of the leaving group. The resultant Brønsted relationship had a β_{1g} value of −0.32, indicating that cleavage of the C1–O1 linkage is rate limiting for those substrates and that
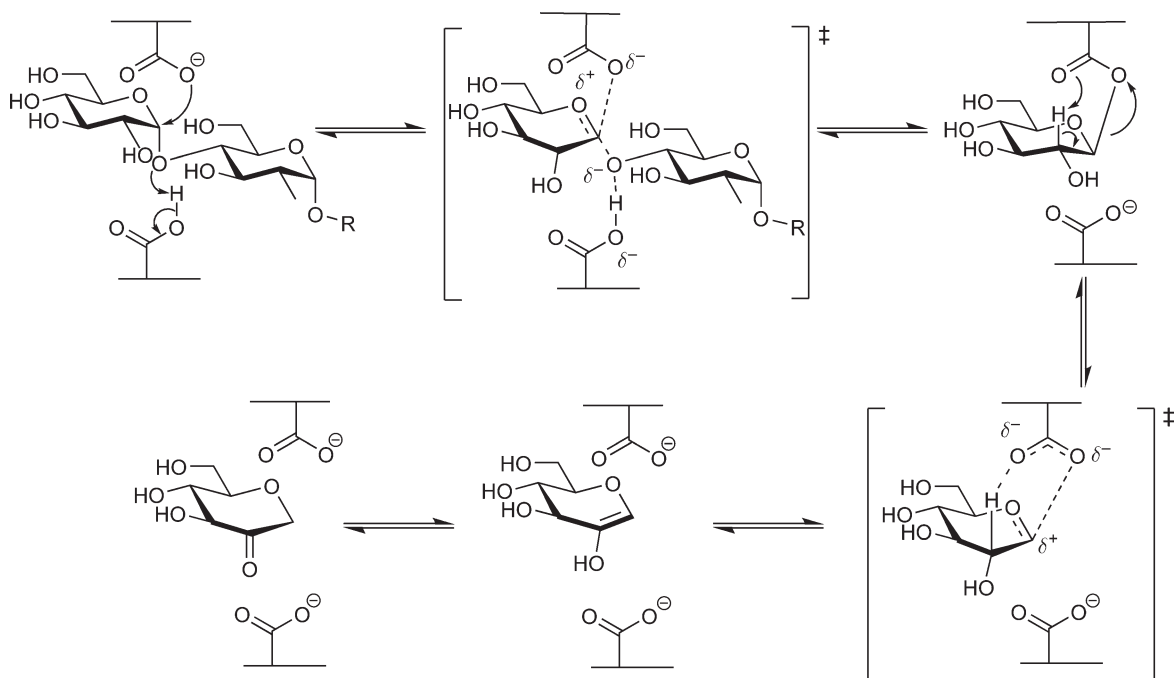
**Scheme 5**   Proposed mechanism of the *Gracilariopsis* α-1,4-glucan lyase from GH31.[162]

the linear free energy plot suggests substantial negative charge development at O1 at the transition state.[162,164] For substrates with poor ability of the leaving group, a large secondary KIE of $(k_{cat})_H/(k_{cat})_D = 1.19$ was measured when a deuterium was substituted at the C1 position, indicative of oxocarbenium ion character at the initial bond cleavage step.[162,164] KIEs were also determined for a substrate for which the elimination step is rate limiting, and the result was a large α-secondary KIE of $(k_{cat})_H/(k_{cat})_D = 1.23$ at C1 and a small primary KIE of $(k_{cat})_H/(k_{cat})_D = 1.92$ at C2.[162,164] Thus, the first part of the proposed mechanism (**Scheme 5**) is identical to that of retaining glycosidases, with formation of the glycosyl–enzyme intermediate proceeding via an oxocarbenium ion-like transition state.[17,162,164] Breakdown of the glycosyl–enzyme intermediate, however, is proposed to proceed via an E2 elimination with substantial E1 character.[17,162,164] The oxocarbenium ion character that develops at the transition state lowers the effective pK_a value of the C2 proton, thereby facilitating elimination of the aglycone to form the 1,5-anhydrofructose product.[17,162,164]

The catalytic residue responsible for C2 deprotonation in this *syn*-elimination has been proposed to be the anionic nucleophilic residue that is released upon cleavage of the C1–O1 linkage of the glycosyl–enzyme intermediate.[17,82,162,164] Unfortunately, because the structure of α-1,4-glucan lyase has not been reported, structural evidence can only be inferred from the homologous GH31 α-xylosidase YicI. Indeed, in the crystal structure of its glycosyl–enzyme intermediate, the carbonyl oxygen of the catalytic nucleophile is found in close proximity to the substrate C2 proton.[82] Furthermore, retaining glycosidases have long been known to catalyze the hydration of glycal substrates in a similar manner.[162,171–174] In those cases, the glycal initially undergoes a *syn*-addition: the catalytic nucleophile protonates C2 and attacks the anomeric center to form the glycosyl–enzyme intermediate.[171–174] This *syn*-addition is essentially the reverse of the elimination step in α-1,4-glucan lyase.[162–164,171–174] Conversely, in the structures of other α-glycosidase families (GH13[6] and GH38[175]) that do not possess α-1,4-glucan lyase or any lyase activity, the carbonyl oxygen of the catalytic nucleophile would not be able to carry out C2 deprotonation since the carboxylic acid functionality is rotated such that the carbonyl oxygen is close to the endocyclic oxygen in these cases.

8.11.3.3 Hexosaminidases

The term hexosaminidase can be used to broadly describe those enzymes that catalyze the hydrolysis of the glycosidic linkages of 2-acetamido-2-deoxy-β-D-glycosides. 2-Acetamido-2-deoxy-β-D-glycosides are found widely in nature. For example, the polymer chitin is present in fungi and exoskeletal organisms, where it provides structural support. These carbohydrate moieties also play important roles as part of gangliosides, which contain a terminal nonreducing GalNAc residue. Gangliosides are present in all mammalian cells and are often involved in signal recognition,[176] regulation of cellular activity,[177,178] and are known to be antigens.[177–203] In addition, essentially all glycoproteins contain N-acetylhexosamine residues, whether N-linked or O-linked.[204–207] Therefore, the enzymes that catalyze the hydrolysis of the 2-acetamido-2-deoxy-β-D-glycosides are of fundamental interest. This group of enzymes includes chitinases, chitobiases, hyaluronidases, and lysozymes, and the enzymes can be found in GH3, GH18, GH20, GH22, GH56, and GH84.[23] The majority of these enzymes are known to utilize a double-displacement mechanism. For some families (GH3, GH22), the classical Koshland double-displacement mechanism already described in Section 8.11.2.1 is employed.[208] Other families utilize a substrate-assisted mechanism of hydrolysis.[208] The substrate-assisted mechanism employed by GH 18, 20, 56, and 84 nonetheless shares mechanistic similarities with the classical retaining glycosidases. The major difference lies in the fact that the nucleophile attacking the anomeric center does not belong to the enzyme. The nucleophile is the carbonyl oxygen of the substrate acetamido group adjacent to the anomeric center, such that an oxazoline intermediate is formed (**Scheme 6**).[47,208–210] Further discussion focuses only on those hexosaminidases that utilize the substrate-assisted mechanism employed by GH 18, 20, 56, 84, and 85. The mechanism in question is essentially that shown to be followed during acid-catalyzed hydrolysis of 2-acetamido-2-deoxy-β-D-glucosides,[48,211] as shown in **Scheme 7**.

The substrate-assisted mechanism was originally proposed for an enzyme, lysozyme in the late 1960s by Lowe,[212,213] but shown not to be followed in that case and essentially forgotten in the interim until several lines of evidence suggested that several families of hexosaminidases might use such a mechanism. The X-ray crystal structure of the bacterial chitobiase from *Serratia marcescens*, showed that the substrate was in fact distorted into a sofa conformation with the carbonyl oxygen of the acetamido group in close proximity to the anomeric
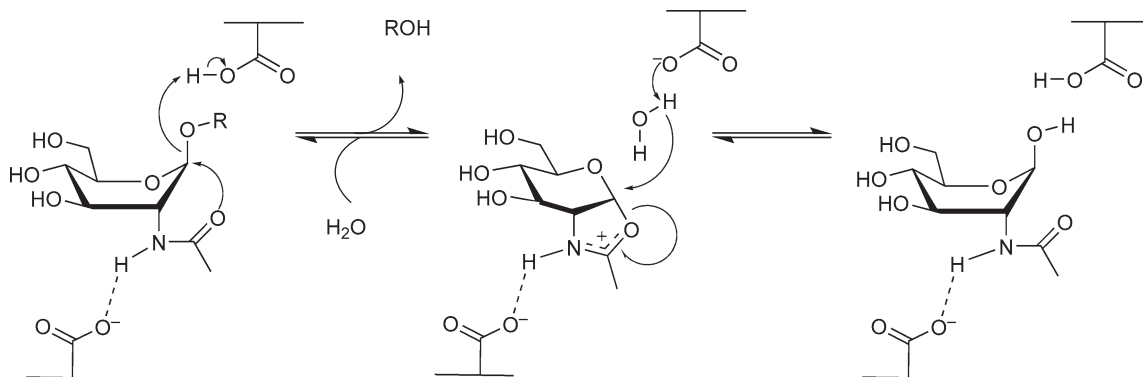
Scheme 6   General mechanism utilized by a retaining N-acetyl-β-hexosaminidase.
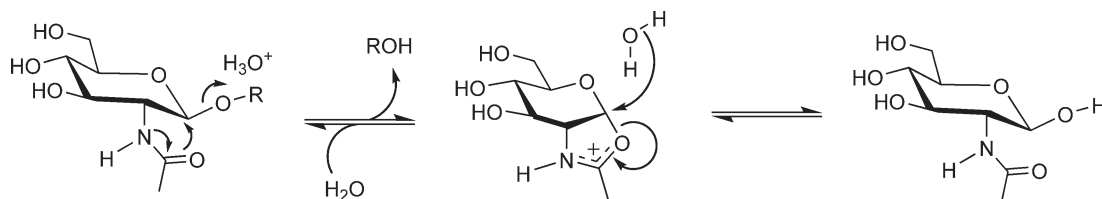
Scheme 7   Acid-catalyzed hydrolysis of a 2-acetamido-2-deoxy-β-D-glucoside.

center,[210,214] as required for the substrate-assisted mechanism. Additional evidence derived from the fact that in the X-ray crystal structures there was no nucleophile or stabilizing carboxylate group close to the anomeric center, even though the enzymes catalyze the hydrolysis of their substrates with net retention of anomeric configuration.[50,215] In an acid-catalyzed process, the anomeric center is nucleophilically attacked by the carbonyl of the 2-acetamido group to form an oxazoline or oxazolinium ion intermediate. Subsequently, a water molecule, activated by general base catalysis, performs a nucleophilic attack at the anomeric center, resulting in the formation of the hydrolysis product with net retention of the anomeric configuration.[47] This mechanism is supported by both structural[47,210,214,216–219] and kinetic studies,[49,209,220] with the most direct evidence being that oxazolines act as excellent substrates for enzymes of this class.[221–223]

Two well-studied examples are used to illustrate this mechanism: the GH84 O-GlcNAcase and the GH20 SpHex from *Streptomyces plicatus*.[209] O-GlcNAcase is particularly interesting due to potential biological applications in transcription, cytokinesis, protein phosphorylation, and tumor suppression.[204,205,224–233] O-GlcNAcase is highly conserved in organisms ranging from *Caenorhabditis elegans* to humans.[157,204,224,226,228–231,233–235] The enzyme is similar to other retaining glycosidases in that two carboxylic acid residues play key catalytic roles.[236] In the first step, namely cyclization, one of the carboxylic acid residues directs and polarizes the 2-acetamido group to act as the nucleophile, forming the oxazoline intermediate. In the cases of both SpHex[209] and O-GlcNAcase,[153] the substitution of the proposed Asp by Ala resulted in mutants of lower enzyme activity and with an altered pH-activity profile relative to the wild-type enzymes. Without the catalytic Asp residue to activate it, the 2-acetamide group acts as a much poorer nucleophile.[209] Interestingly, the addition of exogenous azide rescued the activity of these mutants, presumably via anionic stabilization.[153] A second carboxylic acid acts as the general acid, assisting the departure of the aglycone. When the proposed Glu residue was mutated to either Ala or Gln in O-GlcNAcase[153] and SpHex,[237] the resultant mutants had significantly lower k_{cat} and K_M values. In the second ring-opening step, the carboxylic acid that enhanced the nucleophilicity of the 2-acetamido group in the first step facilitates the departure of the same group. Meanwhile, the general acid residue now acts as a general base, promoting the attack of a water molecule to yield the hydrolysis product.

Evidence for existence of the oxazoline intermediate is found in the ability of GlcNAc-thiazoline, which mimics the oxazoline structure, to inhibit these enzymes.[157,209,220,238–240] Structural analyses of hexosaminidases in complex with these inhibitors have yielded useful mechanistic insights.[47,157,241] In the case of SpHex and O-GlcNAcase, the structures of the two transition states flanking the oxazoline intermediate have been investigated through substrate structure–function studies, including α-deuterium KIEs.[152,208,220,240,241]

Further insights have come from kinetic studies with substrates in which the hydrogen atoms of the 2-acetamido group have been replaced with varying numbers of fluorine atoms to alter the nucleophilicity of the carbonyl oxygen. Taft-like free energy relationships were measured for SpHex and O-GlcNAcase.[152,208] The introduction of electronegative atoms slows down the cyclization step by making the acetamido group less nucleophilic, but the breakdown of the oxazoline intermediate is accelerated since the fluorine substitution renders the acetamido group a better leaving group than the nonfluorinated complement. Similar to other hexosaminidases,[152,157,242] a Taft parameter of −1.29 was measured for SpHex,[208] which is consistent with the participation of the 2-acetamido group at the transition state. The transition state has also been investigated using substrates with leaving phenol groups of different leaving group ability at the anomeric center. Small negative $\beta_{1 g(k_{cat}/K_m)}$ values of −0.29, and −0.11 were measured for SpHex[208] and O-GlcNAcase,[152] respectively, suggesting that little negative charge development occurs at the leaving group oxygen during the glycosidic bond cleavage step either due to an early or to a late transition state. If it is an early transition state, neither the cleavage of the glycosidic linkage nor the protonation of O1 has occurred to a substantial extent. If it is a late transition state with substantial C1–O1 bond cleavage, then significant proton donation must have occurred. The latter case is more likely since substantial α-secondary deuterium KIEs (($k_{cat})_H/(k_{cat})_D$) of 1.14 and 1.07 were measured for O-GlcNAcase[152] and SpHex,[208] respectively, indicating oxocarbenium ion character, but not to the extent seen for the 'classic' Koshland-type retaining glycosidases, presumably as a consequence of the neutral (vs. anionic) nucleophile. The human GH 84 O-GlcNAcase was reported to catalyze the hydrolysis of activated thioglycosides, albeit with approximately 20-fold lower catalytic efficiency, similar to that reported earlier for a GH1 enzyme.[152]

In conclusion, although the majority of retaining hexosaminidases use the oxazoline mechanism, it is important to remember that enzymes in GH3 and GH22, and presumably others, use the classic Koshland

mechanism. The 2-acetamido functionality of the substrates for GH3 and GH22 enzymes is known to be important for catalysis, but plays the role of the catalytic nucleophile. The enzyme supplies both the general acid/base catalysts as well as the nucleophile, as clearly demonstrated by the isolation of trapped glycosyl–enzyme intermediates, with examples including those of hen's egg white lysozyme,[243] and *Vibrio furnisii* N-acetyl-β-D-glucosaminidase.[244] The interested reader is referred to the publication by Vocadlo and Withers,[208] which provides a meticulous analysis of the similarities and differences between the classical retaining glycosidase mechanism and the substrate-assisted mechanism employed by the two distinct groups of hexosaminidases.

8.11.3.4   Sialidases

Sialic acids are a family of nine-carbon monosaccharides which are usually found at the outermost positions of a number of glycoconjugates.[245,246] The key feature of this class of sugars is that it has a carboxylate group at the anomeric center which is generally ionized at the physiological pH. The most famous member of this family is N-acetyl neuraminic acid (Neu5Ac) which was first discovered more than 50 years ago, with the other being 2-keto-3-deoxy-D-*glycero*-D-*galacto*-nonulosonic acid (KDN) (**Figure 4**). These two molecules can be viewed as precursors for the rest of the members in this family since their structural diversity can be generated by subsequent enzymatic modifications such as acetylation, methylation, phosphorylation, and so on at C-4, C-5, C-7, C-8, C-9 positions. The wide occurrence of sialic acid molecules, especially in higher invertebrates and vertebrates, are consistent with their fundamental physiological roles. For example, it has been found that terminal sialic acids are involved in a range of essential cell–cell or cell–molecule recognition processes.[247] The structural diversity of sialic acids allows the mediation of a range of highly specific biological recognition events, which is beneficial to the host organism. However, these terminal structures have been exploited by a wide array of pathogens during pathogenesis. A famous example is the infection by influenza virus being initiated by the recognition and binding of cell surface sialic acid-capping glycoconjugates of the host.[248,249] Given the importance of sialic acids, sialidases (or neuraminidases), and *trans*-sialidases, whose functions are removing or transferring sialyl residues, have been a subject of extensive research over the past two decades.[250] Detailed mechanistic understanding of these enzymes is helpful in modulating their activities and development of novel antiviral drugs.

   Depending on the action mode, sialidases and *trans*-sialidases can be classified as either exo-sialidases or endo-sialidases. Owing to the limited information on GH58 endo-sialidases,[251] the following mechanistic discussions are focused on exo-sialidases, which include enzymes of GH33, 34, and 83. Most of the GH33 sialidases and *trans*-sialidases are from bacteria and eukaryotic organisms while GH34 contains only viral neuraminidases and GH83 contains hemagglutinin neuraminidases. Through rigorous NMR studies, all exo- and *trans*-sialidases discovered to date have been found to catalyze their reactions with retention of anomeric configuration,[252–257] implying a classical double-displacement mechanism.

   Perhaps the best insights into sialidases can be deduced from their crystal structures. To date, the structures of sialidases from eight strains of influenza A viruses[258–263], two strains of influenza B viruses,[264,265] three bacteria (*M. viridifaciens,*[266] *S. typhimurium,*[267] and *V. cholera*[268]) and four eukaryotic organisms (*H. sapiens,*[269] *M. decora,*[270] *T. cruzi,*[271] and *T. rangeli*[272]) have been determined. Despite the lack of high overall sequence similarity and the existence of noncatalytic domains in many sialidases, common catalytic cores have been
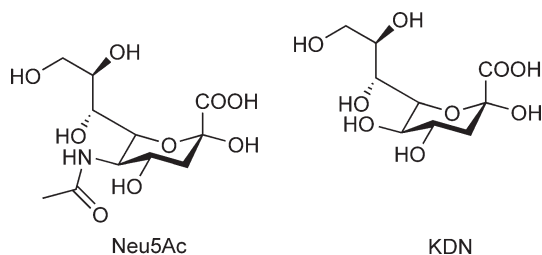
Figure 4   Structures of Neu5Ac and KDN.

found in enzymes from all levels of organisms, implying the sharing of the same ancient sialidase ancestor and adoption of a similar catalytic mechanism for all these enzymes. A six-bladed β-propeller constitutes the classical catalytic domain, with each blade being composed of four antiparallel β-sheets. Soaking these protein crystals with various substrates and inhibitors such as sialic acid conjugates or 2-deoxy-2, 3-dehydro-N-acetylneuraminic acid (DANA),[273] has helped reveal the active site location and topology. The active site is a deep pocket located near the center of the six-bladed β-propeller. Amino acid residues that are responsible for the recognition of substrate or directly involved in catalysis can be found lining this pocket. The carboxylate group attached to the anomeric center of the sialic acid is always surrounded by three Arg residues, which are often referred to as the 'Arg triad'. By contrast, the N-acetyl group at C-5 is buried in a hydrophobic pocket, while the glycerol group at C-6 is held in place by a series of hydrogen bonding interactions with the active site. A tyrosine residue is positioned in the proximity of the anomeric carbon, implying its importance in catalysis. However, the binding pocket around the hydroxyl at C-4 varies between neuraminidases from different sources, as has been exploited in the rational design of antiviral drugs.[249]

The catalytic mechanism employed by sialidases has been under intensive debate over the past two decades. The notion that sialidases or *trans*-sialidases are retaining glycosidases makes the well-established double-displacement mechanism involving two carboxylate residues a likely option[17] (**Scheme 8**). However, careful inspection of the crystal structures of sialidases clearly shows a tyrosine rather than a carboxylate at the corresponding position.[271,272] The relatively high pK_a of the phenol makes tyrosine a less attractive candidate as the nucleophile, therefore a mechanism involving an ion-pair intermediate was long thought to be followed.[252] In the first step of such a mechanism, the aglycone leaves with the assistance of general acid catalysis, yielding a long-lived sialyl cation intermediate, which could be stabilized by the anionic C-1 carboxylate. This sialyl cation then undergoes attack by a water molecule that is deprotonated by the general base. Indeed, mechanistic studies of both acid-catalyzed and anion-catalyzed solvolysis of CMP-N-acetyl-neuraminate indicated that a sialyl cation intermediate with a finite lifetime exists.[274,275] The existence of a sialyl cation was further supported by the observation of a small inverse β-secondary dideuterium KIE on influenza virus A neuraminidase-catalyzed hydrolysis of 4-methylumbelliferyl α-D-neuraminide ($kH/kD = 0.979 \pm 0.007$),[252] which was in agreement with molecular modeling studies.[276] However, these results were not clear-cut and other evidence has emerged to favor the double-displacement mechanism. In one study, the KIE on the hydrolysis of 3,3-dideuterated 4-nitrophenyl α-D-neuraminide by influenza sialidase was reported to be 1.0095 ± 0.011, which brought the earlier inverse KIE into question.[254] Similar results were obtained in the case of *V. cholerae* sialidase.[253] In another case, β-dideuterium and primary ^{13}C isotope effects were measured for two substrates, sialyl-lactose and sialyl-galactose, under both acid hydrolysis and *T. cruzi trans*-sialidase-catalyzed conditions,
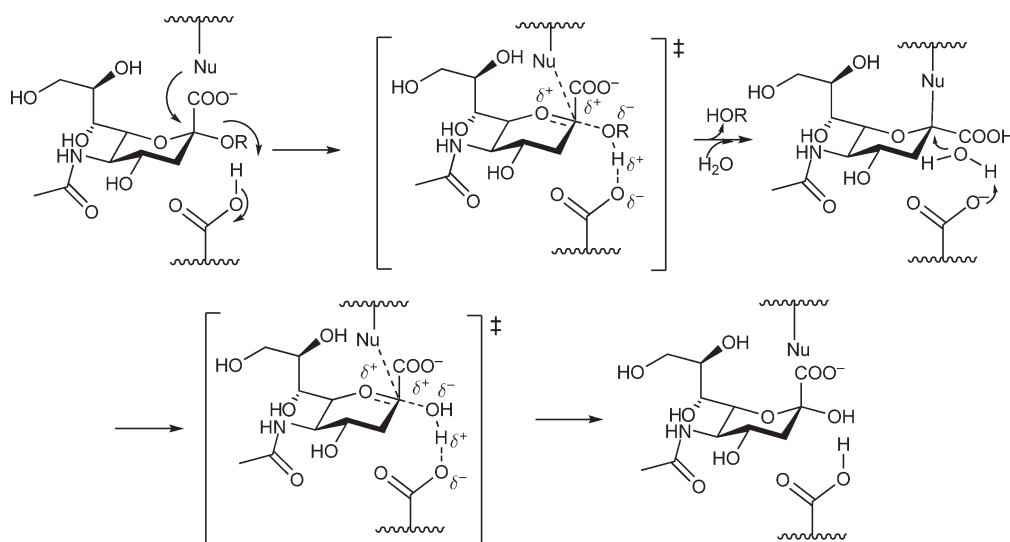

Scheme 8   Catalytic mechanism of a sialidase or a *trans*-sialidase.

respectively. While the results on the acid-catalyzed reaction indicated a reaction pathway involving a dissociative S_N1-like mechanism, all the isotope effects measured on the enzyme-catalyzed reactions are normal (>1).[119] Despite all these kinetic studies, however, definitive evidence concerning the identity of the key intermediate remained elusive until 2003 when the fluorosugar methodologies were applied to sialidases.

Watts *et al.*[54] incubated a 2,3-difluoro sialic acid derivative with a *trans*-sialidase from *T. cruzi* (TcTS) and this resulted in a time-dependent loss of enzymatic activity, consistent with trapping of a covalent sialyl–enzyme intermediate. Removal of the excess inactivator by ultrafiltration followed by incubation of the inactivated enzyme in buffer allowed the turnover of the intermediate to be followed. First-order spontaneous recovery of enzyme activity was found with a half-life of 12.6 min, with reactivation being accelerated considerably in the presence of lactose. This demonstrated the catalytic competence of the trapped intermediate while its structure was first shown by mass spectrometry. The formation of labeled TcTS with a single sialyl residue attached was observed upon incubation of the enzyme with the inactivator. To determine the identity of the nucleophile residue, inactivated TcTS was subjected to pepsin digestion and this sample, in parallel with a pepsin digest of an unlabeled TcTS, was analyzed by LC–MS. Comparison of the two elution profiles identified the labeled peptide, identifying tyrosine as the nucleophile. The same strategy was applied to the GH33 sialidase from *T. rangeli* (TrSA), which is a strictly hydrolytic enzyme, as opposed to TcTS.[277] Two 2,3-difluoro sialic acid analogues, 2,3-difluoro-*N*-acetyl-neuraminic acid (1) and 2,3-difluoro-2-keto-3-deoxy-D-glycero-D-galacto-nonulosonic acid (2, 3-difluoro – KDN) (**Figure 5**) were tested and 2, 3-difluoro – KDN was shown to be a mechanism-based inactivator with the half life of the trapped intermediate being 37 min. Similarly, 2,3-difluoro-*N*-acetyl-neuraminic acid trapped the intermediate on the sialidase from *C. perfringens*, which was also characterized via crystallography.[278] All these studies demonstrated the generality of the double-displacement mechanism for sialidases and that tyrosine is the nucleophile involved. Indeed, inspection of sequences of enzymes from family 33, 34, and 83 clearly shows that this tyrosine is completely conserved,[54] while analysis of the available crystal structures of sialidases revealed that this tyrosine was positioned appropriately.[271,272] Possibly, the tyrosine has evolved to serve this role by avoiding the electrostatic repulsion that would exist between the carboxylate at the anomeric center of the sialic acids and an anionic nucleophile such as Asp or Glu. Instead, sialidases place a highly conserved glutamate residue adjacent to the tyrosine, presumably to function as a base catalyst. Thus sialidases appear to have evolved their own charge relay strategy analogous to that in serine proteases.

The use of fluorinated sialic acid analogues to trap intermediates opens the concern that the reaction mechanism has been altered. This concern was overcome in recent studies wherein classical ping-pong kinetics were observed for TcTS, consistent with the existence of a covalent intermediate.[279] In addition, incubation of the acid/base mutant D59A TcTS with an 'activated' substrate, *p*-nitrophenyl α-sialoside (PNP-SA), resulted in a pre-steady-state 'burst' of PNP release whose magnitude corresponded well with the amount of TcTS used. Further, when PNP-SA was incubated with either D59A TcTS or wild-type TcTS, the accumulation of the sialyl–enzyme intermediate was clearly observed by mass spectrometry.

In parallel with the above, evidence for a tyrosine nucleophile was generated by kinetic analysis of mutants. Bennet and colleagues generated the nucleophile mutants Y370A, Y370D, and Y370G of *M. viridifaciens* sialidase[53] and found significant residual activity (3–86%) when assayed with an activated substrate. NMR analysis revealed that the reaction proceeded with net inversion of the substrate anomeric configuration, even though the wild-type enzyme is a retaining enzyme. This change of mechanism is presumably the result of a
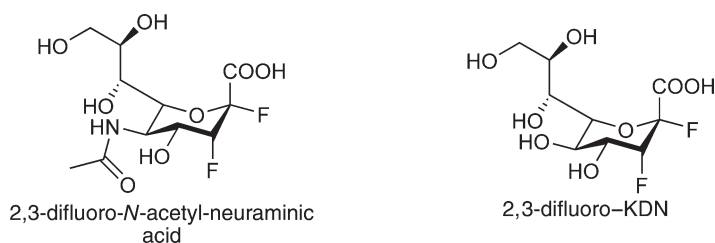
Figure 5   Structures of fluorinated sialic acid analogues.

water molecule fitting into the space generated by the mutation and replacing the role of the missing amino acid residue. The mutant enzyme therefore retains the ability to catalyze the first nucleophilic displacement step but uses water as the nucleophile without formation of a sialyl–enzyme intermediate. Indeed, a similar study had been carried out previously on the retaining β-glucosidase (Abg) from *Agrobacterium faecalis* in which the nucleophile Glu358 was mutated to Ala.[45] While the mutant alone was inactive, the addition of small anions such as azide and formate successfully rescued its activity and converted Abg into an inverting enzyme.

Structural studies on sialidases complexed respectively with the substrate, a putative transition state analogue (DANA) and as the covalent intermediate have provided insights into the reaction pathway.[278,280] Structures of Michaelis complexes[281–283] revealed that the pyranose ring of the sialic acid is distorted into a $B_{2,5}$ conformation, likely driven by the salt bridge between the anomeric carboxylate and the 'Arg triad'. On the α-face of the sugar, a highly conserved carboxylic acid residue forms a hydrogen bond with the glycosidic oxygen. This is the general acid/base catalyst. On the β-face, as expected, the phenolic oxygen of the conserved tyrosine is found perfectly positioned to attack the anomeric center. Such distortion of the pyranose ring in Michaelis complexes is well-precedented in retaining glycosidases.[284] This places the aglycone in the axial position and allows the nucleophilic residue to perform an in-line attack, as favored by stereoelectronic theory. Structures of sialidase–DANA complexes, in contrast, revealed a half chair $^4H_5$ conformation of the pyranose ring for the putative transition state analogue, with the tyrosine oxygen in close proximity to C-2. The next stable species along the reaction coordinate is the covalent sialyl–enzyme intermediate, which was trapped by soaking sialidase crystals with 2,3-difluoro sialic acid derivatives.[277,280] A covalent bond of 1.42 Å length is clearly seen between the nucleophilic tyrosine residue of TcTS and C-2 of the sialic acid. The pyranose ring in this species was bound in a relaxed $^2C_5$ chair conformation, which has been seen in the glycosyl–enzyme intermediates of other families of glycosidases.[6] Interestingly, the number of noncovalent interactions between the substrate and the enzyme increases substantially upon formation of the covalent intermediate. For example, five new hydrogen bonds with the glycerol chain of sialic acid were seen in the sialyl–enzyme formed on TcTS, as well as between TcTS and DANA. This stabilization presumably increases the lifetime of the intermediate, thereby facilitating subsequent transglycosylation.

8.11.3.5   Inverting Glycosidases: Glucoamylases and β-Amylases

Among the approximately 112 currently known families of CAZy enzymes, there are 31 families assigned as inverting glycosidases. Compared with retaining glycosidases, the mechanistic information and structural data on inverting enzymes are relatively sparse, partly due to the simplicity of the catalytic mechanism. As mentioned in the introduction, all of the inverting glycosidases appear to use a single-displacement mechanism to cleave the glycosidic bond. The absence of any covalent glycosyl–enzyme intermediate precludes many of the aforementioned mechanistic approaches. While a number of affinity labeling reagents have been synthesized and applied to the identification of the active site residues, these have not proved generally applicable.[285] The most reliable method to identify catalytic residues in inverting glycosidases currently is by inspection of the crystal structure of the wild-type enzyme complexed with substrate analogues or inhibitors. Subsequent site-directed mutagenesis and kinetic analysis of the mutants can then be used to confirm their roles. Owing to their potential industrial applications, the most well-studied examples among the inverting families include those enzymes involved in amylolytic processes[286] and in the degradation of cellulose.[287,288] Therefore, in this section, the discussion focuses on mechanistic and structural studies on glucoamylases and β-amylases as representative examples of the inverting glycosidases.

α-Amylases, β-amylases, and glucoamylases are the major enzymes responsible for the degradation of starch for the metabolic needs of all organisms. α-Amylases are retaining endo-glycosidases, which hydrolyze the internal α-(1,4) glycosidic linkage in starch.[289] By contrast, β-amylases and glucoamylases are inverting exo-glycosidases removing maltose unit and glucose unit, respectively, from the nonreducing end of starch.[286]

8.11.3.5.1   *Glucoamylases*

Glucoamylases are GH15 inverting glycosidases, thus employ a single-displacement mechanism involving both a general acid and a general base residue. They have been isolated from a wide spectrum of organisms including archaea, bacteria, fungi, and yeast.[286,290–293] Owing to their industrial importance, glucoamylases have been

subjected to intensive structural and engineering efforts. Early examples of three-dimensional structures include the catalytic domains (CD) of glucoamylases from *Aspergillus awamori* var. *X100*[294,295] and the yeast *Saccharomycopsis fibuligera*.[296] The first crystal structure of a prokaryotic glucoamylase, from the species *Thermoanaerobacterium thermosaccharolyticum* was recently reported.[297] The following common features emerge from comparison of the available structures. First, all the reported structures possess a similar $(\alpha/\alpha)_6$ barrel CD. For example, in the CD of the glucoamylase from *Aspergillus awamori* var. *X100*, 12 out of the 13 α-helices are arranged into an $(\alpha/\alpha)_6$ barrel.[294,295] The cleavage site lies at the bottom of this barrel, consistent with the exo-action pattern of the enzyme. Second, besides the catalytic domain, some glucoamylases such as those from *Aspergillus awamori* var. *X100* and *Aspergillus niger*,[298,299] also have a C-terminal starch binding domain (SBD). This domain is thought not only to facilitate the binding of starch to the enzyme, but also to disrupt the surface of insoluble starch and thereby increase the rate of starch degradation.[300] The SBD is linked to the CD via a serine/threonine rich, highly *O*-glycosylated and flexible linker. No X-ray crystal structure of SBD in GH15 glucoamylase has been reported, though the SBD of *A. niger* glucoamylase has been characterized by NMR spectroscopy.[298,299] The recently determined structure of a prokaryotic glucoamylase revealed some differences.[297] In contrast to the single-domain CD of the eukaryotic glucoamylases, it is composed of two domains: an N-terminal super β-sandwich and a C-terminal $(\alpha/\alpha)_6$ barrel structure. It has been speculated that the eukaryotic glucoamylase may have evolved from its prokaryotic counterpart by substitution of the N-terminal β-sandwich domain with a peripheral sub-domain and the acquisition of an SBD. Finally, it should be noted that up to now, no crystal structure of an intact eukaryotic glucoamylase with all the domains present has been reported. Therefore, the exact spatial arrangement of CD, SBD, and the linker domain is still elusive. However, various biophysical techniques[301] and bifunctional inhibitors,[302] which bind to both CD and SBD, have been used to address this issue. After the submission of this manuscript, the first crystal structure of an intact GH15 glucoamylase which includes the catalytic domain, the linker domain and the starch binding domain was published.[303]

As mentioned above, glucoamylases adopt a single-displacement mechanism to hydrolyze the α-(1,4) glucosidic bond, releasing β-glucose as the product. Therefore, the key players in catalysis are the general acid and the general base residues. Earlier studies employing the affinity labeling reagent 1-ethyl-3-(4-azonia-4,4-dimethyl-pentyl)carbodiimide identified Asp176, Glu179, and Glu180 as possible active site catalytic residues in the glucoamylase from *A. niger* (equivalent to the catalytic domain of glucoamylase from *A. awamori*).[304] Subsequent mutagenesis studies indicated that Glu179 is the general acid residue.[305] The crystal structure of the glucoamylase from *A. awamori* var. *X100* complexed with the azasugar inhibitor, DNJ[306] further confirmed Glu179 as the general acid residue. However, a highly ordered water molecule was found to form hydrogen bonds with Glu400 and the 6-hydroxyl of DNJ, with the lone pair of electrons on the water oxygen atom being oriented toward the sugar anomeric center. It was therefore suggested that Glu400 was the actual general base residue, which was subsequently confirmed by kinetic analysis of the mutant E400Q.[307] However, the unusually high residual activity of this mutant (only 35–60-fold lower than wild-type glucoamylase, depending on the specific assay) prompted further analysis since in the case of α-amylase, mutation of any one of the three highly conserved catalytic residues decreases enzyme activity by at least 1000-fold.[308] One explanation for this high residual activity is that the transition state is very late, with a substantial oxocarbenium ion character.[307] Thus very little base catalysis of attack by water is required. This is consistent with KIE measurements with an α-glucosyl fluoride substrate.[309,310] An oxocarbenium ion-like transition state is further supported by the particularly strong inhibition afforded by acarbose toward *A. awamori* var. *X100* glucoamylase ($K_i = 10^{-12}$ mol l^{-1}).[311–313]

8.11.3.5.2 β-Amylases

β-Amylases are another important class of amylolytic enzymes, which catalyze the successive removal of β-maltose units from the nonreducing end of starch with inversion of anomeric configuration.[314] These enzymes have been classified into glycoside hydrolase family GH14 and are found primarily in plants and bacteria. Earlier pH-rate profile studies on sweet potato β-amylase and soybean β-amylase (SBA) demonstrated that both enzymes indeed employ two ionizable groups to carry out catalysis, with pKa values of 3.7 and 7.5 for sweet potato β-amylase and 3.5 and 8.5 for SBA, respectively.[315,316] On the basis of these values, it was initially speculated that the catalytic residues consist of one amino group and one carboxyl group for SBA. In order to confirm the existence of the carboxyl group, the affinity labeling reagent 2′,3′-epoxypropyl α-D-glucopyranoside was designed and shown to irreversibly

inactivate SBA in a stoichiometric fashion by attaching to Glu186.[317,318] The identity of the other catalytic residue was gleaned from crystal structures of wild-type SBA in complex with several ligands. The first structure of β-amylase, at 2-Å resolution, solved in 1993,[319] revealed that SBA is composed of a canonical $(\alpha/\beta)_8$ TIM barrel plus a small globular domain formed from some long loops that connect the alternating α-helices and β-strands of the $(\alpha/\beta)_8$ barrel. The active site was identified from the structure of a SBA/α-cyclodextrin (α-CD) complex. α-CD was found to bind at the entrance of the cleft with the catalytic residue Glu186 located further away at the bottom. It appears to inhibit SBA by blocking the entrance of the catalytic site rather than through an induced-fit mechanism as suggested by Koshland.[320] A detailed view of the active site was obtained from the structure of SBA in complex with β-maltose and maltal.[321] In both cases, two molecules of maltose or 2-deoxy maltose (product of maltal hydration) were seen to bind much deeper into the pocket than α-CD does. On the basis of the spatial arrangement of residues at the active site, Glu380 was identified as the general base residue while Glu186 was confirmed as the general acid residue. Mutation of these two residues did not result in a significant loss of substrate binding, but it did result in a 16 000-fold (E186Q) and a 37 000-fold (E380Q) decrease in activity relative to that of wild-type SBA.[322–324] Sequence alignment of several β-amylases also confirmed that Glu186 and Glu380 were conserved in all cases. Several other conserved regions were also identified and the residues involved therein were subjected to mutagenesis studies coupled with structural analysis to elucidate their roles.[322–323] Among these residues, Asp101 and Leu383 received the most attention.[322–323] Replacement of Asp101 with asparagine or glutamic acid yielded mutants with no detectable activity against soluble starch, while the L383I and L383Q mutants resulted in a significantly increased K_M and a reduced k_{cat}, indicating the importance of these two residues. Asp101 is part of a very flexible 8-residue loop (L3) in SBA.[319,321,325] Comparison of the structures of *Apo* SBA and SBA complexed with various small substrates revealed that L3 exist as two different conformations. In *Apo* SBA, L3 is primarily in an 'open' state. After the substrate binds to the active site, L3 closes over the substrate by forming several interactions with glucosyl residues. This 'open' and 'closed' switch is thought to be critical for SBA catalysis. Leu383 was found to form an inclusion complex with bound α-CD or β-CD via van der Waals interactions in the structure of SBA/α-CD and β-CD complexes.[319,321,325] The CD binding sites at the entrance to the active site, coupled with the interactions between Leu383 and the glucosyl residues, are widely regarded as crucial for the progressive digestion of starch by β-amylases from the nonreducing end. Currently, besides SBA, crystal structures of β-amylases from sweet potato,[326] barley,[327] and the bacterium *B. cereus*[328] have also been solved. The latter structure contains a similar $(\alpha/\beta)_8$ core region, along with an additional maltose-binding site in the core region and a C-terminal starch-binding domain in its structure.[328] These additional structural motifs are believed to be the key to the ability of this bacterial enzyme to digest raw starch granules.

β-Amylases have served as model enzymes for investigating enzymatic promiscuity among glycosidases. In 1969, Hehre *et al.*[329] observed that incubation of maltose with sweet potato β-amylase resulted in a novel maltotetraose product, which was proposed to arise from a condensation mechanism. Ten years later, again from the same laboratory, it was found that the more activated β-maltosyl fluoride is more readily hydrolyzed than is maltose by sweet potato β-amylase,[330] despite the fact that it has the 'wrong' anomeric configuration. Through kinetic and product analyses, a novel transglycosylation–hydrolysis pathway involving two molecules of bound substrate was proposed to rationalize this phenomenon. The general acid residue in the normal hydrolysis mechanism first acts as a general base catalyst, assisting the attack of the 4′-hydroxyl of a molecule of β-maltosyl fluoride bound in the +1 and +2 sites on the anomeric center of the second β-maltosyl fluoride, generating a β-maltotetraosyl fluoride. This transglycosylation product is then hydrolyzed through the normal mechanism, yielding β-maltose as the product. The second 'atypical' reaction catalyzed by β-amylases was the hydration of glycals.[331,332] Maltal could be converted into 2-deoxy-maltose by both sweet potato β-amylase and SBA. Exceptionally large solvent KIEs were observed in both cases, with $V_H/V_D = 8$ for sweet potato β-amylase and $V_H/V_D = 6.5$ for SBA, consistent with the protonation at C-2 being the rate-determining step. Product analysis clearly showed that the protonation at C-2 occurred above the plane of the double bond, while the attack of water also occurred from the β-face. The hydrolysis of glycosyl fluorides of the 'wrong' anomeric configuration has been found in several other inverting glycosidases[333] and in one mutant 'retaining' glycosidase[45] and thus is believed to be a general feature of inverting glycosidases. Similarly, many glycosidases, both inverting and retaining, are able to hydrate glycals. These studies have paved the way to the re-engineering of glycosidases for the synthesis of oligosaccharides, by suggesting ways of converting both retaining and inverting glycosidases into 'glycosynthases'.[124,333,334]
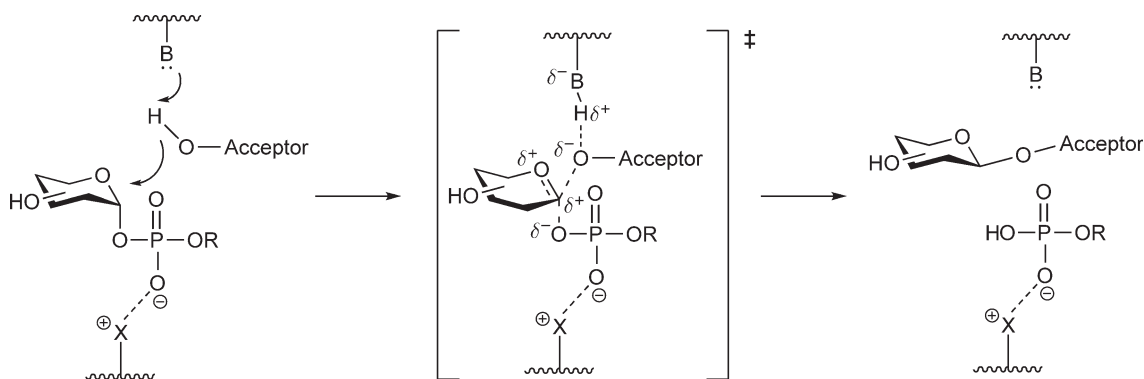
8.11.4   Glycosyltransferases

Glycosyltransferases are the key players in the construction of the vast majority of oligosaccharide structures in nature, yet by comparison with what is known of glycosidases, our mechanistic understanding of the glycosyltransferases is relatively limited. Analogous to the glycosidases, the formation of the glycosidic bonds occurs with either net retention or inversion of the anomeric configuration of the donor substrates, thus glycosyltransferases can be classified mechanistically as either inverting or retaining. The vast majority of the available structures of glycosyltransferases belong to one of the two canonical folds: GT-A and GT-B, and representatives of both inverting and retaining enzymes have been found for each fold, indicating that the general folds do not correlate with the stereochemical outcome. This section is not intended to be comprehensive on the whole area of glycosyltransferases. Instead, small representative groups of enzymes have been selected to illustrate recent structural and mechanistic understanding. The interested reader is referred to several detailed reviews published on the subject of glycosyltransferase structures and mechanisms.[13,22,335–337]

8.11.4.1   Inverting Glycosyltransferases: Fucosyltransferases and Sialyltransferases

As is the case with inverting glycosidases, inverting glycosyltransferases are thought to use a one-step single-displacement mechanism. Their active sites should therefore provide: (1) a general base catalyst in the active site to deprotonate the incoming acceptor; (2) a general acid catalyst or its equivalent to facilitate the release of the nucleoside phosphate leaving group from the donor substrate; (3) an extensive network of noncovalent interactions to stabilize the proposed oxocarbenium ion-like transition state (**Scheme 9**). These features are discussed using the following two examples.

8.11.4.1.1   Fucosyltransferases

Fucosylated glycoconjugates play important roles in a range of biological recognition processes[338] and as a consequence, fucosyltransferases (FucTs) have been the subject of extensive research. FucTs transfer the fucosyl residue from guanosine-diphosphate β-L-fucose (GDP-Fuc, **Figure 6**) to a broad range of acceptors such as oligosaccharides, glycoproteins, and glycolipids, with the formation of α-1,2, α-1,3/4, or α-1,6 linkages.[338] Among the FucTs, human α-1,3/4-fucosyltransferase V (FucT V) and *Helicobacter pylori* α-1,3/4-fucosyltransferase (hpFucT) have been investigated as model systems to understand this class of important enzymes from eukaryotic and prokaryotic organisms, respectively. FucT V is one of the six FucTs identified in the human genome that are involved in the terminal steps of biosynthesis of Lewis antigens and related glycoconjugates. It is a multidomain membrane protein with the CD located at the C-terminus, following an extended stem region in which an aromatic residue (Trp) is thought to be important in determining the acceptor specificity for this specific human FucT.[339,340] This is often refered to as the hypervariable region. Upstream of the stem region are a transmembrane domain and an N-terminal tail, and these two parts can be

Scheme 9   The single-displacement mechanism of inverting glycosyltransferases. R, nucleotide; X, divalent metal ion or charged side chains.
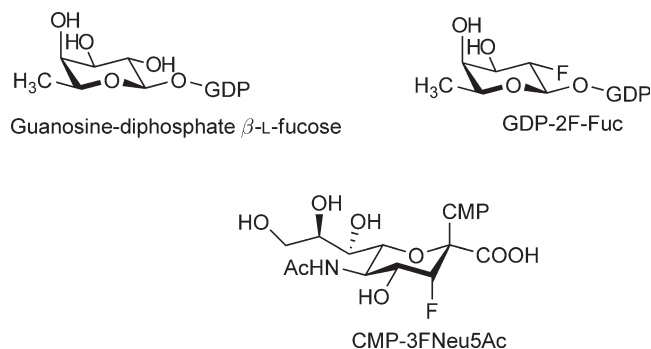
Figure 6 Substrates and fluorinated analogues for FucTs and sialyltransferases.

truncated without seriously affecting enzyme activity.[341] The hpFucT is also a multidomain protein but is arranged in the opposite fashion,[342,343] with the CD situated at the N-terminus. Downstream of the CD is a heptad-repeat, hypervariable region followed by two putative amphipathic helices at the C-terminus. Similar to FucT V, an aromatic residue in the heptad-repeat region has been identified as a key determinant of the acceptor specificity.[344] Deletion of the whole heptad-repeat region abolished all enzyme activity, consistent with its proposed role in promoting dimer formation via a leucine-zipper structure. The two putative C-terminus helices are thought to be membrane anchors for the protein and deletion of these two helices did not significantly affect enzymatic activity, but resulted in a much higher expression level and increased solubility of the truncated protein.[345] Although sequence alignment of these two enzymes only revealed short conserved regions within the CD,[342] it appears that these two enzymes have similar catalytic machinery.

Both FucT V and hpFucT belong to GT 10 according to the CAZy classification of glycosyltransferases, and catalyze the transfer of a fucosyl residue to the GlcNAc moiety of Type II chains (Galβ1,4GlcNAc) or Type I chains (Galβ1,3GlcNAc) with the formation of an α-1,3 or an α-1,4 linkage, respectively.[343] Earlier kinetic studies on FucT V demonstrated that this enzyme followed an ordered, sequential, bi–bi mechanism with the donor GDP-Fuc binding first followed by the acceptor LacNAc. After the transfer, the product Lewis X (Le^x) was first released, followed by GDP.[346] Similar sequential kinetics were seen with hpFucT and with other glycosyltransferases such as a β-1,4-galactosyltransferase,[347] and a β-D-xylosyltransferase.[348] The single-displacement mechanism was consistent with the large solvent isotope effect measured: $V_{max(H_2O)}/V_{max (D_2O)}$ = 2.9 ± 0.1 and the proton inventory study clearly showed that only one proton was transferred during the rate-determining transition state of FucT V.[349] An oxocarbenium ion-like transition state for FucT V was demonstrated in many ways. First, using the deuterated donor substrate GDP-[1-^{2}H]-Fuc, a normal secondary KIE was observed,[349] indicating considerable sp^2 hybridization of the anomeric center at the transition state. Second, the fluorinated donor analogue, GDP-2F-Fuc was synthesized (**Figure 6**) and shown to act as a competitive inhibitor (K_i = 4.2 µmol l^{-1}), rather than a slow substrate or inactivator for the enzyme, consistent with accumulation of positive charge at the transition state.[346,350] A number of deoxygenated Type I and Type II acceptor molecules were synthesized and used to probe the contributions of each hydroxyl to the binding of the acceptor in both FucT V[351] and hpFucT[345] by measuring the relative transfer rate compared with those for the natural acceptor substrates. Interestingly, both studies revealed that the 6-hydroxyl on the galactose in the Type I and Type II structures is crucial for proper binding and transfer. The importance of the 6-OH of galactose in the acceptor has been demonstrated in many other FucTs, including the human milk α-1,3 and α-3,4 FucTs, indicating that the active sites share structural similarities.[352]

The key questions remaining to be answered regarding the mechanism of FucTs are: (1) the identity of the basic residue deprotonating the nucleophile of the incoming acceptor; (2) the mechanism by which FucTs activate the phosphate leaving group of the donor substrates. Earlier pH profile studies on FucT V revealed that a residue with pK_a = 4.1 is essential for catalysis, suggesting that this residue could be a carboxylic acid.[349] However, the definite identity of this key catalytic residue remained elusive until the first crystal structure of hpFucT was obtained in 2007.[345,353,354] The overall structure of hpFucT was found to adopt a GT-B fold with two Rossmann-fold domains, providing the binding sites for the donor and acceptor, respectively. Although sequence alignment of hpFucT with other GT-B glycosyltransferases revealed little homology, significant

similarities have been found in its three-dimensional structure with those of BGT (GT 63), N-acetylglucosaminyltransferase MurG (GT 28) and TDP-vancosaminyltransferase GtfD (GT 1), implying that these GT-B enzymes share a common ancient ancestor. The identity of the general base residue in BGT is well-established,[37,355] thus by superimposing its structure on that of hpFucT, Glu-95 (hpFucT numbering) was identified as the candidate, consistent with earlier pH profile analyses. The crystal structure of hpFucT with intact donor GDP-Fuc bound also clearly showed that this residue was very close to the anomeric center of GDP-Fuc, thus is well suited for this role. Subsequent mutagenesis of this residue yielded mutants E95A and E95D with no detectable activity, fully consistent with its function as the catalytic base. Two other critical residues Arg195 and Lys250 were also identified from this structure. These two basic side chains coordinate the phosphate leaving group and can be viewed as a means of facilitating glycosidic bond cleavage. This was initially confusing since earlier studies showed that both the hpFucT and the highly related FucT V strictly require a metal cofactor for activity, preferentially Mn^{2+}, presumably to chelate the phosphate leaving group and facilitate its departure.[349] Supporting evidence included a 10-fold acceleration of the nonenzymatic fucosyl transfer from GDP to water in the presence of $10\,mmol\,l^{-1}$ Mn^{2+} and the inhibitory tendency of several nucleoside phosphate ligands toward FucT V with $K_i\,(GTP) \approx K_i\,(GDP) < K_i\,(GMP) << K_i\,(G)$ in the case of FucT V.[349] This apparently contradictory scenario has in fact arisen in many other GT-B glycosyltransferases with the conclusion being that the metal cofactor simply facilitates the departure of the phosphate leaving group from the active site, but is not directly involved in the glycosidic bond cleavage step.

8.11.4.1.2   *Sialyltransferases*

The sialyltransferases are another representative group of inverting glycosyltransferases, which catalyze the transfer of a sialic acid moiety from CMP-Neu5Ac to various acceptors[356] (**Figure 1**). Four major types of glycosidic bonds can be formed, namely, Neu5Acα2-6Gal, Neu5Acα2-3Gal, Neu5Acα2-6GalNAc, and Neu5Acα2-8Neu5Ac. As discussed in the sialidase section, terminal sialic acid residue-containing glycoconjugates participate in many important biological recognition processes in higher organisms, while sialylated structures in lower organisms, such as bacteria, are usually involved in virulence. For example, by decorating their cell surfaces with human-like sialylated conjugates, many pathogenic bacteria can escape detection by the immune system.[357] Therefore, blocking the biosynthesis of these sialylated glycans represents a promising therapeutic strategy and detailed structure–function studies of sialyltransferases will undoubtedly benefit the development of such inhibitors. In the current CAZy classification of glycosyltransferases, five families are found to contain sialyltransferases, namely, GT-29, GT-38, GT-42, GT-52, and GT-80. Despite intense interest in these enzymes, only four crystal structures of sialyltransferases have been published to-date, presumably due to the difficulties of handling these membrane-associated proteins. Of the four enzymes with available structures, α-2,3-sialyltransferase (CstI)[358] and α-2,3/α-2,8-sialyltransferase (CstII)[11] from *Campylobacter jejuni* are within GT-42, while *Pasteurella multocida* α-2,3-sialyltransferase (Δ24PmST1)[359,360] and *Photobacterium* sp. JT-ISH-224 α-2,6-sialyltransferase (Δ16psp26ST)[361] are from GT-80. Inspection of their structures and subsequent kinetic characterization all support a single-displacement mechanism for this important class of enzymes, as was seen in the case of FucTs.

The bifunctional CstII transfers a sialic acid residue to the 3-position of a β-galactoside or the 8-position of an α-2,3-sialyl-galactoside acceptor and was the first sialyltransferase to be structurally characterized.[11] Its structure is generally that of a GT-A enzyme, with a Rossmann fold at the N-terminus, followed by a small, lid-like domain, though significant differences exist between the CstII structure and the canonical GT-A fold in terms of both the organization of the secondary structural motifs and the donor binding site. This has led some to propose it as a different fold. A single-step displacement mechanism involving an oxocarbenium ion-like transition state is again supported by its inability to hydrolyze or transfer the CMP-3FNeu5Ac (**Figure 6**). His188 is found to be well positioned to function as the general base catalyst and indeed, mutation of H188 to Ala abolished essentially all enzymatic activity, further supporting its role. In fact, His residues are found to function as the base, rather than Asp/Glu, in a number of GT-B inverting glycosyltransferases.[22] Interestingly, unlike most GT-A glycosyltransferases, CstII contains neither a DxD motif nor a divalent metal to facilitate the departure of the phosphate leaving group. Rather, two highly conserved tyrosine residues, Y156 and Y162 form hydrogen bonds with the oxygen in the CMP and stabilize its negative charge. This has been observed in many metal ion-independent GT-B inverting glycosyltransferases (as is seen in the following) and indicates the convergence of

these two superfamilies. Similar mechanistic information can be deduced from the structure of CstI, another GT-42 sialyltransferase.[358] However, the differences in the acceptor specificity of CstI and CstII can be attributed to the different flexible lid domains. In contrast, Δ24PmST1 and Δ16psp26ST from GT-80 are found to possess a typical GT-B overall fold. Each of them contains two $\alpha/\beta/\alpha$ Rossmann domains with the active site lying in the deep cleft between the two domains. The conformation of the bound substrate analogue, CMP-3F-NeuAc, was found to be $^2C_5$,[360] therefore, the authors proposed that significant conformational changes are needed to achieve the oxocarbenium ion-like transition state. Although many of the interactions between substrate and enzyme are different, the essence of the mechanism of GT-A and GT-B sialyltransferase remains very similar, as is apparent from the structures. Asp141 (of Δ24PmST1) and Asp232 (of Δ16psp26ST) were identified as the general base residues.[360,361] His311 and Ser356 of Δ24PmST1 were found to be hydrogen bonded to the CMP moiety and are believed to facilitate its departure, much as seen in the FucTs and CstII. This use of polar residues to stabilize the phosphate leaving group seems to be a general phenomenon for GT-B glycosyltransferases and some GT-A metal ion-independent enzymes. In contrast, the majority of GT-A glycosyltransferases, which possess a DxD motif, usually employ a metal cofactor to facilitate the departure of the nucleoside phosphate group.

8.11.4.2 Retaining Glycosyltransferases: Galactosyltransferases

Chemical intuition would suggest that the mechanism of retaining glycosyltransferases should be similar to that of retaining glycosidases. However, rigorous investigations on a number of retaining glycosyltransferases have cast serious doubts on the originally preferred double-displacement mechanism, since the presence and the identity of the catalytic nucleophile residue cannot be convincingly demonstrated despite numerous mechanistic and structural studies. Studies aimed at unraveling this mechanism are illustrated through the two examples as follows.

All the galactosyltransferases (GalTs) catalyze essentially the same reaction, the transfer of a galactose residue from the donor substrate, UDP-α-D-galactose, to an acceptor with the formation of either an α or a β-glycosidic linkage to a variety of different acceptor hydroxyls.[362,363] Despite the fact that they catalyze the same reaction, little sequence homology can be found between prokaryotic and eukaryotic GalTs. Most of the eukaryotic GalTs are located in the Golgi apparatus and are type II transmembrane proteins. They are usually composed of an N-terminal cytoplasmic tail followed by a transmembrane domain, a stem region, and finally the CD at the C terminus. The prokaryotic GalTs, by contrast, lack a general topology. In the light of the fact that all GalTs utilize the same donor and some of them have similar acceptor specificity, sensitive computational methods have been applied in an attempt to identify sequence signatures and conserved regions for this group of enzymes.[362] Two retaining GalTs, LgtC from *Neisseria meningitidis* and bovine α-1,3-GalT are the subject of discussion here owing to the extensive mechanistic information available on these two enzymes.

LgtC is an α-1,4-GalT, transfering a galactose to the terminal lactose moiety of lipooligosaccharide (LOS) with retention of anomeric configuration.[364] The LOSs so formed are the major glycolipids found on the cell surface of *Neisseria meningitidis* along with a number of other related mucosal pathogens, leading to the speculation that inhibitors of LgtC could be used as novel antibiotics.

LgtC has been assigned to GT 8 according to Henrissat's classification of glycosyltransferases. Kinetic analysis revealed that LgtC has a reasonably strict donor substrate specificity. Inversion of the configuration of just one hydroxyl in the donor UDP-Gal to give UDP-glucose reduces the second-order rate constant by 200-fold.[365] Interestingly, it has been shown that both α-galactosyl fluoride and 2,4-dinitrophenyl β-D-galactoside also function as donor substrates for LgtC in the presence of UDP, albeit at a lower rate.[366,367] The natural acceptor for LgtC is the lactose element of LOS, but lactose itself has a rather high K_m value. In order to simplify the kinetic assays, a number of galactose and lactose analogues were therefore evaluated as acceptors. Interestingly, anomeric substitutions were shown to lower the K_m values and indeed, by incorporation of aromatic or even aliphatic handles onto a series of acceptor sugars, it proved to be possible to synthesize a variety of glycosidic linkages using a 'substrate engineering' approach.[368] Detailed kinetic analysis demonstrated that this two-substrate enzyme follows an ordered bi–bi mechanism in which UDP-Gal binds first followed by the acceptor. Upon completion of the transfer, the oligosaccharide product is released first, followed by UDP.[365]

The crystal structure of LgtC complexed with the inert donor UDP-2F-Gal and an acceptor analogue: 4′-deoxy-lactose was the first published structure of a retaining glycosyltransferase and remains one of the few ternary complexes.[365,369] The overall topology of LgtC corresponds to that of a GT-A fold, with distinct donor and acceptor binding sites. The donor binding site, as revealed by the complex of LgtC with the inert donor analogue UDP-2F-Gal, is located very deep inside the protein while the acceptor binding site is more solvent accessible, and is only fully formed upon binding of the UDP sugar. This could be viewed as a strategy to minimize unwanted hydrolysis of the donor substrate. As was seen with many GT-A glycosyltransferases, a highly conserved 'DxD' motif is found within the active site of LgtC and binds a divalent Mn^{2+} ion. Besides being coordinated with the 'DxD' motif, this well ordered Mn^{2+} ion also interacts with two phosphate oxygens of the UDP moiety of the donor, presumably to assist the departure of the diphosphate leaving group. On the basis of this structure, the only plausible candidate for the general base catalyst to deprotonate the acceptor hydroxyl is an oxygen atom of the phosphate leaving group. Inspection of the crystal structure for a suitably positioned nucleophile residue on the right trajectory to attack the anomeric center of UDP-Gal, surprisingly, revealed the most probable residue to be Gln189. However, the relatively high residual activity of the mutant Q189A (3% of wild-type enzyme activity) rendered the involvement of this amide as the catalytic nucleophile unlikely. Similarly, the possibility that the 6′-OH of the acceptor lactose could act as the catalytic nucleophile was ruled out when the putative intermediate galactosyl β-1,6-lactose did not function as a substrate.[365,369] Various approaches, which had previously proved to be successful in trapping covalent glycosyl–enzyme intermediates on retaining glycosidases,[370,371] were employed in similar attempts on LgtC, but to no avail.[365] The 2-fluorosugar approach that is so successful with retaining glycosidases is not useful with glycosyltransferases since the success of this approach with glycosidases lies in the ability to modulate the leaving group ability of the aglycone such that the deglycosylation step becomes rate limiting. Owing to the strict donor substrate requirements of glycosyl-transferase, this strategy is not possible. Thus, 2-fluorosugar analogues of the donor act as competitive inhibitors.[372,373] In order to further probe the role of Gln189 in catalysis, a Q189E mutant of LgtC was constructed in the hope that by making it a stronger nucleophile for the glycosylation step, the intermediate might accumulate, if the enzyme adopts the double-displacement mechanism. Results were at first consistent with this since a catalytically competent glycosyl–enzyme intermediate was indeed trapped.[374] However, identification of the site of attachment by peptide mapping revealed that the galactosyl residue was surprisingly covalently linked not to Q189, but to D190, a residue 8.9 Å away from the anomeric center of the donor substrate. If D190 serves as the catalytic nucleophile, a significant conformational change must occur to position it correctly with respect to the donor substrate, implying a very plastic active site.

Similarly puzzling mechanistic information was gleaned from studies on another retaining GalT: the GT 6 bovine α-(1,3)-GalT (α3GalT). This enzyme is expressed in many mammalian species, but not in primates such as humans.[375] It catalyzes the transfer of a galactosyl residue from UDP-Gal to a LacNAc moiety at the nonreducing end of glycoconjugates.[376] Earlier crystallographic studies on this enzyme both in the absence and the presence of substrate UDP-Gal revealed a GT-A fold, similar to that of LgtC.[375] Unexpectedly, in the structure of the complex of α3GalT with UDP-Gal, the authors apparently saw extended electron density from the suitably positioned residue E317 to the anomeric center of the bound galactose residue in the donor site. As E317 is highly conserved in this family, this was interpreted as representing the covalent glycosyl–enzyme intermediate and E317 was proposed to be the catalytic nucleophile. Owing to the limited resolution (2.5 Å) and considerable disorder of this structure, this conclusion was later questioned and re-examined crystallographi-cally at much higher resolution (1.46 Å).[377] Interestingly, in this crystal structure of α3GalT with UDP-Gal, the galactose moiety was not covalently linked to E317, but rather was noncovalently bound as β-galactose, consistent with the low hydrolytic activity of wild-type α3GalT. Residue E317 was believed to assist with proper acceptor binding. This notion was supported by creating α3GalT mutant E317Q, which was subjected to kinetic measurements and structural analysis.[378] The k_{cat} of E317Q for the transfer of galactose to the acceptor lactose and water was reduced by 2400- and 120-fold, respectively. While this represented a significant catalytic contribution, this reduction of activity was much less dramatic than that seen upon mutation of an analogous residue in the glycosidases, which resulted in at least a 10^6-fold decrease in activity.[45,308] Structural analysis of α3GalT E317Q with the acceptor lactose bound clearly showed that the

binding of lactose was perturbed by this mutation, supporting a role for E317 in orienting the acceptor rather than serving as the catalytic nucleophile proposed earlier.

Recently, a panel of α3GalT mutants was constructed in which E317 was changed into alanine, aspartic acid, cysteine, and histidine.[379] While E317D and E317H were expected to retain some enzyme activity due to their nucleophilic character, the relatively high activity of E317A (0.1% activity of wild type) made a role of E317 as the catalytic nucleophile less probable. However, recent evidence in support of the double-displacement mechanism has rekindled the mechanistic debate as follows. It is well known that the activity of the nucleophile or general acid/base mutant of retaining glycosidases with activated substrates can be 'rescued' by the addition of small exogenous anions such as azide and formate.[17] In the case of the nucleophile mutant, the small 'rescue' anion that occupies the cavity resulting from the mutation serves as the new nucleophile and reacts much faster than anything else, yielding a product with anomeric configuration opposite to that of the starting material.[17] Monegal and Planas[380] created the E317A mutant of α3GalT, and subjected this to chemical rescue studies. Interestingly, upon the addition of azide, rescue of activity was indeed observed, with the k_{cat} value increasing by more than 100-fold relative to the transglycosidase activity of the E317A mutant without added azide. The product formed, β-D-galactosyl azide, was isolated confirming reaction with the azide anion. While these observations support a nucleophilic role of E317 for α3GalT, this interpretation must be made with caution since binding of an excellent nucleophile directly adjacent to an electrophilic anomeric center is highly likely to lead to reaction.

LgtC and α3GalT are just two examples of a much larger group of retaining glycosyltransferases possessing GT-A folds. All the available data suggest that they share similar strategies for facilitating the departure of the phosphate leaving group by employing 'DxD' coordinated divalent metals as the Lewis acid catalyst.[22] This has been demonstrated in many cases such as GT6 human α-1,3-N-acetylgalactosaminyltransferase (synthesizing blood group A-antigen),[381] GT6 human α-1,3-galactosyltransferase (synthesizing blood group B-antigen),[381] GT15 α-1,2-mannosyltransferase from *Saccharomyces cerevisiae*[382] and GT64 α-1,4-N-acetylhexosaminyltransferase from *Mus musculus*.[383] However, despite various trapping and kinetic studies, definitive evidence for the nucleophile residue is still elusive. Crystallographic data further complicate this issue as no conserved architectural features on the β-face of the donor substrate have been observed in any GT-A retaining glycosyltransferases, at the location where the nucleophile is expected to be.[22]

A number of mechanistic and structural studies have also been performed on GT-B-type retaining glycosyltransferases. As seen with the GT-B inverting glycosyltransferases, GT-B retaining enzymes utilize positively charged side chains, rather than divalent metal, to facilitate the departure of the phosphate leaving group, as exemplified by GT 4 α-1,3 glucosyltransferase WaaG[384] from *E. coli* and GT 20 α,α-trehalose-phosphate synthase OtsA.[385] However, these studies showed no suitably positioned catalytic nucleophile located on the β-face of the donor substrates,[22] again leading to doubts about the double-displacement mechanism. Several alternative chemical mechanisms have been proposed in order to rationalize the available experimental data.

An '$S_N i$' mechanism for a glycosyltransferase was first proposed during the study of the mechanism of LgtC in order to explain the absence of a proper enzymatic nucleophile residue in the active site.[369] While this mechanism is uncommon in organic chemistry, it does have a few chemical precedents, in each case involving a carbocation transition state.[386] In the area of enzymology, an $S_N i$ mechanism was suggested for glycogen phosphorylase in 1986.[387] Since its proposal for retaining glycosyltransferases,[369] this mechanism has gained a certain level of support in the research community and has been suggested to be operative for several retaining glycosyltransferases, including GT6 α-1,3-galactosyltransferase,[377] GT 8 glycogenin,[388] GT 20 α,α-trehalose-phosphate[385] and GT15 α-1,2-mannosyltransferase.[382] However, unlike other established enzymatic mechanisms, the proposed $S_N i$ mechanism is based on the lack of a suitable enzymatic nucleophile rather than a direct proof. Consequently, significant doubts still exist whether this relatively rare mechanism is actually used by retaining glycosyltransferases.

Another plausible mechanism for retaining glycosyltransferases, interestingly, is the 'Phillips' or 'ion pair' mechanism, which has been largely rejected for retaining glycosidases[243] (**Scheme 10**). The first step of the 'ion pair' mechanism of glycosyltransferases involves the departure of the nucleoside phosphate from the donor substrate, which is facilitated by either a divalent metal cation or by positively charged side residues. The resultant oxocarbenium ion intermediate is stabilized by numerous noncovalent interactions within the enzyme
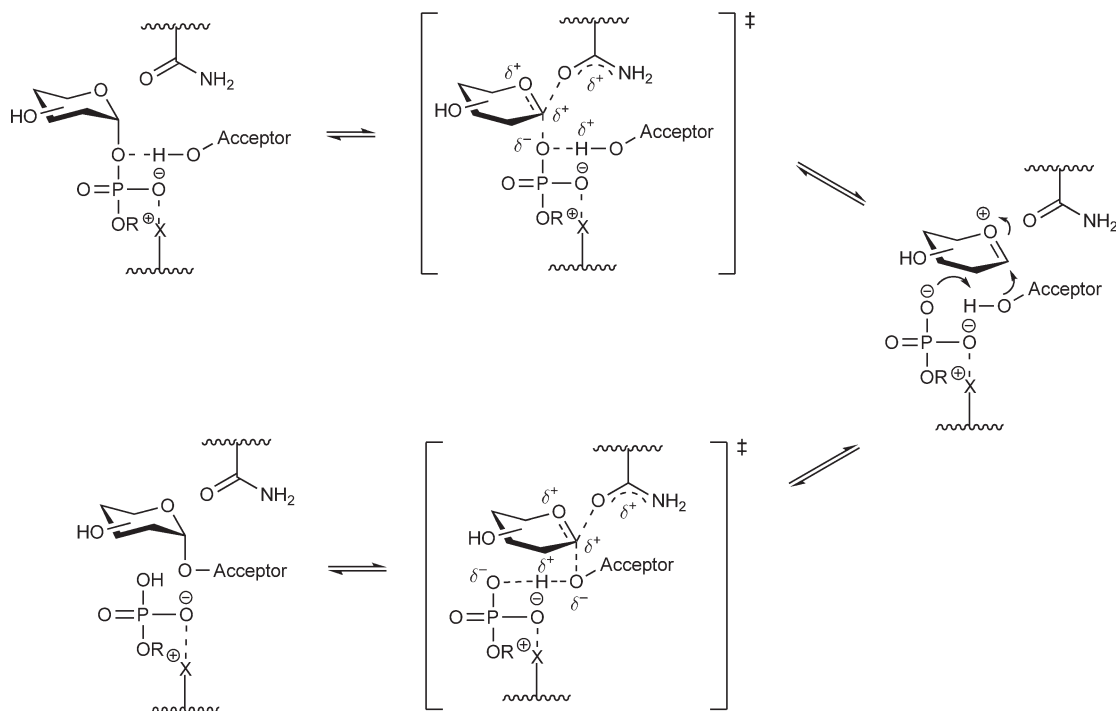
Scheme 10 Proposed 'ion pair' mechanism of retaining glycosyltransferases. R, nucleotide; X, divalent metal ion or charged side chains.

active site as well as by some level of 'back side' stabilization from the amide where present. As it departs, the phosphate leaving group deprotonates the acceptor, which collapses onto the oxocarbenium ion intermediate. This is essentially the mechanism intended by proponents of the $S_N i$ process, but one which respects the need for a true intermediate to avoid violating orbital symmetry principles. However, this raises the question as to why retaining glycosyltransferases do not employ the standard covalent intermediate. One key difference between glycosidases and glycosyltransferases lies in the inherent reactivity of their donor substrates, with a nucleoside phosphate being a much better leaving group than a normal alcohol.[3] As a consequence, the activation barrier for glycosyltransferases is much lower than for glycosidases, thus reaction involving an ion pair intermediate becomes feasible. In contrast, for glycosidases to perform efficient catalysis, the transition state must be stabilized to a much greater extent and this requires a more stabilized 'covalent' intermediate, according to the Hammond postulate.

References

1. A. C. R. Varki; J. Esko; H. Freeze; G. Hart; J. Marth, *Essentials of Glycobiology*; Cold Spring Harbour Laboratory Press: Woodbury NY, 1999.
2. R. Wolfenden, *Chem. Rev.* **2006**, *106*, 3379–3396.
3. R. Wolfenden; X. D. Lu; G. Young, *J. Am. Chem. Soc.* **1998**, *120*, 6814–6815.
4. M. L. Sinnott, *Chem. Rev.* **1990**, *90*, 1171–1202.
5. G. S. Davies; M. L. Sinnott; S. G. Withers, Glycosyl Transfer. In *Comprehensive Biological Catalysis*; M. L. Sinnott, Ed.; Academic Press: New York, 1998; Vol. 1, pp 119–208.
6. J. C. M. Uitdehaag; R. Mosi; K. H. Kalk; B. A. van der Veen; L. Dijkhuizen; S. G. Withers; B. W. Dijkstra, *Nat. Struct. Biol.* **1999**, *6*, 432–436.
7. L. L. Lairson; S. G. Withers, *Chem. Commun.* **2004**, *20*, 2243–2248.
8. I. P. Street; C. R. Armstrong; S. G. Withers, *Biochemistry* **1986**, *25*, 6021–6027.
9. J. A. Campbell; G. J. Davies; V. Bulone; B. Henrissat, *Biochem. J.* **1997**, *326*, 929–939.
10. R. Caputto; L. F. Leloir; C. E. Cardini; A. C. Paladini, *J. Biol. Chem.* **1950**, *184*, 333–350.

11. C. P. C. Chiu; A. G. Watts; L. L. Lairson; M. Gilbert; D. Lim; W. W. Wakarchuk; S. G. Withers; N. C. J. Strynadka, *Nat. Struct. Mol. Biol.* **2004**, *11*, 163–170.
12. A. Wright; M. Dankert; P. Fennessey; P. W. Robbins, *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *57*, 1798–1803.
13. B. Schuman; J. A. Alfaro; S. V. Evans; *Bioactive Conformation I*; Springer-Verlag Berlin: Berlin, 2007; Vol. 272, pp 217–257.
14. H. D. Ly; S. G. Withers, *Annu. Rev. Biochem.* **1999**, *68*, 487–522.
15. G. Davies; B. Henrissat, *Structure* **1995**, *3*, 853–859.
16. V. L. Yip; S. G. Withers, *Curr. Opin. Chem. Biol.* **2006**, *10*, 147–155.
17. D. L. Zechel; S. G. Withers, *Acc. Chem. Res.* **2000**, *33*, 11–18.
18. A. Vasella; G. J. Davies; M. Bohm, *Curr. Opin. Chem. Biol.* **2002**, *6*, 619–629.
19. D. L. Zechel; S. G. Withers, *Curr. Opin. Chem. Biol.* **2001**, *5*, 643–649.
20. C. S. Rye; S. G. Withers, *Curr. Opin. Chem. Biol.* **2000**, *4*, 573–580.
21. J. McCarter; S. G. Withers, *Curr. Opin. Struct. Biol.* **1994**, *4*, 885–892.
22. L. L. Lairson; B. Henrissat; G. J. Davies; S. G. Withers, *Annu. Rev. Biochem.* **2008**, *77*, 521–555.
23. B. Henrissat, *Biochem. J.* **1991**, *280*, 309–316.
24. B. Henrissat; A. Bairoch, *Biochem. J.* **1993**, *293* (Pt 3), 781–788.
25. J. Gebler; N. R. Gilkes; M. Claeyssens; D. B. Wilson; P. Beguin; W. W. Wakarchuk; D. G. Kilburn; R. C. Miller, Jr.; R. A. Warren; S. G. Withers, *J. Biol. Chem.* **1992**, *267*, 12559–12561.
26. M. Claeyssens; B. Henrissat, *Protein Sci.* **1992**, *1*, 1293–1297.
27. B. Henrissat; G. Davies, *Curr. Opin. Struct. Biol.* **1997**, *7*, 637–644.
28. B. Henrissat; I. Callebaut; S. Fabrega; P. Lehn; J. P. Mornon; G. Davies, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7090–7094.
29. B. Henrissat; G. J. Davies, *Plant Physiol.* **2000**, *124*, 1515–1519.
30. B. P. Rempel; L. A. Clarke; S. G. Withers, *Mol. Genet. Metab.* **2005**, *85*, 28–37.
31. J. Jenkins; L. Lo Leggio; G. Harris; R. Pickersgill, *FEBS Lett.* **1995**, *362*, 281–285.
32. M. R. Stam; E. G. Danchin; C. Rancurel; P. M. Coutinho; B. Henrissat, *Protein Eng. Des. Sel.* **2006**, *19*, 555–562.
33. P. M. Coutinho; E. Deleury; G. J. Davies; B. Henrissat, *J. Mol. Biol.* **2003**, *328*, 307–317.
34. Y. Bourne; B. Henrissat, *Curr. Opin. Struct. Biol.* **2001**, *11*, 593–600.
35. G. J. Davies; T. M. Gloster; B. Henrissat, *Curr. Opin. Struct. Biol.* **2005**, *15*, 637–645.
36. J. Rouvinen; T. Bergfors; T. Teeri; J. K. C. Knowles; T. A. Jones, *Science* **1990**, *249*, 380–386.
37. A. Vrielink; W. Ruger; H. P. Driessen; P. S. Freemont, *EMBO J.* **1994**, *13*, 3413–3422.
38. S. J. Charnock; G. J. Davies, *Biochemistry* **1999**, *38*, 6380–6385.
39. Y. N. Hu; S. Walker, *Chem. Biol.* **2002**, *9*, 1287–1296.
40. J. Liu; A. Mushegian, *Protein Sci.* **2003**, *12*, 1418–1431.
41. M. Igura; N. Maita; J. Kamishikiryo; M. Yamada; T. Obita; K. Maenaka; D. Kohda, *EMBO J.* **2008**, *27*, 234–243.
42. A. L. Lovering; L. H. de Castro; D. Lim; N. C. J. Strynadka, *Science* **2007**, *315*, 1402–1405.
43. Y. Q. Yuan; D. Barrett; Y. Zhang; D. Kahne; P. Sliz; S. Walker, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5348–5353.
44. D. E. J. Koshland, *Biol. Rev.* **1953**, *28*, 416–436.
45. Q. Wang; R. W. Graham; D. Trimbur; R. A. J. Warren; S. G. Withers, *J. Am. Chem. Soc.* **1994**, *116*, 11594–11595.
46. V. L. Y. Yip; A. Varrot; G. J. Davies; S. S. Rajan; X. J. Yang; J. Thompson; W. F. Anderson; S. G. Withers, *J. Am. Chem. Soc.* **2004**, *126*, 8354–8355.
47. B. L. Mark; D. J. Vocadlo; S. Knapp; B. L. Triggs-Raine; S. G. Withers; M. N. G. James, *J. Biol. Chem.* **2001**, *276*, 10330–10337.
48. D. Piszkiewicz; T. C. Bruice, *J. Am. Chem. Soc.* **1968**, *90*, 2156–2163.
49. S. Drouillard; S. Armand; G. J. Davies; C. E. Vorgias; B. Henrissat, *Biochem. J.* **1997**, *328*, 945–949.
50. D. M. F. van Aalten; D. Komander; B. Synstad; S. Gaseidnes; M. G. Peter; V. G. H. Eijsink, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8979–8984.
51. B. L. Mark; D. J. Vocadlo; D. Zhao; S. Knapp; S. G. Withers; M. N. G. James, *J. Biol. Chem.* **2001**, *45*, 42131–42137.
52. A. Buschiazzo; O. D. Campetella; A. C. C. Frasch, *Glycobiology* **1997**, *7*, 1167–1173.
53. J. N. Watson; V. Dookhun; T. J. Borgford; A. J. Bennet, *Biochemistry* **2003**, *42*, 12682–12690.
54. A. G. Watts; I. Damager; M. L. Amaya; A. Buschiazzo; P. Alzari; A. C. Frasch; S. G. Withers, *J. Am. Chem. Soc.* **2003**, *125*, 7532–7533.
55. W. P. Burmeister; S. Cottaz; H. Driguez; R. Iori; S. Palmieri; B. Henrissat, *Structure* **1997**, *5*, 663–675.
56. M. Shikita; J. W. Fahey; T. R. Golden; W. D. Holtzclaw; P. Talalay, *Biochem. J.* **1999**, *341*, 725–732.
57. S. Cottaz; B. Henrissat; H. Driguez, *Biochemistry* **1996**, *35*, 15256–15259.
58. N. Tani; M. Ohtsuru; T. Hata, *Agric. Biol. Chem.* **1974**, *38*, 1623–1630.
59. J. Xue; M. Lenman; A. Falk; L. Rask, *Plant Mol. Biol.* **1992**, *18*, 387–398.
60. W. P. Burmeister; S. Cottaz; P. Rollin; A. Vasella; B. Henrissat, *J. Biol. Chem.* **2000**, *275*, 39385–39393.
61. W. Nerinckx; T. Desmet; K. Piens; M. Claeyssens, *FEBS Lett.* **2005**, *579*, 302–312.
62. F. Febbraio; R. Barone; S. D'Auria; M. Rossi; R. Nucci; G. Piccialli; L. DeNapoli; S. Orru; P. Pucci, *Biochemistry* **1997**, *36*, 3068–3075.
63. S. G. Withers; K. Rupitz; I. P. Street, *J. Biol. Chem.* **1988**, *263*, 7929–7932.
64. S. G. Withers; I. P. Street; P. Bird; D. H. Dolphin, *J. Am. Chem. Soc.* **1987**, *109*, 7530–7531.
65. I. P. Street; J. B. Kempton; S. G. Withers, *Biochemistry* **1992**, *31*, 9970–9978.
66. I. P. Street; K. Rupitz; S. G. Withers, *Biochemistry* **1989**, *28*, 1581–1587.
67. S. G. Withers; I. P. Street; M. D. Percival, *ACS Symp. Ser.* **1988**, *374*, 59–77.
68. S. G. Withers; I. P. Street, *J. Am. Chem. Soc.* **1988**, *110*, 8551–8553.
69. J. D. McCarter; D. L. Burgoyne; S. Miao; S. Zhang; J. W. Callahan; S. G. Withers, *J. Biol. Chem.* **1997**, *272*, 396–400.
70. D. J. Vocadlo; L. F. Mackenzie; S. He; G. J. Zeikus; S. G. Withers, *Biochem. J.* **1998**, *335*, 449–455.
71. J. K. Yang; H. J. Yoon; H. J. Ahn; B. I. Lee; J. D. Pedelacq; E. C. Liong; J. Berendzen; M. Laivenieks; C. Vieille; G. J. Zeikus; D. J. Vocadlo; S. G. Withers; S. W. Suh, *J. Mol. Biol.* **2004**, *335*, 155–165.
72. S. Miao; L. Ziser; R. Aebersold; S. G. Withers, *Biochemistry* **1994**, *33*, 7027–7032.

73. D. K. Y. Poon; M. L. Ludwiczek; M. Schubert; E. M. Kwan; S. G. Withers; L. P. McIntosh, *Biochemistry* **2007**, *46*, 1759–1770.
74. G. Hommalai; P. Chaiyen; J. Svasti, *Arch. Biochem. Biophys.* **2005**, *442*, 11–20.
75. M. Hrmova; J. N. Varghese; R. De Gori; B. J. Smith; H. Driguez; G. B. Fincher, *Structure* **2001**, *9*, 1005–1016.
76. J. E. Blanchard; L. Gal; S. M. He; J. Foisy; R. A. J. Warren; S. G. Withers, *Carbohydr. Res.* **2001**, *333*, 7–17.
77. C. Braun; G. Brayer; S. G. Withers, *J. Biol. Chem.* **1995**, *118*, 241.
78. J. D. McCarter; S. G. Withers, *J. Am. Chem. Soc.* **1996**, *118*, 241–242.
79. J. D. McCarter; S. G. Withers, *J. Biol. Chem.* **1996**, *271*, 6889–6894.
80. H. D. Ly; S. Howard; K. Shum; S. M. He; A. Zhu; S. G. Withers, *Carbohydr. Res.* **2000**, *329*, 539–547.
81. C. A. Tarling; S. M. He; G. Sulzenbacher; C. Bignon; Y. Bourne; B. Henrissat; S. G. Withers, *J. Biol. Chem.* **2003**, *278*, 47394–47399.
82. A. L. Lovering; S. S. Lee; Y.-W. Kim; S. G. Withers, *J. Biol. Chem.* **2005**, *280*, 2105–2115.
83. S. Numao; S. M. He; G. Evjen; S. Howard; O. K. Tollersrud; S. G. Withers, *FEBS Lett.* **2000**, *484*, 175–178.
84. S. Howard; S. He; S. G. Withers, *J. Biol. Chem.* **1998**, *273*, 2067.
85. O. Hekmat; C. Florizone; Y. W. Kim; L. D. Eltis; R. A. J. Warren; S. G. Withers, *ChemBioChem* **2007**, *8*, 2125–2132.
86. O. Hekmat; Y. W. Kim; S. J. Williams; S. M. He; S. G. Withers, *J. Biol. Chem.* **2005**, *280*, 35126–35135.
87. T. E. Nelson; J. Larner, *Biochim. Biophys. Acta* **1970**, *198*, 538.
88. X. Guo; W. Laver; E. Vimr; M. Sinnott, *J. Am. Chem. Soc.* **1994**, *116*, 5572.
89. C. Braun; A. Meinke; L. Ziser; S. G. Withers, *Anal. Biochem.* **1993**, *212*, 259.
90. S. S. Rajan; X. Yang; F. Collart; V. L. Y. Yip; S. G. Withers; A. Varrot; J. Thompson; G. J. Davies; W. F. Anderson, *Structure* **2004**, *12*, 1619–1629.
91. J. Gebler; N. R. Gilkes; M. Claeyssens; D. B. Wilson; P. Beguin; W. W. Wakarchuk; D. G. Kilburn; R. C. Miller, Jr.; R. A. J. Warren; S. G. Withers, *J. Biol. Chem.* **1992**, *267*, 12559–12561.
92. S. G. Withers; D. Dombroski; L. A. Berven; D. G. Kilburn; J. R. C. Miller; R. A. J. Warren; N. R. Gilkes, *Biochem. Biophys. Res. Commun.* **1986**, *139*, 487–494.
93. V. L. Y. Yip; J. Thompson; S. G. Withers, *Biochemistry* **2007**, *46*, 9840–9852.
94. M. M. Palcic; C. H. Scaman; A. Otter; A. Szpacenko; A. Romaniouk; Y. X. Li; I. K. Vijay, *Glycoconjugate J.* **1999**, *16*, 351–355.
95. J. B. Kempton; S. G. Withers, *Biochemistry* **1992**, *31*, 9961–9969.
96. S. M. Cutfield; G. J. Davies; G. Murshudov; B. F. Anderson; P. C. E. Moody; P. A. Sullivan; J. F. Cutfield, *J. Mol. Biol.* **1999**, *294*, 771–783.
97. M. Hrmova; R. D. Gori; B. J. Smith; J. K. Fairweather; H. Driguez; J. N. Varghese; G. B. Fincher, *Plant Cell* **2002**, *14*, 1033–1052.
98. W. Nerinckx; T. Desmet; M. Claeyssens, *FEBS Lett.* **2003**, *538*, 1–7.
99. A. Varrot; M. Schulein; S. Fruchard; H. Driguez; G. J. Davies, *Acta Crystallogr.* **2001**, *D57*, 1739–1742.
100. D. L. Zechel; S. P. Reid; D. Stoll; O. Nashiru; R. A. J. Warren; S. G. Withers, *Biochemistry* **2003**, *42*, 7195–7204.
101. D. J. Vocadlo; J. Wicki; K. Rupitz; S. G. Withers, *Biochemistry* **2002**, *41*, 9727–9735.
102. T. Bravman; G. Zolotnitsky; V. Belakhov; G. Shoham; B. Henrissat; T. Baasov; Y. Shoham, *Biochemistry* **2003**, *42*, 10528–10536.
103. M. Vallmitjana; M. Ferrer-Navarro; R. Planell; M. Abel; C. Ausin; E. Querol; A. Planas; J.-A. Perez-Pons, *Biochemistry* **2001**, *40*, 5975–5982.
104. M. W. Bauer; R. M. Kelly, *Biochemistry* **1998**, *37*, 17170–17178.
105. D. Tull; S. G. Withers, *Biochemistry* **1994**, *33*, 6363–6370.
106. M. P. Dale; H. E. Ensley; K. Kern; K. A. R. Sastry; L. D. Byers, *Biochemistry* **1985**, *24*, 3530–3539.
107. D. Shallom; M. Leon; T. Bravman; A. Ben-David; G. Zaide; V. Belakhov; G. Shoham; D. Schomburg; T. Baasov; Y. Shoham, *Biochemistry* **2005**, *44*, 387–397.
108. H. Kersters-Hilderson; E. van Doorslaer; C. K. de Bruyne, *Carbohydr. Res.* **1978**, *65*, 219–227.
109. S. Rosenberg; J. F. Kirsch, *Biochemistry* **1981**, *20*, 3189–3196.
110. S. Rosenberg; J. F. Kirsch, *Biochemisty* **1981**, *20*, 3196–3204.
111. S. Rosenberg; J. F. Kirsch, *Fed. Proc.* **1978**, *37*, 1296.
112. E. A. Taylor Ringia; V. L. Schramm, *Curr. Top. Med. Chem.* **2005**, *5*, 1237–1258.
113. P. C. Kline; V. L. Schramm, *Biochemistry* **1993**, *32*, 13212–13219.
114. A. Lewandowicz; V. L. Schramm, *Biochemistry* **2004**, *43*, 1458–1468.
115. P. J. Berti; K. S. E. Tanaka, *Adv. Phys. Org. Chem.* **2002**, *37*, 239–314.
116. M. Bruner; B. A. Horenstein, *Biochemistry* **1998**, *37*, 289–297.
117. D. Indurugalla; A. J. Bennet, *J. Am. Chem. Soc.* **2001**, *123*, 10889–10898.
118. X. C. Huang; K. S. E. Tanaka; A. J. Bennet, *J. Am. Chem. Soc.* **1997**, *119*, 11147–11154.
119. J. S. Yang; S. Schenkman; B. A. Horenstein, *Biochemistry* **2000**, *39*, 5902–5910.
120. J. K. Lee; A. D. Bain; P. J. Berti, *J. Am. Chem. Soc.* **2004**, *126*, 3769–3776.
121. L. Hosie; M. L. Sinnott, *Biochem. J.* **1985**, *226*, 437–446.
122. E. Vandoorslaer; O. Vanopstal; H. Kerstershilderson; C. K. Debruyne, *Bioorg. Chem.* **1984**, *12*, 158–169.
123. A. H. O'Donnell; X. J. Yao; L. D. Byers, *Biochim. Biophys. Acta* **2004**, *1703*, 63–67.
124. S. M. Hancock; M. D. Vaughan; S. G. Withers, *Curr. Opin. Chem. Biol.* **2006**, *10*, 509–519.
125. A. Varrot; V. L. Y. Yip; Y. Li; S. S. Rajan; X. Yang; W. F. Anderson; J. Thompson; S. G. Withers; G. J. Davies, *J. Mol. Biol.* **2005**, *346*, 423–435.
126. V. L. Y. Yip; S. G. Withers, *Angew. Chem. Int. Ed.* **2006**, *118*, 6325–6328.
127. V. L. Y. Yip; S. G. Withers, *Biochemistry* **2006**, *45*, 571–580.
128. B. Henrissat; A. Bairoch, *Biochem. J.* **1996**, *316*, 695–696.
129. C. Raasch; W. Streit; J. Schanzer; M. Bibel; U. Gosslar; W. Liebl, *Extremophiles* **2000**, *4*, 189–200.
130. J. Thompson; A. Pikis; S. B. Ruvinov; B. Henrissat; H. Yamamoto; J. Sekiguchi, *J. Biol. Chem.* **1998**, *273*, 27347–27356.
131. J. Thompson; S. Hess; A. Pikis, *J. Biol. Chem.* **2004**, *279*, 1553–1561.
132. J. Thompson; C. R. Gentryweeks; N. Y. Nguyen; J. E. Folk; S. A. Robrish, *J. Bacteriol.* **1995**, *177*, 2505–2512.

133. C. L. Bouma; J. Reizer; A. Reizer; S. A. Robrish; J. Thompson, *J. Bacteriol.* **1997**, *179*, 4129–4137.
134. J. Thompson; S. A. Pobrish; S. Immel; F. W. Lichtenthaler; B. G. Hall; A. Pikis, *J. Biol. Chem.* **2001**, *276*, 37415–37425.
135. J. A. Lodge; T. Maier; W. Liebl; V. Hoffmann; N. Strater, *J. Biol. Chem.* **2003**, *278*, 19151–19158.
136. C. Raasch; M. Armbrecht; W. Streit; B. Hocker; N. Strater; W. Liebl, *FEBS Lett.* **2002**, *517*, 267–271.
137. C. Burstein; A. Kepes, *Biochim. Biophys. Acta* **1971**, *230*, 52–63.
138. Y. Nagao; T. Nakada; M. Imoto; T. Shimamoto; S. Sakai; M. Tsuda; T. Tsuchiya, *Biochem. Biophys. Res. Commun.* **1988**, *151*, 236–241.
139. J. Thompson; S. B. Ruvinov; D. I. Freedberg; B. G. Hall, *J. Bacteriol.* **1999**, *181*, 7339–7345.
140. J. Thompson; S. A. Robrish; A. Pikis; A. Brust; F. W. Lichtenthaler, *Carbohydr. Res.* **2001**, *331*, 149–161.
141. M. Bibel; C. Brettl; U. Gosslar; G. Kriegshauser; W. Liebl, *FEMS Microbiol. Lett.* **1998**, *158*, 9–15.
142. C. Suresh; A. A. Rus'd; M. Kitaoka; K. Hayashi, *FEBS Lett.* **2002**, *517*, 159–162.
143. A. Varrot; H. Yamamoto; J. Sekiguchi; J. Thompson; G. J. Davies, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *55*, 1212–1214.
144. A. Pikis; S. Immel; S. A. Robrish; J. Thompson, *Microbiology SGM* **2002**, *148*, 843–852.
145. S. A. Robrish; H. M. Fales; C. Gentry-Weeks; J. Thompson, *Int. J. Bacteriol.* **1994**, *176*, 3250–3256.
146. R. Schmitt; B. Rotman, *Biochem. Biophys. Res. Commun.* **1966**, *22*, 473–479.
147. P. L. Liljestrom; P. Lijestrom, *Nucleic Acids Res.* **1987**, *15*, 2213–2220.
148. M. Buehner; G. C. Ford; K. W. Olsen; D. Moras; M. G. Rossmann, *J. Mol. Biol.* **1974**, *90*, 25–49.
149. J. P. Ferraz; E. H. Cordes, *J. Am. Chem. Soc.* **1980**, *101*, 1488–1491.
150. T. H. Fife; T. J. Przystas, *J. Am. Chem. Soc.* **1980**, *102*, 292–299.
151. J. L. Jensen; W. P. Jencks, *J. Am. Chem. Soc.* **1979**, *101*, 1476–1488.
152. M. S. Macauley; K. A. Stubbs; D. J. Vocadlo, *J. Am. Chem. Soc.* **2005**, *127*, 17202–17203.
153. N. Cetinbas; M. S. Macauley; K. A. Stubbs; R. Drapala; D. J. Vocadlo, *Biochemistry* **2006**, *45*, 3835–3844.
154. A. A. Narine; J. N. Watson; A. J. Bennet, *Biochemistry* **2006**, *45*, 9319–9326.
155. H. Shen; L. D. Byers, *Biochem. Biophys. Res. Commun.* **2007**, *362*, 717–720.
156. A. Day, Purification and Preliminary Characterization of β-glucosidase from *Alcaligenes faecalis* (ATCC 21400). MSc thesis, University of British Columbia, Vancouver, BC, 1985.
157. R. J. Dennis; E. J. Taylor; M. S. Macauley; K. A. Stubbs; J. P. Turkenburg; S. J. Hart; G. N. Black; D. J. Vocadlo; G. J. Davies, *Nat. Struct. Mol. Biol.* **2006**, *13*, 365–371.
158. S. Newstead; J. N. Watson; T. L. Knoll; A. J. Bennet; G. Taylor, *Biochemistry* **2005**, *44*, 9117–9122.
159. Q. P. Liu; G. Sulzenbacher; H. Yuan; E. P. Bennet; K. Saunders; J. Spence; E. Nudelman; S. B. Levery; T. White; J. M. Neveu; W. S. Lane; Y. Bourne; M. L. Olsson; B. Henrissat; H. Clausen, *Nat. Biotechnol.* **2007**, *25*, 454–464.
160. M. Smar; S. A. Short; R. Wolfenden, *Biochemistry* **1991**, *30*, 7908–7912.
161. K. Hirano; M. Ziak; K. Kamoshita; Y. Sukenaga; S. Kametani; Y. Shiga; J. Roth; H. Akanuma, *Glycobiology* **2000**, *10*, 1283–1289.
162. S. S. Lee; S. Yu; S. G. Withers, *Biochemistry* **2003**, *42*, 13081–13090.
163. S. S. Lee; S. K. Yu; S. G. Withers, *J. Am. Chem. Soc.* **2002**, *124*, 4948–4949.
164. S. S. Lee; S. K. Yu; S. G. Withers, *Biologia* **2005**, *60*, 137–148.
165. S. S. Lee; S. M. He; S. G. Withers, *Biochem. J.* **2001**, *359*, 381–386.
166. S. Yu; K. Bojsen; B. Svensson; J. Marcussen, *Biochim. Biophys. Acta* **1999**, *1433*, 1–15.
167. P. Nyvall; M. Pedersen; L. Kenne; P. Gacesa, *Phytochemistry* **2000**, *54*, 139–145.
168. K.-A. Lee; M.-S. Kim, *Can. J. Microbiol.* **2000**, *46*, 1077–1081.
169. M.-J. Papandreou; R. Barbouche; R. Guieu; M. P. Kieny; E. Fenouillet, *Mol. Pharmacol.* **2002**, *61*, 186–193.
170. I. Przylas; Y. Terada; K. Fujii; T. Takaha; W. Saenger; N. Strater, *Eur. J. Biochem.* **2000**, *267*, 6903–6913.
171. O. M. Viratelle; J. M. Yon, *Biochemistry* **1980**, *19*, 4143–4149.
172. G. Legler, *Adv. Carbohydr. Chem. Biochem.* **1990**, *48*, 319–384.
173. E. J. Hehre; D. S. Genghof; H. Sternlicht; C. F. Brewer, *Biochemistry* **1977**, *16*, 1780–1787.
174. D. F. Wentworth; R. Wolfenden, *Biochemistry* **1974**, *13*, 4715–4720.
175. S. Numao; D. A. Kuntz; S. G. Withers; D. R. Rose, *J. Biol. Chem.* **2003**, *278*, 48074–48083.
176. M. Sorice; A. Longo; T. Garofalo; V. Mattei; R. Misasi; A. Pavan, *Glycoconjugate J.* **2003**, *20*, 63–70.
177. E. G. Bremer; J. Schlessinger; S. Hakomori, *J. Biol. Chem.* **1986**, *261*, 2434–2440.
178. E. G. Bremer; S. Hakomori; D. F. Bowenpope; E. Raines; R. Ross, *J. Biol. Chem.* **1984**, *259*, 6818–6825.
179. H. Clausen; S. Hakomori, *Vox Sang UINIS* **1989**, *56*, 1–20.
180. W. T. J. Morgan; W. M. Watkins, *Glycoconjugate J.* **2000**, *17*, 501–530.
181. F. Yamamoto; H. Clausen; T. White; J. Marken; S. I. Hakomori, *Nature* **1990**, *345*, 229–233.
182. W. T. J. Morgan, Blood Group Specific Mucopolysaccharides. In *Methods in Carbohydrate Chemistry*; R. L. Whistler, Ed.; Academic Press: New York, 1965; Vol. 5, pp 95–98.
183. B. Rigat; W. Wang; A. Leung; D. J. Mahuran, *Biochemistry* **1997**, *36*, 8325–8331.
184. T. Garofalo; M. Sorice; R. Misasi; A. Cinque; V. Mattei; M. Pontieri; M. G. Cifone; A. Pavan, *J. Lipid Res.* **2002**, *43*, 971–978.
185. K. Minoguchi; W. D. Swaim; E. H. Berenstein; R. P. Siraganian, *J. Biol. Chem.* **1994**, *269*, 5249–5254.
186. K. Kasahara; Y. Watanabe; T. Yamamoto; Y. Sanai, *J. Biol. Chem.* **1997**, *272*, 29947–29953.
187. K. Iwabuchi; S. Yamamura; A. Prinetti; K. Handa; S. Hakomori, *FASEB J.* **1998**, *273*, 9130–9138.
188. M. Sorice; T. Garofalo; R. Misasi; V. Dolo; G. Lucania; T. Sansolini; I. Parolini; M. Sargiacomo; M. R. Torrisi; A. Pavan, *Biosci. Rep.* **1999**, *19*, 197–208.
189. M. Sorice; I. Parolini; T. Sansolini; T. Garofalo; V. Dolo; M. Sargiacomo; T. Tai; C. Peschle; M. R. Torrisi; A. Pavan, *J. Lipid Res.* **1997**, *38*, 969–980.
190. B. Pettmann; M. Manthorpe; S. Varon, *J. Neurosci. Res.* **1988**, *20*, 442–450.
191. H. E. Saqr; M. C. Lee; A. M. Burkman; A. J. Yates, *J. Neurosci. Res.* **1995**, *41*, 491–500.
192. R. DeMaria; L. Lenti; F. Malisan; F. dAgostino; B. Tomassini; A. Zeuner; M. R. Rippo; R. Testi, *Science* **1997**, *277*, 1652–1655.
193. W. Malorni; A. M. Giammarioli; T. Garofalo; M. Sorice, *Apoptosis* **2007**, *12*, 941–949.

194. T. Garofalo; R. Misasi; V. Mattei; A. M. Giammarioli; W. Malorni; G. M. Pontieri; A. Pavan; M. Sorice, *J. Biol. Chem.* **2003**, *278*, 8309–8315.
195. A. M. Giammarioli; T. Garofalo; M. Sorice; R. Misasi; L. Gambardella; R. Gradini; S. Fais; A. Pavan; W. Malorni, *FEBS Lett.* **2001**, *508*, 494.
196. A. M. Giammarioli; T. Garofalo; M. Sorice; R. Misasi; L. Gambardella; R. Gradini; S. Fais; A. Pavan; W. Malorni, *FEBS Lett.* **2001**, *506*, 45–50.
197. C. Cametti; F. DeLuca; M. A. Macri; B. Maraviglia; R. Misasi; M. Sorice; A. Pavan; T. Garofalo; G. M. Pontieri; F. Bordi; G. Zimatore, *Colloids Surf. B* **1996**, *7*, 39–46.
198. E. G. Bremer; S. Hakomori, *Biochem. Biophys. Res. Commun.* **1982**, *106*, 711–718.
199. M. Hirayama; Y. Kitagawa; S. Yamamoto; A. Tokuda; T. Mutoh; T. Hamano; T. Aita; M. Kuriyama, *J. Neurol. Sci.* **1997**, *152*, 99–101.
200. T. Mutoh; A. Tokuda; T. Miyadai; M. Hamaguchi; N. Fujiki, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5087–5091.
201. H. Gouy; P. Deterre; P. Debre; G. Bismuth, *J. Immunol.* **1994**, *152*, 3271–3281.
202. B. Ravichandra; P. G. Joshi, *J. Neurochem.* **1999**, *73*, 557–567.
203. B. Ravichandra; P. G. Joshi, *Biophys. Chem.* **1999**, *76*, 117–132.
204. N. E. Zachara; G. W. Hart, *Biochim. Biophys. Acta* **2006**, *1761*, 599–617.
205. G. W. Hart; C. Slawson; M. Housley; Q. Zeidan; S. Whelan; W. Dias; W. Cheung; K. Sakabe; P. Butkinaree; K. Park; S. Shimoji; Z. Wang; J. Bullen, *Glycobiology* **2006**, *16*, 1105.
206. C. Slawson; G. W. Hart, *FASEB J.* **2006**, *20*, A37–A37.
207. W. L. Du; D. Edelstein; L. Rossetti; I. G. Fantus; H. Goldberg; F. Ziyadeh; J. Wu; M. Brownlee, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12222–12226.
208. D. J. Vocadlo; S. G. Withers, *Biochemistry* **2005**, *44*, 12809–12818.
209. S. J. Williams; B. L. Mark; D. J. Vocadlo; M. N. G. James; S. G. Withers, *J. Biol. Chem.* **2002**, *277*, 40055–40065.
210. I. Tews; A. C. T. vanScheltinga; A. Perrakis; K. S. Wilson; B. W. Dijkstra, *J. Am. Chem. Soc.* **1997**, *119*, 7954–7959.
211. D. Piszkiewicz; T. C. Bruice, *J. Am. Chem. Soc.* **1967**, *89*, 6237–6243.
212. G. Lowe; G. Sheppard, *Chem. Commun.* **1968**, *9*, 529–530.
213. G. Lowe; G. Sheppard; M. L. Sinnott; A. Williams, *Biochem. J.* **1967**, *104*, 893.
214. I. Tews; A. Perrakis; A. Oppenheim; Z. Dauter; K. S. Wilson; C. E. Vorgias, *Nat. Struct. Biol.* **1996**, *3*, 638–648.
215. A. C. T. Vanscheltinga; S. Armand; K. H. Kalk; A. Isogai; B. Henrissat; B. W. Dijkstra, *Biochemistry* **1995**, *34*, 15619–15623.
216. C. E. Vorgias; A. Perrakis; I. Tews, Structure-Function Studies on the Chitinolytic Enzymes of *Serratia marcescens* Chitinase and Chitobiase. In *Enzyme Engineering Xiii*; J. S. Dordick; A. J. Russell, Eds.; New York Academy of Sciences: New York, 1996; Vol. 799, pp 190–192.
217. M. J. Lemieux; B. L. Mark; M. M. Cherney; S. G. Withers; D. J. Mahuran; M. N. G. James, *J. Mol. Biol.* **2006**, *359*, 913–929.
218. B. L. Mark; D. J. Mahuran; M. M. Cherney; D. L. Zhao; S. Knapp; M. N. G. James, *J. Mol. Biol.* **2003**, *327*, 1093–1109.
219. T. Maier; N. Strater; C. G. Schuette; R. Klingenstein; K. Sandhoff; W. Saenger, *J. Mol. Biol.* **2003**, *328*, 669–681.
220. S. Knapp; D. Vocadlo; Z. N. Gao; B. Kirk; J. P. Lou; S. G. Withers, *J. Am. Chem. Soc.* **1996**, *118*, 6804–6805.
221. M. Fujita; S. Shoda; T. Haneda; K. Inaza, *Biochim. Biophys. Acta* **2001**, *1528*, 9–14.
222. B. Li; Y. Zeng; S. Hauser; H. Song; L. X. Wang, *J. Am. Chem. Soc.* **2005**, *127*, 9692–9693.
223. Y. Zeng; J. Wang; B. Li; S. Hauser; H. Li; L. X. Wang, *Chem. Eur. J.* **2006**, *12*, 3355–3364.
224. Y. Gao; L. Wells; F. I. Comer; G. J. Parker; G. W. Hart, *J. Biol. Chem.* **2001**, *276*, 9838–9845.
225. L. Wells; K. Vosseller; G. W. Hart, *Science* **2001**, *291*, 2376–2378.
226. Z. Wang; A. Pandey; G. W. Hart, *Mol. Cell. Proteomics* **2007**, *6*, 1365–1379.
227. G. Z. Tao; C. Kirby; S. A. Whelan; F. Rossi; X. H. Bi; M. MacLaren; E. Gentalen; R. A. O'Neill; G. W. Hart; M. B. Omary, *Biochem. Biophys. Res. Commun.* **2006**, *351*, 708–712.
228. F. Liu; K. Iqbal; I. Grundke-Iqbal; G. W. Hart; C. X. Gong, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10804–10809.
229. N. E. Zachara; G. W. Hart, *Biochim. Biophys. Acta* **2004**, *1673*, 13–28.
230. N. E. Zachara; G. W. Hart, *Trends Cell Biol.* **2004**, *14*, 218–221.
231. N. E. Zachara; C. Butkinaree; G. W. Hart, *Glycobiology* **2003**, *13*, 833.
232. F. I. Comer; G. W. Hart, *J. Biol. Chem.* **2000**, *275*, 29179–29182.
233. G. W. Hart, *Annu. Rev. Biochem.* **1997**, *66*, 315–335.
234. L. K. Kreppel; M. A. Blomberg; G. W. Hart, *J. Biol. Chem.* **1997**, *272*, 9308.
235. W. A. Lubas; D. W. Frank; M. Krause; J. A. Hanover, *J. Biol. Chem.* **1997**, *272*, 9316.
236. E. J. Hehre, *Carbohydr. Res.* **2001**, *331*, 347–368.
237. B. L. Mark; G. A. Wasney; T. J. S. Salo; A. R. Khan; Z. M. Cao; P. W. Robbins; M. N. G. James; B. L. Triggs-Raine, *J. Biol. Chem.* **1998**, *273*, 19618–19624.
238. A. C. Terwischa van Scheltinga; S. Armand; K. H. Kalk; A. Isogai; B. Henrissat; B. W. Dijkstra, *Biochemistry* **1995**, *34*, 15619.
239. G. Vaaje-Kolstad; D. R. Houston; F. V. Rao; M. G. Peter; B. Synstad; D. M. F. Van Aalten; V. G. H. Eijsink, *Biochim. Biophys. Acta* **2004**, *1696*, 103–111.
240. G. E. Whitworth; M. S. Macauley; K. A. Stubbs; R. J. Dennis; E. J. Taylor; G. J. Davies; I. R. Greig; D. J. Vocadlo, *J. Am. Chem. Soc.* **2007**, *129*, 635–644.
241. M. S. Macauley; G. E. Whitworth; A. W. Debowski; D. Chin; D. J. Vocadlo, *J. Biol. Chem.* **2005**, *280*, 25313–25322.
242. C. S. Jones; D. J. Kosman, *J. Biol. Chem.* **1980**, *255*, 11861–11869.
243. D. J. Vocadlo; G. J. Davies; R. Laine; S. G. Withers, *Nature* **2001**, *412*, 835–838.
244. D. J. Vocadlo; C. Mayer; S. M. He; S. G. Withers, *Biochemistry* **2000**, *39*, 117–126.
245. A. Varki, *Glycobiology* **1992**, *2*, 25–40.
246. T. Angata; A. Varki, *Chem. Rev.* **2002**, *102*, 439–469.
247. A. Varki, *FASEB J.* **1997**, *11*, 248–255.
248. L. V. Gubareva; L. Kaiser; F. G. Hayden, *Lancet* **2000**, *355*, 827–835.

249. M. Vonitzstein; W. Y. Wu; G. B. Kok; M. S. Pegg; J. C. Dyason; B. Jin; T. V. Phan; M. L. Smythe; H. F. White; S. W. Oliver; P. M. Colman; J. N. Varghese; D. M. Ryan; J. M. Woods; R. C. Bethell; V. J. Hotham; J. M. Cameron; C. R. Penn, *Nature* **1993**, *363*, 418–423.
250. G. Taylor, *Curr. Opin. Struct. Biol.* **1996**, *6*, 830–837.
251. K. Stummeyer; A. Dickmanns; M. Muhlenhoff; R. Gerardy-Schahn; R. Ficner, *Nat. Struct. Mol. Biol.* **2005**, *12*, 90–96.
252. A. K. J. Chong; M. S. Pegg; N. R. Taylor; M. Vonitzstein, *Eur. J. Biochem.* **1992**, *207*, 335–343.
253. X. M. Guo; M. L. Sinnott, *Biochem. J.* **1993**, *296*, 291–292.
254. X. M. Guo; W. G. Laver; E. Vimr; M. L. Sinnott, *J. Am. Chem. Soc.* **1994**, *116*, 5572–5578.
255. J. C. Wilson; D. I. Angus; M. Vonitzstein, *J. Am. Chem. Soc.* **1995**, *117*, 4214–4217.
256. Y. H. Kao; L. Lerner; T. G. Warner, *Glycobiology* **1997**, *7*, 559–563.
257. A. R. Todeschini; L. Mendonca-Previato; J. O. Previato; A. Varki; H. van Halbeek, *Glycobiology* **2000**, *10*, 213–221.
258. R. J. Russell; L. F. Haire; D. J. Stevens; P. J. Collins; Y. P. Lin; G. M. Blackburn; A. J. Hay; S. J. Gamblin; J. J. Skehel, *Nature* **2006**, *443*, 45–49.
259. L. Venkatramani; E. Bochkareva; J. T. Lee; U. Gulati; W. Graeme Laver; A. Bochkarev; G. M. Air, *J. Mol. Biol.* **2006**, *356*, 651–663.
260. P. Bossart-Whitaker; M. Carson; Y. S. Babu; C. D. Smith; W. G. Laver; G. M. Air, *J. Mol. Biol.* **1993**, *232*, 1069–1083.
261. M. J. Jedrzejas; S. Singh; W. J. Brouillette; W. G. Laver; G. M. Air; M. Luo, *Biochemistry* **1995**, *34*, 3144–3151.
262. W. R. Tulip; J. N. Varghese; W. G. Laver; R. G. Webster; P. M. Colman, *J. Mol. Biol.* **1992**, *227*, 122–148.
263. E. Rudino-Pinera; P. Tunnah; S. J. Crennell; R. G.Webster; W. G. Laver; E. F. Garman, *PDB Code 1w20*.
264. J. B. Finley; V. R. Atigadda; F. Duarte; J. J. Zhao; W. J. Brouillette; G. M. Air; M. Luo, *J. Mol. Biol.* **1999**, *293*, 1107–1119.
265. N. R. Taylor; A. Cleasby; O. Singh; T. Skarzynski; A. J. Wonacott; P. W. Smith; S. L. Sollis; P. D. Howes; P. C. Cherry; R. Bethell; P. Colman; J. Varghese, *J. Med. Chem.* **1998**, *41*, 798–807.
266. A. Gaskell; S. Crennell; G. Taylor, *Structure* **1995**, *3*, 1197–1205.
267. S. J. Crennell; E. F. Garman; C. Philippon; A. Vasella; W. G. Laver; E. R. Vimr; G. L. Taylor, *J. Mol. Biol.* **1996**, *259*, 264–280.
268. I. Moustafa; H. Connaris; M. Taylor; V. Zaitsev; J. C. Wilson; M. J. Kiefel; M. von Itzstein; G. Taylor, *J. Biol. Chem.* **2004**, *279*, 40819–40826.
269. L. M. Chavas; C. Tringali; P. Fusi; B. Venerando; G. Tettamanti; R. Kato; E. Monti; S. Wakatsuki, *J. Biol. Chem.* **2005**, *280*, 469–475.
270. Y. Luo; S. C. Li; M. Y. Chou; Y. T. Li; M. Luo, *Structure* **1998**, *6*, 521–530.
271. A. Buschiazzo; M. F. Amaya; M. L. Cremona; A. C. Frasch; P. M. Alzari, *Mol. Cell* **2002**, *10*, 757–768.
272. M. F. Amaya; A. Buschiazzo; T. Nguyen; P. M. Alzari, *J. Mol. Biol.* **2003**, *325*, 773–784.
273. W. P. Burmeister; B. Henrissat; C. Bosso; S. Cusack; R. W. Ruigrok, *Structure* **1993**, *1*, 19–26.
274. B. A. Horenstein; M. Bruner, *J. Am. Chem. Soc.* **1996**, *118*, 10371–10379.
275. B. A. Horenstein; M. Bruner, *J. Am. Chem. Soc.* **1998**, *120*, 1357–1362.
276. N. R. Taylor; M. Vonitzstein, *J. Med. Chem.* **1994**, *37*, 616–624.
277. A. G. Watts; P. Oppezzo; S. G. Withers; P. M. Alzari; A. Buschiazzo, *J. Biol. Chem.* **2006**, *281*, 4149–4155.
278. S. L. Newstead; J. A. Potter; J. C. Wilson; G. G. Xu; C. H. Chien; A. G. Watts; S. G. Withers; G. L. Taylor, *J. Biol. Chem.* **2008**, *283*, 9080–9088.
279. I. Damager; S. Buchini; M. F. Amaya; A. Buschiazzo; P. Alzari; A. C. Frasch; A. Watts; S. G. Withers, *Biochemistry* **2008**, *47*, 3507–3512.
280. M. F. Amaya; A. G. Watts; T. Damager; A. Wehenkel; T. Nguyen; A. Buschiazzo; G. Paris; A. C. Frasch; S. G. Withers; P. M. Alzari, *Structure* **2004**, *12*, 775–784.
281. J. N. Varghese; J. L. McKimm-Breschkin; J. B. Caldwell; A. A. Kortt; P. M. Colman, *Proteins* **1992**, *14*, 327–332.
282. W. P. Burmeister; R. W. Ruigrok; S. Cusack, *EMBO J.* **1992**, *11*, 49–56.
283. M. N. Janakiraman; C. L. White; W. G. Laver; G. M. Air; M. Luo, *Biochemistry* **1994**, *33*, 8172–8179.
284. G. J. Davies; L. Mackenzie; A. Varrot; M. Dauter; A. M. Brzozowski; M. Schulein; S. G. Withers, *Biochemistry* **1998**, *37*, 11707–11713.
285. S. G. Withers; R. Aebersold, *Protein Sci.* **1995**, *4*, 361–372.
286. J. Sauer; B. W. Sigurskjold; U. Christensen; T. P. Frandsen; E. Mirgorodskaya; M. Harrison; P. Roepstorff; B. Svensson, *Biochim. Biophys. Acta* **2000**, *1543*, 275–293.
287. A. Varrot; T. P. Frandsen; I. von Ossowski; V. Boyer; S. Cottaz; H. Driguez; M. Schulein; G. J. Davies, *Structure* **2003**, *11*, 855–864.
288. A. Varrot; S. Hastrup; M. Schulein; G. J. Davies, *Biochem. J.* **1999**, *337*, 297–304.
289. B. Svensson, *Plant Mol. Biol.* **1994**, *25*, 141–157.
290. D. Norouzian; A. Akbarzadeh; J. M. Scharer; M. Moo Young, *Biotechnol. Adv.* **2006**, *24*, 80–85.
291. P. J. Reilly, *Starch-Stärke* **1999**, *51*, 269–274.
292. P. M. Coutinho; P. J. Reilly, *Protein Eng.* **1994**, *7*, 749–760.
293. P. M. Coutinho; P. J. Reilly, *Proteins* **1997**, *29*, 334–347.
294. A. Aleshin; A. Golubev; L. M. Firsov; R. B. Honzatko, *J. Biol. Chem.* **1992**, *267*, 19291–19298.
295. A. E. Aleshin; C. Hoffman; L. M. Firsov; R. B. Honzatko, *J. Mol. Biol.* **1994**, *238*, 575–591.
296. J. Sevcik; A. Solovicova; E. Hostinova; J. Gasperik; K. S. Wilson; Z. Dauter, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 854–866.
297. A. E. Aleshin; P. H. Feng; R. B. Honzatko; P. J. Reilly, *J. Mol. Biol.* **2003**, *327*, 61–73.
298. K. Sorimachi; A. J. Jacks; M. F. LeGalCoeffet; G. Williamson; D. B. Archer; M. P. Williamson, *J. Mol. Biol.* **1996**, *259*, 970–987.
299. K. Sorimachi; M. F. LeGalCoeffet; G. Williamson; D. B. Archer; M. P. Williamson, *Structure* **1997**, *5*, 647–661.
300. S. M. Southall; P. J. Simpson; H. J. Gilbert; G. Williamson; M. P. Williamson, *FEBS Lett.* **1999**, *447*, 58–60.
301. G. F. H. Kramer; A. P. Gunning; V. J. Morris; N. J. Belshaw; G. Williamson, *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 2595–2602.
302. N. Payre; S. Cottaz; C. Boisset; R. Borsali; B. Svensson; B. Henrissat; H. Driguez, *Angew. Chem. Int. Ed.* **1999**, *38*, 974–977.

303. R. Bott; M. Saldajeno; W. Cuevas; D. Ward; M. Scheffers; W. Aehle; S. Karkehabadi; M. Sandgren; H. Hansson, *Biochemistry* **2008**, *47*, 5746–5754.
304. B. Svensson; A. J. Clarke; I. Svendsen; H. Moller, *Eur. J. Biochem.* **1990**, *188*, 29–38.
305. M. R. Sierks; C. Ford; P. J. Reilly; B. Svensson, *Protein Eng.* **1990**, *3*, 193–198.
306. E. M. Harris; A. E. Aleshin; L. M. Firsov; R. B. Honzatko, *Biochemistry* **1993**, *32*, 1618–1626.
307. T. P. Frandsen; C. Dupont; J. Lehmbeck; B. Stoffer; M. R. Sierks; R. B. Honzatko; B. Svensson, *Biochemistry* **1994**, *33*, 13808–13816.
308. G. D. Brayer; G. Sidhu; R. Maurus; E. H. Rydberg; C. Braun; Y. L. Wang; N. T. Nguyen; C. H. Overall; S. G. Withers, *Biochemistry* **2000**, *39*, 4778–4791.
309. H. Matsui; J. S. Blanchard; C. F. Brewer; E. J. Hehre, *J. Biol. Chem.* **1989**, *264*, 8714–8716.
310. Y. Tanaka; W. Tao; J. S. Blanchard; E. J. Hehre, *J. Biol. Chem.* **1994**, *269*, 32306–32312.
311. B. W. Sigurskjold; C. R. Berland; B. Svensson, *Biochemistry* **1994**, *33*, 10191–10199.
312. A. E. Aleshin; B. Stoffer; L. M. Firsov; B. Svensson; R. B. Honzatko, *Biochemistry* **1996**, *35*, 8319–8328.
313. C. R. Berland; B. W. Sigurskjold; B. Stoffer; T. P. Frandsen; B. Svensson, *Biochemistry* **1995**, *34*, 10153–10161.
314. J. A. Thoma; D. E. Koshland, Jr., *J. Biol. Chem.* **1960**, *235*, 2511–2517.
315. J. A. Thoma; D. E. Koshland, Jr., *Biochemistry* **1965**, *4*, 714–722.
316. Y. Nitta; T. Kunikata; T. Watanabe, *J. Biochem.* **1979**, *85*, 41–45.
317. Y. Isoda; Y. Nitta, *J. Biochem.* **1986**, *99*, 1631–1637.
318. Y. Nitta; Y. Isoda; H. Toda; F. Sakiyama, *J. Biochem.* **1989**, *105*, 573–576.
319. B. Mikami; E. J. Hehre; M. Sato; Y. Katsube; M. Hirose; Y. Morita; J. C. Sacchettini, *Biochemistry* **1993**, *32*, 6836–6845.
320. J. A. Thoma; D. E. Koshland, Jr., *J. Am. Chem. Soc.* **1960**, *82*, 3329–3333.
321. B. Mikami; M. Degano; E. J. Hehre; J. C. Sacchettini, *Biochemistry* **1994**, *33*, 7779–7787.
322. A. Totsuka; V. H. Nong; H. Kadokawa; C. S. Kim; Y. Itoh; C. Fukazawa, *Eur. J. Biochem.* **1994**, *221*, 649–654.
323. A. Totsuka; C. Fukazawa, *Eur. J. Biochem.* **1996**, *240*, 655–659.
324. Y. N. Kang; M. Adachi; S. Utsumi; B. Mikami, *J. Mol. Biol.* **2004**, *339*, 1129–1140.
325. M. Adachi; B. Mikami; T. Katsube; S. Utsumi, *J. Biol. Chem.* **1998**, *273*, 19859–19865.
326. C. G. Cheong; S. H. Eom; C. Chang; D. H. Shin; H. K. Song; K. Min; J. H. Moon; K. K. Kim; K. Y. Hwang; S. W. Suh, *Proteins* **1995**, *21*, 105–117.
327. B. Mikami; H. J. Yoon; N. Yoshigi, *J. Mol. Biol.* **1999**, *285*, 1235–1243.
328. B. Mikami; M. Adachi; T. Kage; E. Sarikaya; T. Nanmori; R. Shinke; S. Utsumi, *Biochemistry* **1999**, *38*, 7050–7061.
329. E. J. Hehre; G. Okada; D. S. Genghof, *Arch. Biochem. Biophys.* **1969**, *135*, 75–89.
330. E. J. Hehre; C. F. Brewer; D. S. Genghof, *J. Biol. Chem.* **1979**, *254*, 5942–5950.
331. E. J. Hehre; S. Kitahata; C. F. Brewer, *J. Biol. Chem.* **1986**, *261*, 2147–2153.
332. S. Kitahata; S. Chiba; C. F. Brewer; E. J. Hehre, *Biochemistry* **1991**, *30*, 6769–6775.
333. Y. Honda; M. Kitaoka, *J. Biol. Chem.* **2006**, *281*, 1426–1431.
334. L. F. Mackenzie; Q. P. Wang; R. A. J. Warren; S. G. Withers, *J. Am. Chem. Soc.* **1998**, *120*, 5583–5584.
335. U. M. Unligil; J. M. Rini, *Curr. Opin. Struct. Biol.* **2000**, *10*, 510–517.
336. Y. Hu; S. Walker, *Chem. Biol.* **2002**, *9*, 1287–1296.
337. C. Breton; L. Snajdrova; C. Jeanneau; J. Koca; A. Imberty, *Glycobiology* **2006**, *16*, 29r–37r.
338. B. Ma; J. L. Simala-Grant; D. E. Taylor, *Glycobiology* **2006**, *16*, 158r–184r.
339. F. Dupuy; J. M. Petit; R. Mollicone; R. Oriol; R. Julien; A. Maftah, *J. Biol. Chem.* **1999**, *274*, 12257–12262.
340. F. Dupuy; A. Germot; R. Julien; A. Maftah, *Glycobiology* **2004**, *14*, 347–356.
341. Z. Xu; L. Vo; B. A. Macher, *J. Biol. Chem.* **1996**, *271*, 8818–8823.
342. Z. M. Ge; D. E. Taylor, *J. Bacteriol.* **1997**, *179*, 4970–4976.
343. D. A. Rasko; G. Wang; M. M. Palcic; D. E. Taylor, *J. Biol. Chem.* **2000**, *275*, 4988–4994.
344. B. Ma; L. H. Lau; M. M. Palcic; B. Hazes; D. E. Taylor, *J. Biol. Chem.* **2006**, *281*, 22428.
345. B. Ma; G. F. Audette; S. J. Lin; M. M. Palcic; B. Hazes; D. E. Taylor, *J. Biol. Chem.* **2006**, *281*, 6385–6394.
346. L. Qiao; B. W. Murray; M. Shimazaki; J. Schultz; C. H. Wong, *J. Am. Chem. Soc.* **1996**, *118*, 7653–7662.
347. B. S. Khatra; D. G. Herries; K. Brew, *Eur. J. Biochem.* **1974**, *44*, 537–560.
348. A. E. Kearns; S. C. Campbell; J. Westley; N. B. Schwartz, *Biochemistry* **1991**, *30*, 7477–7483.
349. B. W. Murray; S. Takayama; J. Schultz; C. H. Wong, *Biochemistry* **1996**, *35*, 11183–11195.
350. M. L. Mitchell; F. Tian; L. V. Lee; C. H. Wong, *Angew. Chem. Int. Ed.* **2002**, *41*, 3041–3044.
351. T. de Vries; C. A. Srnka; M. M. Palcic; S. J. Swiedler; D. H. van den Eijnden; B. A. Macher, *J. Biol. Chem.* **1995**, *270*, 8712–8722.
352. S. Gosselin; M. M. Palcic, *Bioorg. Med. Chem.* **1996**, *4*, 2023–2028.
353. S. W. Lin; T. M. Yuan; J. R. Li; C. H. Lin, *Biochemistry* **2006**, *45*, 8108–8116.
354. H. Y. Sun; S. W. Lin; T. P. Ko; J. F. Pan; C. L. Liu; C. N. Lin; A. H. J. Wang; C. H. Lin, *J. Biol. Chem.* **2007**, *282*, 9973–9982.
355. L. Lariviere; V. Gueguen-Chaignon; S. Morera, *J. Mol. Biol.* **2003**, *330*, 1077–1086.
356. A. HarduinLepers; M. A. Recchi; P. Delannoy, *Glycobiology* **1995**, *5*, 741–758.
357. A. P. Moran; M. M. Prendergast; B. J. Appelmelk, *FEMS Immunol. Med. Microbiol.* **1996**, *16*, 105–115.
358. C. P. C. Chiu; L. L. Lairson; M. Gilbert; W. W. Wakarchuk; S. G. Withers; N. C. J. Strynadka, *Biochemistry* **2007**, *46*, 7196–7204.
359. L. S. Ni; M. C. Sung; H. Yu; H. Chokhawala; X. Chen; A. J. Fisher, *Biochemistry* **2006**, *45*, 2139–2148.
360. L. S. Ni; H. A. Chokhawala; H. Z. Cao; R. Henning; L. Ng; S. S. Huang; H. Yu; X. Chen; A. J. Fisher, *Biochemistry* **2007**, *46*, 6288–6298.
361. Y. Kakuta; N. Okino; H. Kajiwara; M. Ichikawa; Y. Takakura; M. Ito; T. Yamamoto, *Glycobiology* **2008**, *18*, 66–73.
362. C. Breton; E. Bettler; D. H. Joziasse; R. A. Geremia; A. Imberty, *J. Biochem.* **1998**, *123*, 1000–1009.
363. T. Hennet, *Cell. Mol. Life Sci.* **2002**, *59*, 1081–1095.
364. W. W. Wakarchuk; A. Cunningham; D. C. Watson; N. M. Young, *Protein Eng.* **1998**, *11*, 295–302.
365. H. D. Ly; B. Lougheed; W. W. Wakarchuk; S. G. Withers, *Biochemistry* **2002**, *41*, 5075–5085.
366. B. Lougheed; H. D. Ly; W. W. Wakarchuk; S. G. Withers, *J. Biol. Chem.* **1999**, *274*, 37717–37722.

367. L. L. Lairson; W. W. Wakarchuk; S. G. Withers, *Chem. Commun.* **2007**, *48*, 365–367.
368. L. L. Lairson; A. G. Watts; W. W. Wakarchuk; S. G. Withers, *Nat. Chem. Biol.* **2006**, *2*, 724–728.
369. K. Persson; H. D. Ly; M. Dieckelmann; W. W. Wakarchuk; S. G. Withers; N. C. J. Strynadka, *Nat. Struct. Biol.* **2001**, *8*, 166–175.
370. J. Wicki; D. R. Rose; S. G. Withers, *Methods Enzymol.* **2002**, *354*, 84–105.
371. R. M. Mosi; S. G. Withers, *Methods Enzymol.* **2002**, *354*, 64–84.
372. R. P. Gibson; C. A. Tarling; S. Roberts; S. G. Withers; G. J. Davies, *J. Biol. Chem.* **2004**, *279*, 1950–1955.
373. H. Jamaluddin; P. Tumbale; S. G. Withers; K. R. Acharya; K. Brew, *J. Mol. Biol.* **2007**, *369*, 1270–1281.
374. L. L. Lairson; C. P. C. Chiu; H. D. Ly; S. M. He; W. W. Wakarchuk; N. C. J. Strynadka; S. G. Withers, *J. Biol. Chem.* **2004**, *279*, 28339–28344.
375. L. N. Gastinel; C. Bignon; A. K. Misra; O. Hindsgaul; J. H. Shaper; D. H. Joziasse, *EMBO J.* **2001**, *20*, 638–649.
376. K. Sujino; T. Uchiyama; O. Hindsgaul; N. O. L. Seto; W. W. Wakarchuk; M. M. Palcic, *J. Am. Chem. Soc.* **2000**, *122*, 1261–1269.
377. E. Boix; Y. N. Zhang; G. J. Swaminathan; K. Brew; K. R. Acharya, *J. Biol. Chem.* **2002**, *277*, 28310–28318.
378. Y. N. Zhang; G. J. Swaminathan; A. Deshpande; E. Boix; R. Natesh; Z. H. Xie; K. R. Acharya; K. Brew, *Biochemistry* **2003**, *42*, 13512–13521.
379. P. Molina; R. M. A. Knegtel; B. A. Macher, *Biochim. Biophys. Acta* **2007**, *1770*, 1266–1273.
380. A. Monegal; A. Planas, *J. Am. Chem. Soc.* **2006**, *128*, 16030–16031.
381. S. I. Patenaude; N. O. L. Seto; S. N. Borisova; A. Szpacenko; S. L. Marcus; M. M. Palcic; S. V. Evans, *Nat. Struct. Biol.* **2002**, *9*, 685–690.
382. Y. D. Lobsanov; P. A. Romero; B. Sleno; B. M. Yu; P. Yip; A. Herscovics; P. L. Howell, *J. Biol. Chem.* **2004**, *279*, 17921–17931.
383. L. C. Pedersen; J. Dong; F. Taniguchi; H. Kitagawa; J. M. Krahn; L. G. Pedersen; K. Sugahara; M. Negishi, *J. Biol. Chem.* **2003**, *278*, 14420–14428.
384. C. Martinez-Fleites; M. Proctor; S. Roberts; D. N. Bolam; H. J. Gilbert; G. J. Davies, *Chem. Biol.* **2006**, *13*, 1143–1152.
385. R. P. Gibson; J. P. Turkenburg; S. J. Charnock; R. Lloyd; G. J. Davies, *Chem. Biol.* **2002**, *9*, 1337–1346.
386. E. S. L. C.E. Boozer, *J. Am. Chem. Soc.* **1952**, *74*, 308–311.
387. H. W. Klein; M. J. Im; D. Palm, *Eur. J. Biochem.* **1986**, *157*, 107–114.
388. B. J. Gibbons; P. J. Roach; T. D. Hurley, *J. Mol. Biol.* **2002**, *319*, 463–477.

Biographical Sketches
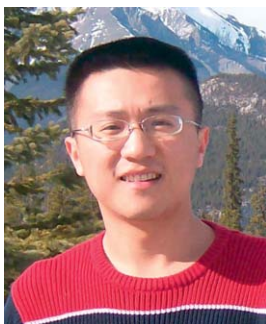
Ran Zhang was born in Jiangsu Province, China. He received his B.Sc. from Peking University, P. R. China while working in the laboratory of Professor Zhongfan Liu. In 2004 he joined Professor Stephen G. Withers' group at the University of British Columbia as a Ph.D. candidate. His research focuses on the mechanistic studies of α-glycosidases and is funded by the British Columbia Innovation Council.

Vivian L. Y. Yip received her undergraduate and graduate training at the University of British Columbia. She completed her B.Sc. degree in 2002 and her Ph.D. under the supervision of Dr. Stephen G. Withers in 2007. Her graduate work focused on the mechanistic elucidation of unusual glycosidases. She was funded by NSERC and the Michael Smith Foundation for Health Research during this time. She is also the recipient of the CCUCC Chemistry Doctoral Award.

Stephen G. Withers (B.Sc. and Ph.D.) was trained at the University of Bristol, UK, where he obtained his Ph.D. under the supervision of Dr. Michael Sinnott. Later he moved to Canada as a postdoctoral fellow, applying heteronuclear NMR to the study of enzymatic catalysis with Drs. Brian Sykes and Neil Madsen in the Department of Biochemistry at the University of Alberta. In 1982 he moved to the University of British Columbia as assistant professor of chemistry. He now holds the Khorana Chair of Chemistry and Biochemistry at UBC and serves as the director of CHiBi, the Centre for High-throughput Biology at UBC.

8.12 Synthesis of Alginate in Bacteria

Peter A. Tipton, University of Missouri, Columbia, MO, USA

8.12.1 Introduction

Alginate is a linear polysaccharide that is produced by the bacterial genera *Pseudomonas* and *Azotobacter*, and by marine algae. Bacterial alginates are secreted to form a protective capsule, and in algae alginate plays a variety of structural roles. The viscosity of aqueous alginate solutions varies with the composition of the polymer and its size, and in the presence of divalent metal ions some alginate solutions can form a gel. The physicochemical properties of alginate have led to its ubiquitous use as an additive in the food industry, and its role in bacterial adherence, colonization, and survival in infections has attracted a great deal of attention from biomedical researchers. In particular, *Pseudomonas aeruginosa* colonizes the lung tissue of patients with cystic fibrosis (CF), and the production of alginate has an adverse effect on pulmonary function, so that the leading cause of morbidity and mortality in CF patients is *P. aeruginosa* infection.[1] Other patient populations are also at risk for *P. aeruginosa* infections; a recent study in Canada found that *P. aeruginosa* accounted for 10% of the organisms isolated from ICU patients, and a significant fraction of those isolates were multidrug resistant.[2] Within the realm of extracellular polysaccharides alginate has a relatively simple structure, and the genes required for alginate biosynthesis in several organisms have been identified, so it offers a tractable system for studying the complex processes by which polysaccharides are made, modified, and secreted.

8.12.2 Alginate Structure

There are two different monosaccharide units in alginate, β-D-mannuronate (M) and α-L-guluronate (G). The units are linked through 1,4-glycosidic bonds, and the polymers are unbranched, extending for hundreds of residues, having molecular weights ranging from 100 000 to 500 000[3] (**Figure 1**). The relative amounts of M and G and the pattern of their distribution in the polymer differ in alginates from different sources. Alginates isolated from the genera *Sargassum* (seaweed) contain between 15 and 45% M,[4] and alginates from *Laminaria* (kelp) contain 30–65% M.[5] The marine alginates are composed of block structures made up of long sequences of consecutive M residues (M blocks), G residues (G blocks), and alternating M and G residues (MG blocks). The bacterial alginates are generally characterized by a higher M content. *Azotobacter* alginate typically contains 80–90% M, and the M content of *P. aeruginosa* alginate isolated from the sputum of CF patients varies from 50
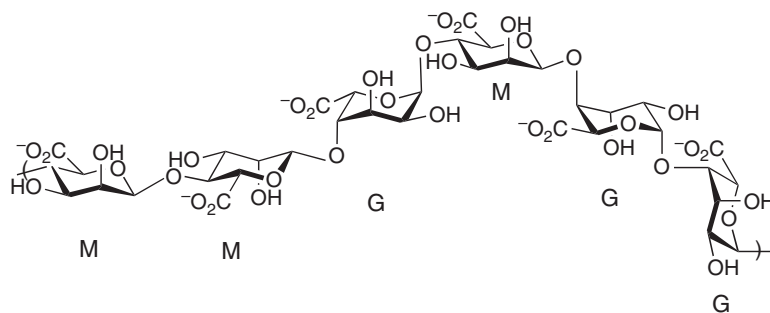
Figure 1   Structure of alginate. A section of the polysaccharide with an arbitrary sequence is shown. Residues labeled M are mannuronate residues and those labeled G are guluronate residues.

to 90%.[3,6] Bacterial alginates differ from the marine algae alginates in that the bacterial alginates are partially acetylated. Acetylation occurs on approximately 10% of the M residues at O-2 or O-3. G residues are not acetylated. *Azotobacter* alginate has a block structure like the marine algae alginates, but *Pseudomonas* alginate does not have G blocks, and the G residues are randomly dispersed throughout the polymer.

The M residues in alginate adopt the $^4C_1$ conformation, which places the carboxylate group attached to C5 in an equatorial position. Guluronate is the C5 epimer of M, and the pyranose shifts to the $^1C_4$ conformation so the C5 substituent is again in an equatorial position.[7] The conformations of the monomers dictate that M–M pairs are linked by diequatorial glycosidic bonds and G–G pairs are linked by diaxial glycosidic bonds. The different conformations of M and G residues also means that the conformation of the polymer undergoes a dramatic local change as a result of the enzymatic reaction that converts a M residue into a G residue.

As a carboxylate-containing polyol, alginate coordinates with metal ions; the ability of alginate in seaweed to chelate heavy metals including gold, cadmium, copper, zinc, nickel, and lead, has attracted interest in using it as a biosorbent for toxic heavy metals.[8,9] Coordination with Ca^{2+} is physiologically relevant in the lung tissue of CF patients. Because of the defect in the CF transmembrane receptor protein, the extracellular milieu of the lung tissue is dehydrated, which impedes mucociliary clearance of foreign objects. Mucoid (alginate-producing) *P. aeruginosa* infections exacerbate the situation because of the viscous nature of alginate.[10] The viscosity of alginate solutions increases with the G content, and in the presence of Ca^{2+}, G-rich alginate can form a gel.

The structure of the Ca^{2+}-alginate gel has been the subject of some speculation. The 'eggbox' model[11] is widely cited. In the eggbox model adjacent G residues in one polymer strand form a pocket that contributes ligands to occupy half of the coordination sites of Ca^{2+}; a second strand provides the remaining ligands, thus causing interstrand association. The interactions provided by a single Ca^{2+} are not sufficient to enforce association of the strands; however, gelation occurs when sequential Gs (at least 20) coordinate Ca^{2+}. A schematic of the Ca^{2+}-mediated association between alginate strands that emphasizes the pocket formed between adjacent G residues and occupied by Ca^{2+} looks somewhat like a carton of eggs viewed from the top. Adjacent M residues with their diequatorial glycosidic linkages do not form the deep Ca^{2+}-binding pocket, and so do not have specific Ca^{2+}-binding sites. Computational studies of the Ca^{2+}-mediated association between poly(G) strands suggests that the eggbox model is largely correct, but that it misassigns the coordination of Ca^{2+} and neglects the importance of hydrogen bonding between carbohydrate residues.[12] The structure proposed for Ca^{2+}-poly(G) complexes is a 2_1 helix composed of antiparallel strands in which Ca^{2+} is coordinated to the carboxylate and 2-OH from each chain.

One of the great advantages alginate offers the biochemical experimentalist is that the polymer can be 'sequenced' by ^1H-NMR spectroscopy.[5,13,14] Although the positions of specific residues in the chain cannot be identified, the identity of the neighbors of a given residue influence the chemical shift of some of the ^1H resonances, so a statistical picture of the composition of the polymer can be developed. The composition of the polymer is established by comparison of the anomeric ^1H signals. H-1 of G appears at 5.05 ppm, regardless of the identities of the neighboring residues. Two signals are observed for H-1 of M; a peak at 4.67 ppm for M residues that are connected to another M residue on the reducing end, and a peak at 4.70 ppm for those M residues that have a G residue on the reducing end. The relative intensities of the two anomeric proton signals

for M residues indicate the frequencies of the MG and MM diads. The chemical shift of H-5 of G residues is exquisitely sensitive to the identity of the flanking residues. An M residue at the reducing end of the G residue places G H-5 approximately 3 Å away from the 3-OH of M; on the contrary, when the G residue has another G at its reducing end, the nearest OH group to H-5 is about 4 Å away. As a result, the H-5 signal for G is deshielded when an M residue is at the reducing end, and appears 0.3 ppm downfield from H-5 on a G that has a G residue at its reducing end.[13] Small chemical shift changes are induced by the preceding residue, so H-5 from a G residue flanked by G and G (the GGG triad) appears at 4.46 ppm, and H-5 from a G residue flanked by M and G (the MGG triad) appears at 4.44 ppm. When the G residue is flanked by G and M (the GGM triad), the H-5 signal is at 4.75 ppm, and when G is flanked by M and M (the MGM triad), H-5 appears at 4.73 ppm. Measurement of the MG and MM diad frequencies, and the GGG, MGG, GGM, and MGM triad frequencies, and consideration of the relationships between the triads and diads based on mass conservation, gives a picture of the composition of the alginate in terms of G blocks, M blocks, and MG blocks.

8.12.3 Overview of Alginate Biosynthesis

Although alginate is a polymer of six-carbon uronic acids, bacteria grown on glucose incorporate few of the hexose units intact. Isotope labeling experiments demonstrate that C6 is incorporated into alginate 10–100 times more frequently than C1, indicating that most of the glucose is metabolized through the Entner–Doudoroff pathway which converts glucose 6-phosphate into pyruvate and glyceraldehyde 3-phosphate.[15,16] The glyceraldehyde 3-phosphate equilibrates with dihydroxyacetone phosphate, and the two species recondense to form fructose 1,6-bisphosphate. Fructose 6-phosphate generated by this pathway is the starting point for alginate biosynthesis. The level of labeling in alginate from glucose-derived pyruvate is low because the pyruvate equilibrates with many metabolite pools before it is converted into another glycolytic intermediate. Experiments with *Pseudomonas mendocina* indicate that the carbon skeleton of fructose can be incorporated intact into alginate.[17]

Alginate synthesis requires seven chemical transformations from fructose 6-phosphate (**Figure 2**). An excellent review that summarizes early work on the pathway in *P. aeruginosa* is available.[18] The first part of the pathway, in which the activated monomer unit is prepared, occurs in the cytoplasm. The formation of the first polymeric intermediate, mannuronan, is catalyzed by proteins that are localized in the membrane separating the cytoplasm from the periplasmic space. Epimerization and acetylation of alginate occurs in the periplasmic space, and mature alginate passes through the outer membrane via a porin-like protein. Twelve genes that encode proteins involved in alginate biosynthesis are located in an operon at 35 min on the *P. aeruginosa* chromosomal map, and one additional structural gene is well outside the operon at about 10 min.[19]

The first steps in the pathway accomplish the synthesis of the nucleotide sugar that donates the monosaccharide units to the polymer. Fructose 6-phosphate is isomerized to mannose 6-phosphate; the phosphoryl group is then transferred from the 6-position to the 1-position to generate mannose 1-phosphate. Mannose 1-phosphate and GTP are substrates for a pyrophosphorylase that catalyzes formation of GDP-mannose. GDP-mannose is oxidized to GDP-mannuronate in the final step that is catalyzed by a cytoplasmic enzyme.

GDP-mannuronate is the activated precursor that donates M residues to mannuronan. The proteins required for this step are localized in the inner membrane and subsequent transformations of the incipient alginate occur in the periplasmic space. The final steps of alginate synthesis are epimerization and acetylation. These transformations are not carried out at every residue. Epimerization occurs at C5, converting a β-D-mannuronate residue to a α-L-guluronate residue. The number and pattern of distribution of G residues in mature alginate differs across species. Acetylation occurs at O2 or O3, and only on M residues. Acetylation and epimerization appear to be mutually exclusive, such that acetylation precludes epimerization and epimerization prevents acetylation. Secretion of mature alginate is mediated by AlgE, a porin-like protein.

The alginate pathway and the activities of the alginate enzymes overlap with a few other metabolic endpoints. The committed step in alginate biosynthesis is the formation of GDP-mannuronate. GDP-mannose is the precursor of GDP-rhamnose, a constituent of the A-band in lipopolysaccharide (LPS). The GDP-mannose that is required for A-band LPS biosynthesis is formed by the action of phosphomannose isomerase (PMI)/GDP-mannose pyrophosphorylase (GMP) activities that are separate from those used in
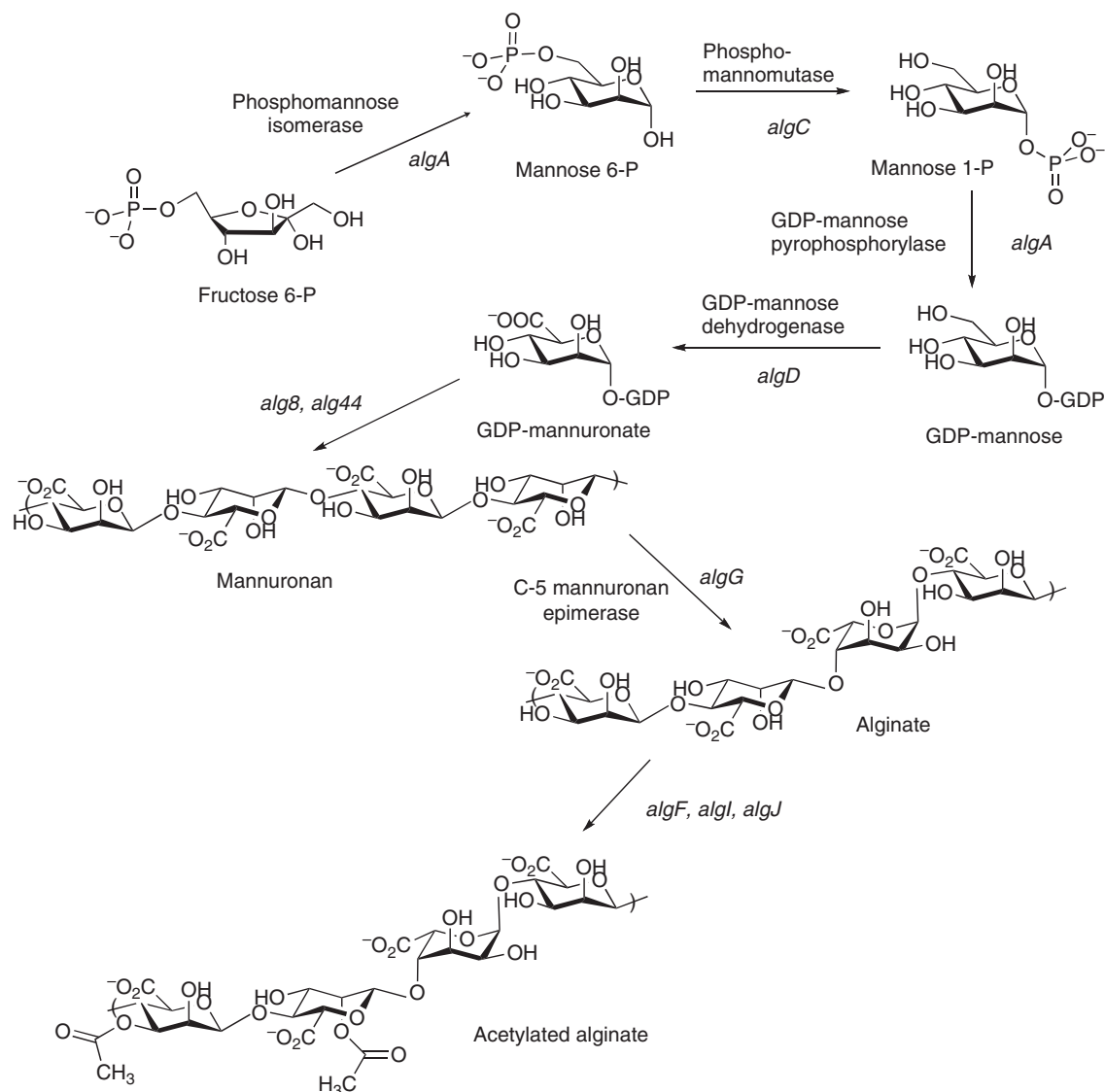
Figure 2   Pathway of alginate synthesis in *Pseudomonas aeruginosa*.

alginate biosynthesis. The *algA* gene product is used in the alginate pathway and the *wbpW* gene product operates in the LPS pathway. The two genes complement each other imperfectly, which allowed for their identification by the creation of knockout mutants.[20] The core structure of LPS contains dTDP-rhamnose, which is synthesized in several steps from glucose 1-phosphate. Although glucose 1-phosphate does not lie on the alginate biosynthetic pathway, its formation from glucose 6-phosphate is linked to alginate biosynthesis because the same enzyme, phosphomannomutase (PMM)/phosphoglucomutase (PGM), catalyzes both transformations.[21]

8.12.4   Phosphomannose Isomerase

The *algA* gene encodes a bifunctional enzyme that catalyzes the first and third steps of alginate biosynthesis. The bifunctional nature of the enzyme distinguishes it as a member of the Type II family of PMIs.[22] Although detailed mechanistic studies have not been conducted, and no structure is available, the protein has been

purified to homogeneity and some functional characteristics have been determined.[23] One of the most interesting observations is that catalysis requires a divalent metal ion. The activity is greatest with cobalt, but nickel, manganese, magnesium, calcium, and zinc also support catalysis. Ketose–aldose isomerases that are metal ion-dependent and catalyze their reaction via direct hydride transfer, and those that are metal ion-independent and catalyze their reaction using proton transfer reactions have been well characterized.[24,25] Solution model studies have established that the transition states for both pathways are similar in energy,[26] so both mechanisms should be considered viable possibilities for the PMI reaction (**Figure 3**). The metal ion specificity that is reported for PMI does not match what one would expect for metal-assisted hydride transfer (Zn^{2+}) or for a proton transfer mechanism in which the metal ion was important for binding the substrate phosphate group (Mg^{2+} or Mn^{2+}). The enzyme in *Burkholderia cepacia*, which is 50% identical and 67% similar in its amino acid sequence to the *P. aeruginosa* enzyme, also exhibits unusual metal ion specificity: $Ca^{2+} > Mn^{2+} > Mg^{2+} > Co^{2+} > Ni^{2+}$.[27] 5-Phospho-D-arabinonohydroxamic acid is a potent inhibitor of PMI, exhibiting a K_i of $0.1\,\mu mol\,l^{-1}$. The K_m for fructose 6-phosphate has not been reported, but the K_m for mannose 6-phosphate in the reverse reaction is $3\,mmol\,l^{-1}$, so it has been proposed that the tight binding of the inhibitor signifies that it is a close structural mimic of the reaction intermediate or transition state.[28] However, hydroxamate analogues have been proposed as reaction intermediate or transition state analogues for ketose–aldose isomerases that operate via proton transfer mechanisms[29] as well as those that proceed via direct hydride transfer,[30] so the potent inhibition afforded by the hydroxamate analogue does not help to determine the chemical mechanism.

8.12.5   Phosphomannomutase

The conversion of mannose 6-phosphate into mannose 1-phosphate is accomplished by the enzyme encoded by the *algC* gene, which is designated PMM/PGM in recognition of its dual role in preparing both mannose and glucose for conversion into the nucleotide sugars required for alginate and LPS core biosynthesis. The enzyme is comprised of 462 amino acids, requires Mg^{2+} for activity, and is phosphorylated at Ser108 in the resting state.

Deletion mutants of *algC* in *P. aeruginosa* are defective in alginate and LPS biosynthesis, establishing the role of PMM/PGM in both biosynthetic pathways.[21] The values for the steady-state kinetic parameter V/K, which reflects substrate specificity, are about $5\,mmol\,l^{-1}\,s^{-1}$ for both mannose 6-phosphate and glucose 6-phosphate, demonstrating that the enzyme exhibits dual substrate specificity.[31] Most kinetic studies of PMM/PGM have been conducted using glucose 1-phosphate as the substrate. Although this is the reverse of the biosynthetic direction, conversion of the 1-phosphohexose into the 6-phosphohexose is thermodynamically favored, and in the case of glucose 6-phosphate formation, the reaction can be coupled to the glucose 6-phosphate dehydrogenase reaction, providing a convenient spectrophotometric assay.

The equilibrium constant for the conversion of glucose 1-phosphate to glucose 6-phosphate is 17.3.[32] This value is the product of the equilibrium constants for interconversion of the α-D-glucose phosphoesters and the spontaneous anomerization of D-glucose 6-phosphate. The PMM/PGM is specific for the α-anomer of its substrates and products. Factoring out the equilibrium constant for anomerization, K_{eq} for the interconversion of α-D-glucose 1-phosphate and α-D-glucose 6-phosphate is about 7. The ratio of enzyme-bound α-D-glucose 6-phosphate to α-D-glucose 1-phosphate at equilibrium is about 3, indicating that the enzyme provides a small degree of stabilization to α-D-glucose 1-P.[33]

The chemical reaction catalyzed by PMM/PGM adheres to the mechanism characterized for other phosphomutases including the well-studied rabbit muscle phosphoglucomutase.[34] Following binding of mannose 6-phosphate at the active site the phosphoryl group on Ser108 is transferred onto the substrate to generate mannose 1,6-bisphosphate. The bisphosphorylated intermediate undergoes a 180° reorientation, exchanging the positions of the phosphoryl groups relative to Ser108. The phosphoryl group at C6 is transferred to Ser108, regenerating active enzyme and releasing mannose 1-phosphate (**Figure 4**).

The three-dimensional structure of PMM/PGM has been studied in detail, providing a framework for understanding its specificity and mechanism. The protein is heart-shaped and the active site is located in a deep cleft between the two lobes of the heart.[35] Comparison of the unliganded enzyme and various ligand-bound
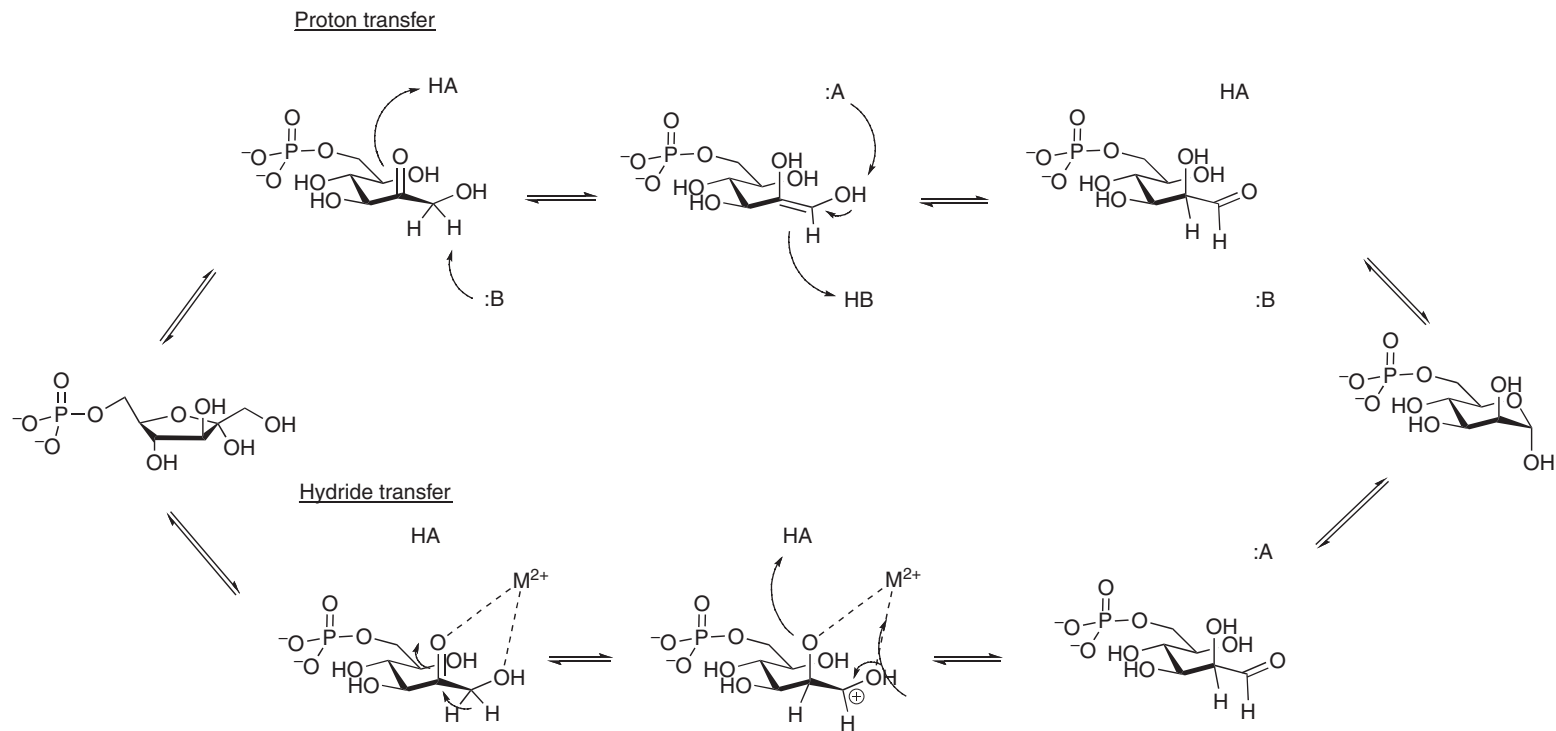
Figure 3 Potential chemical mechanisms for the reaction catalyzed by phosphomannose isomerase. The top mechanism uses general acid–base catalysis to transfer a proton between C2 and C1. The bottom mechanism uses a metal ion to assist in hydride transfer between C2 and C1.
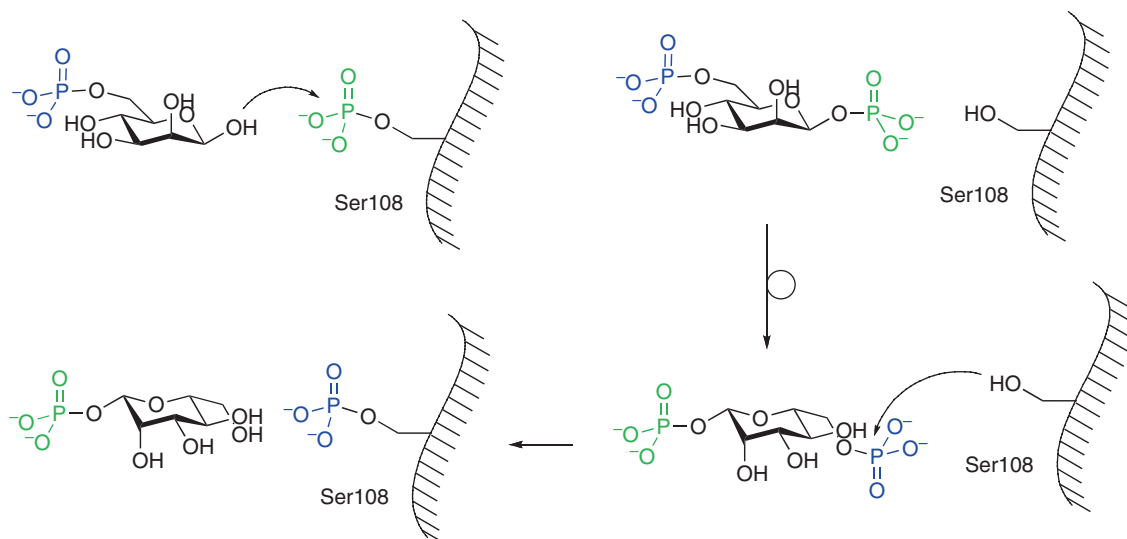
Figure 4 Mechanism of the reaction catalyzed by phosphomannomutase.

forms reveals that the enzyme cycles between an open conformation in the absence of the substrate, and a closed form when the substrate is bound. The closed conformation forms by rotation of one domain of the protein by 9°, moving some residues as much as 4.5 Å, and essentially sealing the active site from the solvent.[36]

Although PMM/PGM exhibits dual substrate specificity, it is not nonspecific. Allose 1-phosphate (the C3 epimer of glucose 1-phosphate) and galactose 1-phosphate (C4 epimer of glucose 1-phosphate) are not recognized by the enzyme as either substrates or inhibitors.[36] In mannose and glucose the hydroxyl groups at C3 and C4 are equatorial, and they exchange positions with one another when the intermediate reorients. The axial substituents at C3 in allose and at C4 in galactose preclude such a conservative positional exchange upon reorientation.

The PMM/PGM is activated by glucose 1,6-bisphosphate, and exhibits substrate inhibition in steady-state kinetic assays. The substrate inhibition can be relieved by increased concentrations of glucose 1,6-bisphosphate, which is, of course, the intermediate in the reaction with phosphoglucose.[31] Presumably, mannose 1,6-bisphosphate activates PMM/PGM as well, but this has not been tested. The substrate inhibition and its relief by the intermediate provide strong evidence for the proposed chemical mechanism. The inhibition arises when the bisphosphorylated intermediate dissociates prematurely from the enzyme during the catalytic cycle, and substrate binding to the unphosphorylated enzyme creates a dead-end complex.

Perhaps the most interesting mechanistic feature of the PMM/PGM reaction is the physical reorientation of the intermediate that must occur. The enzyme faces the challenge of how to reorient the ligand without allowing it to diffuse away. A crystal structure of glucose 1,6-bisphosphate bound to phosphorylated enzyme (a nonproductive complex) shows that the enzyme is in a conformation intermediate between the open and closed states that characterize the ligand-free and ligand-bound states, respectively.[37] Surprisingly, pre-steady-state kinetic studies suggest that the reorientation is a kinetically detectable event, that is, there is a point in the catalytic cycle when the intermediate is unable to phosphorylate the enzyme.[33] This point represents the time that the intermediate is midway through its reorientation and Ser108 is not accessible to either phosphoryl group.

PMM/PGM is unable to prevent loss of the intermediate during every catalytic cycle. Isotope-trapping studies demonstrated that approximately 1 out of 15 times that the intermediate is produced, it is lost from the active site.[33] It is not known whether the reorientation of the intermediate is a purely stochastic event or if the enzyme directs it in any way. Random rotational motion of the intermediate would be expected to occur rapidly enough that it would not impede the reaction. However, 50% of the time the intermediate would reseat in the orientation in which it began, so there would be a loss of efficiency. On the contrary, it is difficult to

envision a mechanism by which the enzyme could actively direct the reorientation of the intermediate. This question remains open, although site-directed mutagenesis studies have identified several residues that influence the partitioning between dissociation and production reorientation.[37]

8.12.6   Guanosine Diphosphate-Mannose Pyrophosphorylase

The GMP activity that is used in alginate synthesis is associated with the *algA* gene product, which also catalyzes the PMI reaction. The first indications that AlgA is bifunctional came when the protein was purified.[23] The isomerase and pyrophosphorylase activities co-purified, and their activities remained in constant ratio throughout the purification. Brief treatment with chymotrypsin cleaved approximately 1 kDa from the C-terminus of AlgA; the remaining protein retained GDP-pyrophosphorylase activity, but the phosphomannose isomerase activity was lost.[38]

The bifunctional nature of the AlgA protein can be clearly discerned in the structure of the gene that encodes it (**Figure 5**). The region that encodes amino acids 2–287 is a member of the Pfam family of nucleotidyl transferases, and the region from 298 to 464 is a member of the Pfam family of PMIs. Amino acids 396–404 contain the sequence that was identified as a conserved motif in type M I and II PMIs.[39]

In the alginate pathway the catalytic activities of the AlgA protein flank that of PMM. It is appealing to speculate that the arrangement of the activities in the pathway is a hint that the PMI/GMP and PMM/PGM proteins form a functionally associated complex *in vivo*. Neither the product of the PMI reaction, nor the PMM reaction are unstable or require protection from solvent, and the turnover numbers of the enzymes are slow enough that it seems unlikely that substrate channeling is required to gain a kinetic advantage. However, direct channeling of the products between the enzymes could serve as a mechanism to ensure that the metabolite is not drawn off into another pathway. As tempting as it is to invoke substrate channeling between PMI/GMP and PMM/PGM, it has not been examined experimentally.

8.12.7   Guanosine Diphosphate-Mannose Dehydrogenase

The mannose portion of GDP-mannose is oxidized in an NAD^+-dependent reaction to generate GDP-mannuronate. The reaction is catalyzed by GDP-mannose dehydrogenase (GMD), a member of the small family of enzymes that catalyze the four-electron oxidation of their substrates. The reaction thus requires two equivalents of NAD^+ per catalytic cycle, and proceeds through an intermediate that is at the oxidation state of an aldehyde. GMD is encoded by the *algD* gene in *P. aeruginosa*, and the subunit contains 435 amino acids. Early studies of GMD were conducted by Chakrabarty and co-workers.[40]

Two mechanisms for the NAD^+-dependent oxidation of an alcohol to a carboxylate have been characterized in enzymatic reactions. In the first mechanism, an active-site cysteine plays a crucial role in the reaction. A hydride is transferred to NAD^+ from the alcohol substrate to generate an aldehyde intermediate, then the cysteine thiolate attacks the aldehyde to form a thiohemiacetal intermediate. The thiohemiacetal is oxidized by the second NAD^+ to form a thioester, which is hydrolyzed to generate the carboxylate product. The second mechanism is similar to the first, except that the aldehyde undergoes hydration instead of thiohemiacetal formation. The aldehyde hydrate is oxidized by NAD^+ to form the observed product. This reaction proceeds
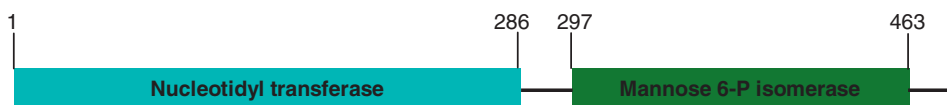
Figure 5   Schematic of the *P. aeruginosa algA* gene encoding the bifunctional protein phosphomannose isomerase-GDP-mannose pyrophosphorylase, indicating the portions of the gene that are recognized as Pfam family members.

without the formation of any covalent enzyme intermediates, and can accommodate, but does not require, participation by a cysteine residue acting as a general base.

Elegant studies of UDP-glucose dehydrogenase have demonstrated convincingly that its reaction proceeds via a thiohemiacetal intermediate. Mutation of the active-site cysteine to serine allowed the ester intermediate to be trapped and characterized by mass spectrometry.[41] On the contrary, histidinol dehydrogenase, which was long assumed to utilize the same catalytic mechanism, was found to be unaffected by mutation of every one of its conserved cysteine residues, suggesting that the reaction proceeded via the aldehyde hydrate.[42] In GMD cysteine 268 resides in the active site with the thiol(ate) group 3–3.5 Å from the carboxylate oxygens in GDP-mannuronic acid. Thus, the residue is positioned appropriately to play a role as a nucleophile, and V_{max} is reduced by 250-fold in the C268A mutant. The reduction of activity demonstrates that C268 plays a critical role in the catalytic reaction, but the mutant retains considerably more activity than the comparable mutant of UDP-glucose dehydrogenase, in which the residual activity was decreased relative to wild type by at least 10 000-fold. The possibility remains that C268 in GMD serves not as a nucleophile to attack the intermediate aldehyde, but as a general base to activate a water molecule for addition to the aldehyde.

The crystal structure of GMD shows that it is a domain-swapped dimer, with residues from each chain contributing to the active sites.[43] The overall topology of the protein is very similar to that of UDP-glucose dehydrogenase,[44] although the two proteins share only 22% sequence identity. Residues 247–267 form a loop that appears to serve as a flap to hold GDP-mannose in the active site. GMD is inactivated by penicillic acid, and the mechanism appears to involve covalent adduct formation that prevents closure of the loop.[45] The side chain of C213 is 3.6 Å from and pointed toward the side chain of N252. A hydrogen bonding interaction between these residues may serve to stabilize the loop in the conformation that helps to hold the substrate at the active site. When N252 is mutated to alanine, V_{max} increases almost twofold, probably because the rate of partially rate-limiting product release has been increased.

The kinetic behavior of GMD is quite complex and displays exquisite sensitivity to reaction conditions including the nature of the buffer and even the order of addition of the substrates.[46] In phosphate buffer GMD exhibits Michaelis–Menten kinetic behavior, and the kinetic mechanism is bi uni uni bi ping-pong, with GDP-mannose binding first and GDP-mannuronate dissociating last. There is a single binding site for the pyridine nucleotide cofactor, so after oxidation of GDP-mannose to the aldehyde, NADH dissociates from the enzyme and is replaced by NAD^+ so the second oxidative step can take place.

GMD exhibits sigmoidal kinetics under most conditions, and it appears that the Michaelis–Menten kinetic behavior that was observed in phosphate buffer arose serendipitously because phosphate acted as an allosteric effector, which decreased the Hill coefficient to unity. A probable physiologically relevant allosteric effector was sought, and GMP was identified. In the absence of GMP the GMD reaction exhibits a Hill coefficient of up to six when NAD^+ is the variable substrate. Increasing concentrations of GMP depress the Hill coefficient so that it is 1 at GMP concentrations above 80 μmol l^{-1}. GMP also causes V_{max} and $K_{1/2}$ for NAD^+ to decrease. Whether GMP actually serves as an *in vivo* regulator of GMD activity is not known, but it would provide a link between the tricarboxylic acid cycle and alginate synthesis by making both processes sensitive to the guanine nucleotide pool in the cell. In the tricarboxylic acid cycle, which functions to provide energy under aerobic conditions, and in an anapleurotic manner to provide many central metabolites, GDP is required by succinyl-CoA synthetase. Alginate biosynthesis places a heavy demand on GTP pools because of the requirement for GDP-mannose, so attenuation of alginate synthesis by GMP would spare guanine nucleotides so that they would be available as required by the tricarboxylic acid cycle.

The quaternary structure of GMD is probably responsible for at least some of the complexity in the kinetic behavior. Under various solution conditions the enzyme behaves as a trimer or a hexamer, as characterized by gel filtration chromatography and dynamic light scattering. The asymmetric unit in the GMD crystal contained four molecules, and a small interface between dimers was observed, which suggests that GMD could form a weak tetramer in solution.[43] Because of the plasticity of the quaternary structure in GMD it has been suggested that it may be an example of the recently characterized 'morpheein' proteins that utilize rearrangements in quaternary structure as a mechanism of allosteric regulation.[47]

8.12.8   Mannuronan Synthesis

The proteins encoded by *alg8* and *alg44* are required for mannuronan synthesis. Although one or both of these proteins have been referred to as alginate polymerases, the designation 'polymerase' is usually reserved for enzymes with processive properties, which has not been demonstrated yet for either protein, and it is probably more helpful to think about mannuronan synthesis as a glycosyltransferase reaction. Much work remains to be done before the reaction can be described in molecular detail. *In vitro* synthesis of mannuronan using purified components has not been demonstrated yet. Sequence analysis of Alg8 identifies it as a member of the family 2 glycosyltransferases, and cellular fractions containing cytoplasmic and outer membrane fractions have been demonstrated to catalyze the formation of mannuronan from GDP-mannuronate. No alginate production was observed in an *alg8* deletion strain, but it was restored when the *alg8* gene was reintroduced, which established the essentiality of Alg8.[48] Family 2 glycosyltransferases catalyze the formation of β-glycosidic bonds with inversion of configuration at the anomeric carbon of the acceptor molecule, consistent with the known configurations of mannuronan and GDP-mannuronate. Alg8 contains several transmembrane helices and a large cytoplasmic domain which contains the active site. Site-directed mutagenesis of residues that are conserved in family 2 glycosyltransferases resulted in inactive protein; in several cases the mutant proteins were unstable.[49]

Strains lacking *alg44* are also unable to synthesize mannuronan.[50] One study has indicated that Alg44 is a periplasmic protein,[50] while a second study using hydropathy analysis and the construction and characterization of PhoA fusion proteins concluded that Alg44 has an N-terminal domain that is cytoplasmic, a single transmembrane helix, and a C-terminal periplasmic domain.[49] The N-terminal portion contains a PilZ domain that binds bis-(3′-5′)-cyclic dimeric GMP (c-di-GMP), which has recently been shown to regulate activities in *P. aeruginosa* that are involved in cytotoxicity and biofilm formation.[51] Mutations in Alg44 that prevented binding of c-di-GMP also resulted in the loss of the ability to synthesize alginate.[52] The exact role of Alg44 remains unclear. The C-terminal portion of the protein shows homology with proteins of multi-drug efflux systems, and deletion of *alg44* causes decreased expression of AlgE, the putative porin-like protein that is involved in secretion of alginate, suggesting that there may be an important physical association between Alg44 and AlgE. The fact that Alg44 binds c-di-GMP would seem to suggest that it plays a regulatory role, but it has been suggested that it may also have a catalytic role in mannuronan synthesis.[49]

8.12.9   C5-Mannuronan Epimerase

Mannuronan epimerase (ME) catalyzes the inversion of the stereochemical configuration at C5 of some of the residues in mannuronan. The epimerization converts a M residue into a G, and also induces a shift in the conformation of the carbohydrate unit from $^4C_1$ to $^1C_4$. Because the change in stereochemistry occurs at C5, the stereochemical designation for the epimerized residue changes from D to L. Also, although the configuration at the anomeric carbon does not change, epimerization at C5 changes the relationship between the anomeric substituent and the substituent at C5 from *cis* to *trans*, so the anomeric designation changes, and the result of the epimerization is to convert M to G .

C5-Mannuronan epimerases have been characterized in detail from *P. aeruginosa* and *Azotobacter vinelandii*. The AlgE enzymes from *A. vinelandii* and the AlgG enzyme from *P. aeruginosa* are quite distinct despite catalyzing the same chemical reaction. In each case, however, several of the same questions are raised. First is the mechanism of the chemical transformation itself, that is, how is the epimerization accomplished? Second, what is the specificity of the enzyme? 'Specificity' encompasses the size of the substrate and the requirements for the identity of the residues neighboring the one that is acted upon. Other fascinating questions include whether the epimerase behaves processively, and what roles are played in the reaction by ancillary proteins.

Epimerization of M is a deceptively simple reaction that requires abstraction of the proton at C5 and reprotonation on the opposite face of the carbohydrate. The transformation is difficult to accomplish because the proton that is abstracted is adjacent to a carboxylate group and therefore has a very high pK, probably >30. In order to make the reaction energetically feasible the carbanion that results from proton abstraction must be

stabilized. Other enzymes that abstract protons alpha to a carboxylate, notably enolase, stabilize the carbanionic intermediate by delocalizing electrons out onto the carboxylate to form an *aci*-acid, which is stabilized by coordination to a metal ion.[53] Although the *A. vinelandii* enzyme is Ca^{2+}-dependent, structural studies discussed below indicate that the metal ions do not interact with the substrate. It is quite unlikely that *aci*-acid formation occurs in the *P. aeruginosa* epimerase reaction, because the enzyme neither contains nor requires a metal ion for catalysis.[54]

An alternative to *aci*-acid formation is to stabilize the carbanion by forming a glycal with unsaturation at the C4–C5 position, transiently cleaving the glycosidic bond with the adjacent residue (**Figure 6**). Testing this mechanism is straightforward since it predicts that the substrate will be cleaved transiently during the course of the reaction. Rapid-mixing chemical quench studies conducted with a size-homogeneous population of mannuronan revealed that indeed, the substrate was converted to a group of shorter oligomers during the reaction. Colorimetric detection of the presumptive glycal intermediate yielded consistent results, and also demonstrated the disappearance of the intermediate over time.[55]

The steady-state kinetic parameters for *P. aeruginosa* epimerase demonstrate that the enzyme acts preferentially on oligomeric or polymer substrates. No reaction was detectable with substrates containing fewer than nine residues, and K_{cat} increased while K_m decreased, as the substrate lengthened to approximately 30 residues. For substrates greater than 30 residues in length K_{cat} and K_m remained relatively constant.[54] The specificity revealed by the steady-state kinetics was consistent with the intermediates characterized in the rapid-mixing chemical quench experiment, which showed that the oligomeric substrate was cleaved near its middle, never near the ends. The physical basis for the size specificity is unclear; that is, how does the enzyme distinguish between substrate molecules that are much larger than the binding site? If each substrate occupies the same binding surface on the enzyme, why do larger substrates react more favorably? The answer may lie in how the solution behavior of the oligosaccharide affects binding and dissociation, or how the enzyme moves from one site of epimerization to the next.
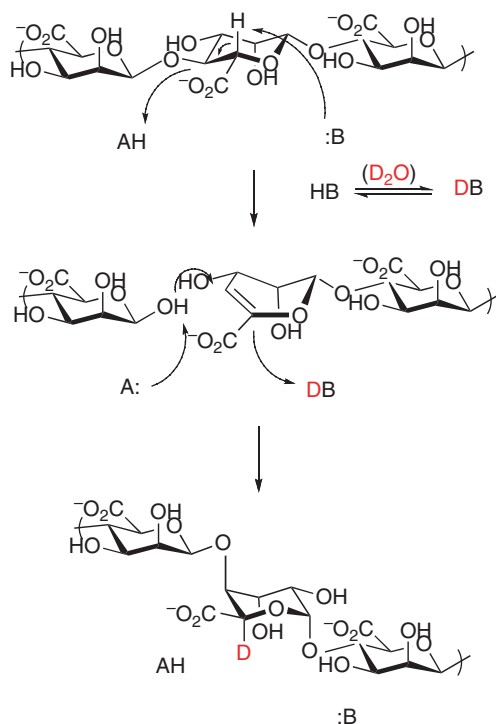
Figure 6　The chemical mechanism of the C5-mannuronan epimerase reaction. The exchange of solvent deuterium into the product is indicated. Note the conformation change in the polymeric substrate that is induced by the epimerization reaction.

The number and distribution of G residues in the alginate polymer must reflect, at some level, the activity and specificity of the mannuronan epimerase. Alginate from *P. aeruginosa* differs from that isolated from *A. vinelandii* and seaweed, in that the G content is lower, and the G residues are randomly distributed throughout the polymer. Alginate from *A. vinelandii* and seaweed has a block structure in which tracts of M residues are interspersed with tracts of G residues, and some regions of the polymer have diads of alternating M and G.

In order to determine the pattern of G incorporation into alginate by *P. aeruginosa* epimerase the product of the reaction was characterized as a function of the extent of the reaction. Although alginate isolated from the lung tissue of CF patients does not contain GG diads, it was observed that *in vitro* the epimerase was capable of producing GG diads and GGG triads.[55] Therefore, the absence of sequential G residues in *P. aeruginosa* alginate is not due to an intrinsic property of the epimerase. Upon extended incubation with mannuronan, ME converted 75% of the residues into Gs. The frequency with which GG diads were introduced in the alginate as a function of the overall extent of incorporation of G residues was examined in order to determine if the enzyme discriminated between M residues that were adjacent to Gs and those that were adjacent to Ms. The pattern of G incorporation that was observed was consistent with random action by the epimerase; in other words, M residues that were adjacent to G residues were just as likely to be acted upon by the epimerase as those that were not adjacent to Gs. It is important to note that these results do not directly address the question of processivity in the epimerase reaction. Processivity refers to the number of catalytic cycles (in this case, epimerizations) that the enzyme catalyzes before dissociating from the substrate. The random pattern of epimerization may indicate that the epimerase is not processive, that is, it dissociates after each catalytic cycle and reassociates at a random location, but it is also consistent with translocation by the enzyme along the alginate polymer for a random distance before catalyzing another epimerization event.

It is interesting that the reaction epimerizes 75% of the M residues, but not more. No solvent deuterium incorporation into M residues during the epimerization reaction could be detected, which indicates that the proton abstraction that is the necessary first chemical step in the reaction is irreversible. However, the enzyme does not epimerize all of the M residues in the substrate. The equilibrium constant for epimerizations is usually close to one for reactions that involve simple substrates, so one might expect the epimerase reaction to reach equilibrium when the M content of the alginate is equal to the G content. The fact that the reaction occurs on a polymeric substrate made up of chiral monomers complicates considerations of the energetics of the reaction somewhat. Nonetheless, it is difficult to reconcile the apparent irreversibility of proton abstraction with the 1:3 M:G ratio in the product, unless one proposes that the conformation of the polymer becomes such that it can no longer bind to the enzyme.

The fact remains that alginate produced *in vitro* by purified *Pseudomonas* epimerase differs in its G content from the alginate isolated from the lungs of CF patients. What, then, is different from how the epimerase acts *in vivo* and how it behaves in the test tube? One difference is that the epimerase appears to be associated with other proteins *in vivo*. This possibility was first suggested by the behavior of *P. aeruginosa* strains in which either of two proteins that are encoded in the *alg* operon, AlgX and AlgK, are knocked out. It was observed that neither *algX(−)* nor *algK(−)* strains secreted alginate, nor did they secrete mannuronan. The media in which the strains were grown contained elevated levels of M and MM dimers.[56,57] These results were explained by proposing that AlgX and AlgK associated with ME in the periplasm and protected alginate from degradation by alginate lyase (AlgL). In the absence of either AlgX or AlgK, it was proposed that the protein conduit through which alginate moved to the outer membrane was lost, and the alginate was degraded.[58]

The structure of the *Pseudomonas* epimerase has not been determined, but a homology model of the C-terminal domain has been described.[59] Sequence analysis clearly shows that the C-terminal portion of the epimerase contains at least six and as many as nine 24-residue repeats which are characteristic of carbohydrate-binding and sugar hydrolase (CASH) domains that are frequently found in proteins that bind or act upon carbohydrates. The CASH domain forms a unique right-handed β-helix that was first characterized in pectate lyase C.[60] A model for *P. aeruginosa* epimerase was generated by homology modeling based on the structure of pectate lyase. One face of the right-handed β-helix forms a long shallow groove that can be envisioned to provide a binding site for alginate. Mutations of residues in the *P. aeruginosa* protein and the homologous protein from *P. fluorescens* that abrogated epimerase activity were all located along one face of the β-helix. The model suggests that the helix is long enough to bind nine uronic acid residues, which corresponds with the experimental observation that oligomers of mannuronan containing fewer than nine residues were not substrates.
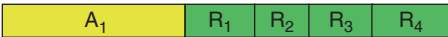
Azotobacter vinelandii has a periplasmic C5-mannuronan epimerase similar to the *P. aeruginosa* enzyme discussed above. In addition, *A. vinelandii* encodes a family of seven C5-mannuronan epimerases designated AlgE1–7 that are secreted. The AlgE epimerases do not show significant sequence homology with AlgG and they require Ca^{2+} for activity. Extensive and elegant work by groups at the Norwegian Biopolymer Laboratory and the Department of Biotechnology at the Norwegian University of Science and Technology has been done to characterize the AlgE epimerases[61,62]. Each protein is made up of one or more of each of two modules, the A module, which is approximately 385 residues in length, and is sufficient for catalysis, and the R module, which is about 150 residues long, and has no catalytic activity by itself, but enhances catalysis when fused to A modules. Recombinant A modules and R modules each bind Ca^{2+}, so it is likely that the wild-type proteins bind multiple Ca^{2+} ions.[63]

The modular structures of AlgE1–7 are shown in **Figure 7**, along with the nature of the epimerization product. Although all of the enzymes except AlgE4 produce both G blocks and MG blocks, the lengths of the blocks and their relative distributions differ among the different enzymes. AlgE4 produces MG blocks predominantly while the other AlgE enzymes can form G and MG blocks.[64]

The biological function of the different epimerases is not certain. However, it is known that *A. vinelandii* can form metabolically inactive cysts, and that alginate is the primary carbohydrate constituent of the cyst. Alginates with different compositions have different physical properties, so the multiple AlgE epimerases may allow the bacteria to tailor the mechanical properties of alginate for different functions in different locations in the cyst.

Since the A modules alone exhibit catalytic activity and give rise to distinct products that match those formed by the intact multimodule proteins, it appears that the determinants for the nature of the product residue is in the A module. This expectation was borne out by studies in which the A modules from AlgE2 and AlgE4 were each divided into nine regions, and swapped in over 40 different combinations.[65] The hybrid enzymes were all catalytically active and produced alginates that were intermediate in their block structures compared to those formed by the parent proteins. The primary determinant of the block structure resided in one region of the A module where 15 out of 48 residues differed between AlgE4 and AlgE2. However, it was observed that other regions of the A module also influenced the structure of the product.

One of the interesting features of the epimerase reaction is the potential for processivity. After the enzyme catalyzes the epimerization of one residue in the polymeric substrate, it could dissociate and rebind at another location, or it could translocate to another residue without releasing the polymer. To investigate whether AlgE operated in a processive manner, the composition of the product as a function of the extent of epimerization of the polymer was characterized.[66] The experimental results, determined by ^1H and ^{13}C NMR spectroscopy,

Figure 7 Schematic of the AlgE family of C5-mannuronan epimerases. The active site for epimerization is in the A module. F_{GG} is the fraction of product containing GG diads, and F_{MG}, F_{GM} is the fraction of product containing MG (or GM) diads; data are taken from H. Ertesvag; H. K. Hoidal; H. Schjerven; B. I. G. Svanem; S. Valla, *Metab. Eng.* **1999**, *1*, 262–269.

were compared with Monte Carlo simulations of the reaction occurring by a processive process, by a nonprocessive process in which the binding of enzyme and substrate at each round of catalysis occurred without any discrimination for the neighboring residues (random attack), and a preferred attack model, which was not processive, but in which the binding affinity between enzyme and substrate was dependent on the identification of neighboring residues. It was striking that the simulations showed that neither the time courses for the reactions nor the frequency of appearance of different triads allowed for discrimination between the processive and preferred attack models. However, the authors suggested that characterization of the mean length of the MG or G blocks resulting from the action of the epimerase could be used to determine the mechanism of epimerase action. In the case of AlgE4, which forms MG blocks, the authors suggested that the data most closely matched the simulations from the processive model. The data from the AlgE2 reaction, which forms G blocks, did not closely match the outputs of any of the models.

In a second study, the products derived from AlgE4 action on polymannuronan were degraded with an alginate lyase that was specific for cleaving the substrate adjacent to G residues.[67] It was reasoned that since AlgE4 produces predominantly MG blocks, degradation of the product at early stages of the epimerization reaction would provide insight into the mode of action of the epimerase. If the epimerase acted in a processive manner the product would contain strings of alternating MG residues interspersed between blocks of Ms. Upon degradation with the G-specific lyase, a distribution of products consisting of dimers and larger oligomers would result. If, however, the epimerase introduced a single G, dissociated from the polymer and then rebound to catalyze epimerization at another location, degradation of the polymer with G-specific lyase would produce a random distribution of oligomers. A nonrandom size distribution of products resulting from degradation of the mannuronan that had been acted upon by AlgE4 was observed, and it was suggested that the enzyme epimerized about 10 residues each time it bound to the substrate before dissociating. A caveat that would confound the interpretation of these results is that if the enzyme dissociated from the polymer at each round of catalysis but exhibited a strong preference for binding adjacent to a G residue, the same product distribution would be predicted.

There is no experimentally determined structure available for a complete C5-mannuronan epimerase. However, the structure of the R module from *A. vinelandii* AlgE4 epimerase has been determined by NMR spectroscopy[68] (PDB code 2agm), and the A module has been determined by X-ray crystallography (PDB codes 2pyh and 2pyg). Both domains contain CASH domains.

The R module of AlgE4 has a β-roll structure, in this case made up of 11 short β-strands. It was found that the protein was both unstable and insoluble when Ca^{2+} was removed. Therefore, the Ca^{2+}-binding sites were identified by titration with the paramagnetic lanthanide Tm^{3+}. These studies revealed that the metal binding sites were in the loops that connected the β-strands. The A module has not been described in the literature yet, but two structures are available in the PDB. Essentially, the entire structure is a right-handed β-helix. One face forms a shallow groove that is lined with arginine and lysine residues to form an electrostatically positive surface to which the anionic substrate binds (**Figure 8**). Several Ca^{2+} ions are visible in the structure. Most of them are associated with loops in the protein and do not appear to be involved in substrate binding.

8.12.10   Alginate Acetylation

Alginate produced by *Azotobacter* and *Pseudomonas* is acetylated at the 2-OH and/or 3-OH. Not every residue is acetylated, and very little is known about the process. Since acetylation occurs in the periplasm where there is no acetyl-CoA, an important question is whether the acetyl groups derive from acetyl-CoA, and if so, how they are delivered in a chemically activated form to the site of alginate acetylation. Strains of *P. aeruginosa* that do not acetylate alginate have diminished virulence,[69,70] so there is a clinical imperative for understanding acetylation.

The genes for the proteins required for alginate acetylation have been identified by analysis of deletion mutants; strains in which *algF*, *algI*, or *algJ* have been deleted secrete unacetylated alginate.[71,72] Sequence analysis sheds little light on the roles of the AlgF, AlgI, or AlgJ proteins; none of them are recognized as belonging to a characterized protein superfamily.

Hydropathy analysis indicates that AlgI contains at least 10 transmembrane helices, and experimentally it has been demonstrated to be associated with the inner membrane. These properties suggest that AlgI may be
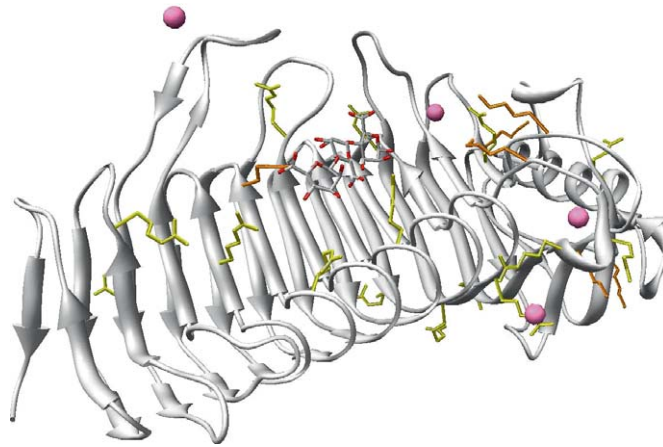
Figure 8 Structure of the A module of AlgE4 (PDB code 2agm). A trisaccharide is bound in the shallow groove formed by the β-helix. Lysine (orange) and arginine (yellow) residues believed to play an important role in binding the anionic substrate. Calcium ions are shown as purple spheres.

involved in the transport of the acetyl group from the cytoplasm to the periplasmic space. AlgI shares 20% sequence identity with DltB, a protein from *Lactobacillus* whose exact function is not known, but which is involved in lipoteichoic acid biosynthesis. It has been suggested that DltB serves as a transport protein to move alanyl residues from alanylated acyl carrier protein across the membrane,[73] and one could propose an analogous role in acetyl transfer for AlgI.

AlgF and AlgJ are localized in the periplasm. AlgJ has a signal sequence at its N-terminal end that is apparently uncleaved, and it has been suggested that AlgJ is associated with the periplasmic side of the inner membrane.

8.12.11 Alginate Lyase

Paradoxically, alginate synthesis in *P. aeruginosa* requires functional alginate lyase.[74] The gene encoding alginate lyase, *algL*, is located within the *alg* operon and its deletion causes a loss of the ability to secrete alginate. Alginate appears to accumulate in the periplasmic space in the deletion mutants, ultimately causing rupture of the cell wall.[58]

Alginate lyase catalyzes cleavage of the glycosidic bond through β-elimination. The similarity between the lyase reaction and the C-5 mannuronan epimerase reaction has been noted.[75] Interestingly, the proteins share little sequence homology.

A number of alginate lyases from different species with distinct substrate specificities have been characterized, and are very useful for analyzing the composition of alginate.[76] *Pseudomonas aeruginosa* alginate lyase is specific for cleavage of the glycosidic bond connecting M residues.[77] Incubation of mannuronan with *P. aeruginosa* alginate lyase results in the production of trimeric oligomannuronate predominantly, suggesting that the enzyme binds near the terminus of the polymeric substrate. The enzyme was reported to exhibit maximal activity with hexameric oligomannuronate, although the activity with oligomeric substrates was not compared directly with mannuronan.[78]

One model has been put forward to explain the role of alginate lyase in alginate synthesis. Accumulating evidence suggests that a multiprotein complex composed of AlgG, AlgK, and or AlgX and perhaps AlgL forms a conduit in the periplasmic space through which the polymer is sheparded on its way to the outer membrane for secretion. Bakkevig *et al.*[79] have suggested that AlgL is required to degrade any polymer that fails to traverse the conduit successfully and is stranded in the periplasmic space.

8.12.12   Secretion of Alginate

The *alg* operon encodes a protein called AlgE, which is an integral outer membrane protein that has characteristics of an anion-specific channel protein.[80] Interactions between AlgE and Alg44 have been reported, suggesting that AlgE may be physically associated with the complex of proteins involved in synthesis of the alginate polymer and transport through the periplasmic space.[49]

8.12.13   Regulation of Alginate Synthesis

Alginate production (mucoidy) is typically not a stable phenotype in *P. aeruginosa*, and a great deal of effort has gone into characterizing the regulation of alginate synthesis. The lung tissue of a CF patient is colonized initially by bacteria that do not produce alginate. However, in response to some environmental signal alginate production is turned on. The onset of alginate production coincides with a deteriorating prognosis for the patient. Mucoid *P. aeruginosa* cells that are isolated from CF lung tissue frequently revert to nonmucoid status upon culturing.

All of the genes for alginate synthesis except *algC* are under the control of the *algD* promoter. Transcription requires the alternative σ factor σ^{22}, also known as AlgT or AlgU.[81] The activity of σ^{22} is regulated by the MucA and MucB proteins. The mucoid phenotype has been correlated with the accumulation of mutations in the *mucAB* genes. MucA contains a single transmembrane helix, and it is believed to span the periplasmic membrane so that it can interact with MucB in the periplasm and σ^{22} in the cytoplasm. Active MucA and MucB lead to rapid turnover of σ^{22}. When frameshift or deletion mutations in *mucAB* result in the loss of functional MucA and MucB, σ^{22} activates the transcription of the *algR* and *algB* genes, as well as the genes in the alginate operon.[82] AlgR and AlgB bind to the *algD* promoter and further activate transcription.[83,84]

8.12.14   Future Directions

It is clear that many fundamental questions remain about alginate biosynthesis, primarily in the latter stages of the pathway. The functions of accessory proteins such as AlgX and AlgK remain to be defined in detail. Available evidence suggests that multiple Alg proteins interact in the periplasm and in association with the inner and outer membranes. It is unknown how the enzymes that act upon the alginate polymer in mutually exclusive ways to catalyze epimerization, acetylation, or cleavage of the glycosidic bond, are all able to access the substrate. The mechanism by which directionality in alginate secretion has not been addressed. Coming to grip with these and other issues will require developing methods for reconstituting biologically relevant multiprotein complexes *in vitro*, and the difficult task of working with membrane proteins.

The observations which suggest that the recently characterized regulatory molecule cyclic dimeric GMP may be involved in alginate biosynthesis are exciting, and exploration of this topic will likely lead to new insights into the initiation and establishment of mucoidy. This may provide another link between the complex topics of capsule production, cell-to-cell signaling, and virulence. Infections caused by *P. aeruginosa* continue to be a major health hazard, and one can hope that careful investigation of these topics will provide hints of how to treat these largely antibiotic-resistant infections.

Bacterial biofilm formation in *Pseudomonas* and many other bacterial species has been recognized as an important aspect of the microbial lifestyle, and one that has direct implications for the treatment of infections and the efficiency of many industrial processes. Although alginate was once considered to be a major constituent of the biofilm it is now recognized that *Pseudomonas* biofilms can form without alginate, and other polysaccharides such as those produced by the enzymes encoded in the *pel* and *psl* operons may play more important roles in many circumstances.[85,86] Although the role of alginate in biofilms may be more circumscribed than once thought, it remains the best experimental system available for characterizing the complex and important process of polysaccharide biosynthesis and export.

Abbreviations

CF	cystic fibrosis
dTDP	deoxythymidine diphosphate
G	guluronate
GDP	guanosine diphosphate
GMD	GDP-mannose dehydrogenase
GMP	GDP-mannose pyrophosphorylase
GTP	guanosine triphosphate
LPS	lipopolysaccharide
M	mannuronate
NAD	nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
PMI	phosphomannose isomerase
PMM/PGM	phosphomannomutase/phosphoglucomutase
UDP	uridine diphosphate

References

1. D. M. Ramsey; D. J. Wozniak, *Mol. Microbiol.* **2005**, *56*, 309–322.
2. G. G. Zhanel; M. DeCorby; N. Laing; B. Weshnoweski; R. Vashisht; F. Tailor; K. A. Nichol; A. Wierzbowski; P. J. Baudry; J. A. Karlowsky; P. Lagace-Wiens; A. Walkty; M. McCracken; M. R. Mulvey; J. Johnson; Canadian Antimicrobial Resistance Alliance (CARA); D. J. Hoban, *Antimicrob. Agents Chemother.* **2008**, *52*, 1430–1437.
3. L. R. Evans; A. Linker, *J. Bacteriol.* **1973**, *116*, 915–924.
4. T. A. Davis; F. Llanes; B. Volesky; A. Mucci, *Environ. Sci. Technol.* **2003**, *37*, 261–267.
5. H. Grasdalen; B. Larsen; O. Smidsrod, *Carbohydrate Res.* **1979**, *68*, 23–31.
6. P. J. Tatnell; N. J. Russell; J. R. W. Govan; P. Gacesa, *Biochem. Soc. Trans.* **1996**, *24*, 404S.
7. C. A. Steginsky; J. M. Beale; H. G. Floss; R. M. Mayer, *Carbohydrate Res.* **1992**, *225*, 11–26.
8. T. A. Davis; B. Volesky; R. H. S. F. Vieira, *Water Res.* **2000**, *34*, 4270–4278.
9. E. Fourest; B. Volesky, *Environ. Sci. Technol.* **1996**, *30*, 277–282.
10. R. B. Parad; C. J. Gerard; D. Zurakowski; D. P. Nichols; G. B. Pier, *Infect. Immun.* **1999**, *67*, 4744–4750.
11. D. A. Rees; E. J. Welsh, *Angew. Chem. Int. Ed. Engl.* **1977**, *16*, 214–224.
12. I. Braccini; S. Perez, *Biomacromolecules* **2001**, *2*, 1089–1096.
13. H. Grasdalen, *Carbohydr. Res.* **1983**, *118*, 255–260.
14. A. Heyraud; C. Gey; C. Leonard; C. Rochas; S. Girond; B. Kloareg, *Carbohydr. Res.* **1996**, *289*, 11–23.
15. J. M. Beale, Jr.; J. L. Foster, *Biochemistry* **1996**, *35*, 4492–4501.
16. A. R. Lynn; J. R. Sokatch, *J. Bacteriol.* **1984**, *158*, 1161–1162.
17. A. J. Anderson; A. J. Hacking; E. A. Dawes, *J. Gen. Microbiol.* **1987**, *133*, 1045–1052.
18. S. Shankar; R. W. Ye; D. Schlictman; A. M. Chakrabarty, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1995**, *70*, 221–255.
19. C. E. Chitnis; D. E. Ohman, *Mol. Microbiol.* **1993**, *8*, 583–590.
20. H. L. Rocchetta; J. C. Pacan; J. S. Lam, *Mol. Microbiol.* **1998**, *29*, 1419–1434.
21. R. W. Ye; N. A. Zielinski; A. M. Chakrabarty, *J. Bacteriol.* **1994**, *176*, 4851–4857.
22. A. E. I. Proudfoot; G. Turcatti; T. N. C. Wells; M. A. Payton; D. J. Smith, *Eur. J. Biochem.* **1994**, *219*, 415–423.
23. D. Shinabarger; A. Berry; T. B. May; R. Rothmel; A. Fialho; A. M. Chakrabarty, *J. Biol. Chem.* **1991**, *266*, 2080–2088.
24. W. J. Albery; J. R. Knowles, *Biochemistry* **1976**, *15*, 5627–5631.
25. K. N. Allen; A. Lavie; G. K. Farber; A. Glasfeld; G. A. Petsko; D. Ringe, *Biochemistry* **1994**, *33*, 1481–1487.
26. J. P. Richard; J. Crugeiras; R. W. Nagorski, *J. Phys. Org. Chem.* **1998**, *11*, 512–518.
27. S. A. Sousa; L. M. Moreira; J. Wopperer; L. Eberl; I. Sa-Correia; J. H. Leitao, *Biochem. Biophys. Res. Commun.* **2007**, *353*, 200–206.
28. C. Roux; J. H. Lee; C. J. Jeffery; L. Salmon, *Biochemistry* **2004**, *43*, 2926–2934.
29. K. D. Collins, *J. Biol. Chem.* **1974**, *249*, 136–142.
30. K. N. Allen; A. Lavie; G. A. Petsko; D. Ringe, *Biochemistry* **1995**, *34*, 3742–3749.
31. N. E. Naught; P. A. Tipton, *Arch. Biochem. Biophys.* **2001**, *396*, 111–118.
32. M. R. Atkinson; E. Johnson; R. K. Morton, *Biochem. J.* **1961**, *79*, 12–15.
33. L. Naught; P. A. Tipton, *Biochemistry* **2005**, *44*, 6831–6836.
34. W. J. J. Ray; J. W. Long, *Biochemistry* **1976**, *15*, 3993–4006.
35. C. Regni; P. A. Tipton; L. J. Beamer, *Structure* **2002**, *10*, 269–279.
36. C. Regni; L. Naught; P. A. Tipton; L. J. Beamer, *Structure* **2004**, *12*, 1–20.
37. C. Regni; A. M. Schramm; L. J. Beamer, *J. Biol. Chem.* **2006**, *281*, 15564–15571.
38. T. B. May; D. Shinabarger; A. Boyd; A. M. Chakrabarty, *J. Biol. Chem.* **1994**, *269*, 4872–4877.

39. S. O. Jensen; P. R. Reeves, *Biochim. Biophys. Acta* **1998**, *1382*, 5–7.
40. S. Roychoudhury; T. B. May; J. F. Gill; S. K. Singh; D. S. Feingold; A. M. Chakrabarty, *J. Biol. Chem.* **1989**, *264*, 9380–9385.
41. X. Ge; R. E. Campbell; I. van de Rijn; M. E. Tanner, *J. Am. Chem. Soc.* **1998**, *120*, 6613–6614.
42. H. Teng; E. Segura; C. Grubmeyer, *J. Biol. Chem.* **1993**, *268*, 14182–14188.
43. C. F. Snook; P. A. Tipton; L. J. Beamer, *Biochemistry* **2003**, *42*, 4658–4668.
44. R. E. Campbell; S. C. Mosimann; I. van De Rijn; M. E. Tanner; N. C. Strynadka, *Biochemistry* **2000**, *39*, 7012–7023.
45. J. L. Kimmel; P. A. Tipton, *Arch. Biochem. Biophys.* **2004**, *441*, 132–140.
46. L. E. Naught; S. Gilbert; R. Imhoff; C. Snook; L. Beamer; P. Tipton, *Biochemistry* **2002**, *41*, 9637–9645.
47. E. K. Jaffe, *Trends Bio. Sci.* **2005**, *30*, 490–497.
48. U. Remminghorst; B. H. A. Rehm, *Appl. Environ. Microbiol.* **2006**, *72*, 298–305.
49. L. L. Oglesby; S. Jain; D. E. Ohman, *Microbiology* **2008**, *154*, 1605–1615.
50. U. Remminghorst; B. H. Rehm, *FEBS Lett.* **2006**, *580*, 3883–3888.
51. H. Kulasakara; V. Lee; A. Brencic; N. Liberati; J. Urbach; S. Miyata; D. G. Lee; A. N. Neely; M. Hyodo; Y. Hayakawa; F. M. Ausubel; S. Lory, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2839–2844.
52. M. Merighi; V. T. Lee; M. Hyodo; Y. Hayakawa; S. Lory, *Mol. Microbiol.* **2007**, *65*, 876–895.
53. R. R. Poyner; W. W. Cleland; G. H. Reed, *Biochemistry* **2001**, *40*, 8009–8017.
54. A. Jerga; A. Raychaudhuri; P. A. Tipton, *Biochemistry* **2006**, *45*, 552–560.
55. A. Jerga; M. D. Stanley; P. A. Tipton, *Biochemistry* **2006**, *45*, 9138–9144.
56. S. Jain; D. E. Ohman, *J. Bacteriol.* **1998**, *180*, 634–641.
57. A. Robles-Price; T. Y. Wong; H. Sletta; S. Valla; N. L. Schiller, *J. Bacteriol.* **2004**, *186*, 7369–7377.
58. S. Jain; D. E. Ohman, *Infect. Immun.* **2005**, *73*, 6429–6436.
59. S. A. Douthit; M. Dlakic; D. E. Ohman; M. J. Franklin, *J. Bacteriol.* **2005**, *187*, 4573–4583.
60. M. D. Yoder; N. T. Keen; F. Jernak, *Science* **1993**, *260*, 1503–1507.
61. H. Ertesvag; H. K. Hoidal; H. Schjerven; B. I. G. Svanem; S. Valla, *Metab. Eng.* **1999**, *1*, 262–269.
62. M. Hartmann; A. S. Duun; S. Markussen; H. Grasdalen; S. Valla; G. Skjak-Braek, *Biochim. Biophys. Acta* **2002**, *1570*, 104–122.
63. H. Ertesvag; S. Valla, *J. Bacteriol.* **1999**, *181*, 3033–3038.
64. H. K. Hoidal; H. Ertesvag; G. Skjak-Braek; B. T. Stokke; S. Valla, *J. Biol. Chem.* **1999**, *274*, 12316–12322.
65. T. M. Bjerkan; B. E. Lillehov; W. I. Strand; G. Skjak-Braek; S. Valla; H. Ertesvag, *Biochem. J.* **2004**, *381*, 813–821.
66. M. Hartmann; O. B. Holm; G. A. B. Johansen; G. Skjak-Braek; B. T. Stokke, *Biopolymers* **2002**, *63*, 77–88.
67. C. Campa; S. Holtan; N. Nilsen; T. M. Bjerkan; B. T. Stokke; G. Skjak-Braek, *Biochem. J.* **2004**, *381*, 155–164.
68. F. L. Aachmann; B. I. G. Svanem; P. Guntert; S. B. Petersen; S. Valla; R. Wimmer, *J. Biol. Chem.* **2006**, *281*, 7350–7356.
69. D. E. Nivens; D. E. Ohman; J. Williams; M. J. Franklin, *J. Bacteriol.* **2001**, *183*, 1047–1057.
70. G. B. Pier; F. Coleman; M. Grout; M. Franklin; D. E. Ohman, *Infect. Immun.* **2001**, *69*, 1895–1901.
71. M. J. Franklin; D. E. Ohman, *J. Bacteriol.* **1996**, *178*, 2186–2195.
72. M. J. Franklin; D. E. Ohman, *J. Bacteriol.* **2002**, *184*, 3000–3007.
73. M. Y. Kiriukhin; F. C. Neuhaus, *J. Bacteriol.* **2001**, *183*, 2051–2058.
74. M. T. Albrecht; N. L. Schiller, *J. Bacteriol.* **2005**, *187*, 3869–3872.
75. P. Gacesa, *FEBS Lett.* **1987**, *212*, 199–202.
76. T. Y. Wong; L. A. Preston; N. L. Schiller, *Annu. Rev. Microbiol.* **2000**, *54*, 289–340.
77. A. Linker; L. R. Evans, *J. Bacteriol.* **1984**, *159*, 958–964.
78. B. H. A. Rehm, *FEMS Microbiol. Lett.* **1998**, *165*, 175–180.
79. K. Bakkevig; H. Sletta; M. Gimmestad; R. Aune; H. Ertesvag; K. Degnes; B. E. Christensen; T. E. Ellingsen; S. Valla, *J. Bacteriol.* **2005**, *187*, 8375–8384.
80. B. H. Rehm; G. Boheim; J. Tommassen; U. K. Winkler, *J. Bacteriol.* **1994**, *176*, 5639–5647.
81. D. W. Martin; B. W. Holloway; V. Deretic, *J. Bacteriol.* **1993**, *175*, 1153–1164.
82. D. W. Martin; M. J. Schurr; M. H. Mudd; J. R. W. Govan; B. W. Holloway; V. Deretic, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8377–8381.
83. J. Kato; A. M. Chakrabarty, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 1760–1764.
84. A. J. Leech; A. Sprinkle; L. Wood; D. J. Wozniak; D. E. Ohman, *J. Bacteriol.* **2008**, *190*, 581–589.
85. L. Friedman; R. Kolter, *J. Bacteriol.* **2004**, *186*, 4457–4465.
86. L. Ma; K. D. Jackson; R. M. Landry; M. Parsek; D. J. Wozniak, *J. Bacteriol.* **2006**, *188*, 8213–8221.

Biographical Sketch

Peter A. Tipton received his undergraduate degree in chemistry from Cornell University, and his Ph.D. in biochemistry at the University of Wisconsin-Madison, where he worked with Professor W. W. Cleland on the mechanism of biotin carboxylase. He spent a postdoctoral stint in Professor Jack Peisach's laboratory at the Albert Einstein College of Medicine, where he used pulsed EPR methods to study active-site geometries in Mn^{2+}-utilizing enzymes. He assumed a position at the University of Missouri, where he is currently professor of biochemistry. His research interests are in metabolic pathways and enzyme mechanisms in bacterial pathogens, and enzyme inhibitor design.

8.13 Enzymology of Bacterial Resistance

Jed F. Fisher and Shahriar Mobashery, University of Notre Dame, IN, USA

8.13.1 Introduction

Notwithstanding incredible discoveries in both biology and chemistry, the scientific assessment of the Age of Antibiotics as it enters its seventh decade is whether present events will direct this Age to crisis and then extinction, or through crisis to renaissance.[1–6] The events deciding between these two fates for the antibiotics are diverse. The foremost of these is the realization that the clinical use of the antibiotic is the concurrent clinical selection, from among the extraordinary, diverse, and robust counter-responses devised over the eons of times by microorganisms,[7–9] of decisive biochemical methods for the neutralization of the efficacy of the antibiotic. To use an antibiotic is to condemn it to eventual obsolescence.[1,10] The velocity with which the antibiotic succumbs to obsolescence is determined by a host of factors including policy at the medical (how should antibiotics be used?),[11–19] chemical (are there new biologically active natural products to be discovered[20–23] and new biochemical targets for screening identification of new synthetic structure?),[24–27] and business (how can the economics of antibiotic discovery be transformed to encourage innovation?)[28–32] While credible proposals exist for the extension of the Age of Antibiotics into the foreseeable future, the deliberate and inexorable but seemingly evolutionary in terms of time, movement of anti-infective resistance mechanisms into new microorganisms and new settings have failed to galvanize action. The progression, for example, of vancomycin resistance into the Enterococci, the transition of β-lactam-resistant *Staphylococcus aureus* from the hospital into the community, and the spread among the Gram-negative bacteria of mechanistically perfected hydrolytic enzymes for β-lactam antibiotic destruction, are documented. The 'superbugs' are, quite literally, upon us.[33–39] Nonetheless, in the absence of pandemic neither existing resources nor existing reward have elevated the problem beyond recognized scientific and medical concern, to societal recognition to support the creation (and preservation) of new anti-infectives as a long-term research endeavor of both difficulty and importance.

Within the scientific and medical communities, the challenges associated with this endeavor are well appreciated. It is not simply that the resistance mechanisms are already evolutionarily selected, and that the barrier is only one of acquisition and distribution. This circumstance alone presents challenge. The problem is deeper. Anti-infective resistance is intimately connected to pathogen virulence,[22,40–42] and thus anti-infective use selects for both phenomena, as evidenced by the increasing appearance in the community of β-lactam (methicillin)-resistant *S. aureus* (MRSA) expressing Panton-Valentine leukocidin cytolytic peptide virulence factors.[39,43,44] Moreover, successful resistance to anti-infectives is often the result of selection of an ensemble of adjustments, including especially reduced ingress of the anti-infective agent, active anti-infective egress by transporters, structural adjustment at the target site to diminish the effectiveness of the anti-infective, and structural modification of the anti-infective to diminish its recognition at the target. There is fitness cost to the

microorganism for some of these adjustments, but negligible cost for many. Of these adjustments, the boldest strokes are enzymatic: enzymatic transformation of the target, and/or of the anti-infective, to neutralize the mechanism exerted against the microorganism by the anti-infective. These bold strokes, as used by the clinically useful antibacterial agents against the major human bacterial pathogens, are the focus of this review. Our emphases within this review are the key structural aspects relevant to bacterial resistance, as seen by the enzymology of target modification and by anti-infective modification of the antibacterial by resistance enzymes. The clinical aspects of this medical challenge – which necessarily encompass the entirety of the adjustments used by bacterial pathogens – have been reviewed recently for many representative bacterial pathogens including *Escherichia coli*,[45,46] *Neisseria* spp.,[47] *Pseudomonas aeruginosa*,[48–50] *Shigella* spp.,[51] *Campylobacter* spp.,[52] *Haemophilus influenzae*,[53] *Bacteroides* spp.,[54] *Acinetobacter baumanni*,[55,56] *Salmonella* and *Streptococcus* spp.,[57,58] *Clostridium difficile*,[59] *Helicobacter pylori*,[60] *Enterococci* spp.,[33,61,62] and *S. aureus*.[39,43,63–69]

A review of the enzymatic resistance mechanisms used by bacterial pathogens may be organized from the perspective of: microbiology, the genetics of the antibacterial resistome,[7,8,70–75] the enzyme mechanism used to modify the antibacterial,[76] and the antibacterial structural class. This latter perspective is the organizational basis of this review. The antibacterial classes discussed are the primary antibacterials in clinical use.[77] These antibacterials are the β-lactams (penicillins, cephalosporins, carbapenems, and β-lactamase inhibitors), the glycopeptides (vancomycin), the aminoglycosides, the quinolones, and the macrolides. The enzymes relevant to antibacterial resistance to these sub-classes of anti-infectives represent both target modification (β-lactams, aminoglycosides, quinolones) and target modification (β-lactams, glycopeptides, and macrolides).
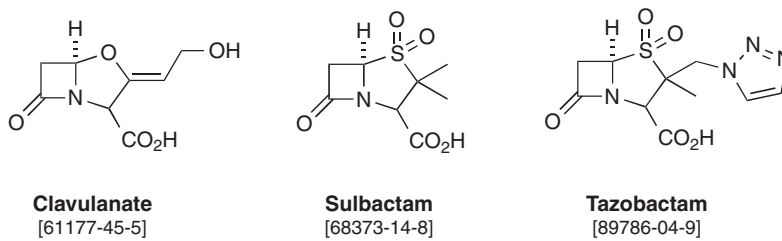
8.13.2   Enzymatic Basis for β-Lactam Resistance

8.13.2.1   Overview

The β-lactam antibiotics have been used clinically for over 60 years, and remain to this day the dominant class of antibacterials. During this interval, medicinal chemistry development of the β-lactams has progressed through the penicillin, cephalosporin, cephamycin, and carbapenem sub-classes of the β-lactams, with the consequence that the β-lactam structures with potent biological activity represent a rich structural diversity around the eponymous β-lactam (bicyclic structures with a 2-azetidinone core: **Figure 1**). A considerable driving force to their structural devlopment is, however, the truly extraordinary resistance responses to these antibiotics that have developed over this identical time interval and are now widely disseminated – continuing even as this chapter is written – among bacterial pathogens. The value of the β-lactams is their safety and efficacy as bacteriocidal agents, targeting a uniquely prokaryotic structural entity (the bacterial cell wall) that is the cytoskeleton of bacteria. Eukaryotes have no such structure. For the Gram-positive bacteria, the cell wall is a polymeric exoskeleton. Underneath this is the periplasmic space and the cell membrane, all surrounding the cell cytoplasm. For the Gram-negative bacteria, the cell wall is a polymeric endoskeleton located underneath the outer membrane, and above (in order) the periplasmic space, inner cell membrane, and cytoplasm. The chemical structures of the Gram-negative and Gram-positive cell walls are nearly identical: each consists of relatively rigid polysaccharide strands interconnected by peptide stems. Because of cross-linking of the peptide stems of each glycan strand, a single peptidoglycan polymer is biosynthesized as the cell wall edifice. Any impression of stasis with respect to this polymer is wholly mistaken. As bacteria grow, so must their cell walls and membranes grow in parallel; as bacteria divide, so must their cell walls and membranes partition; as bacteria depend on functional proteins such as sensors, transporters, and pores for solute ingress and egress, so must their cell walls provide gaps and structural supports for these proteins. The enzyme catalysts of peptidoglycan biosynthesis and peptidoglycan remodeling are the penicillin binding proteins (the PBPs).[78,79] This terminology – that of 'penicillin binding proteins' – is historical. The realization that these membrane-bound proteins were the target of the β-lactam antibacterials preceded the recognition that the target of the β-lactams was not mere proteins, but enzymes integral to cell wall biosynthesis. Each bacterium has a family of PBPs, and each β-lactam targets (depending on its structure) one or more of this family. With respect to cell wall biosynthesis, the most important PBPs are the bifunctional PBPs. The first function of these PBPs is transglycosylase assembly of the glycan strands from the immediate disaccharide precursor (Lipid II, shown in **Figure 2**),[80] and the second function is the transpeptidase linking of the peptide stems of adjacent glycan strands. The two
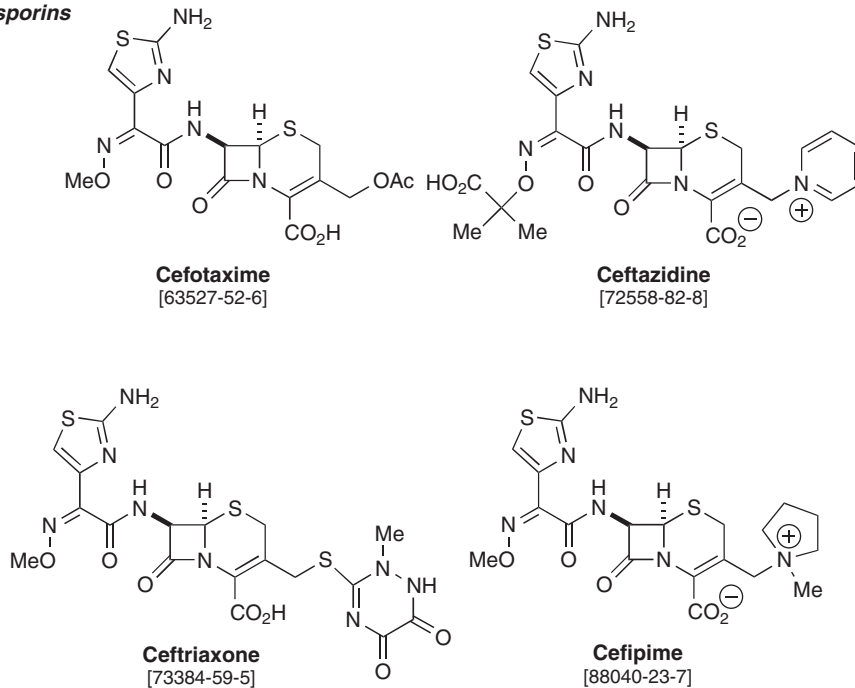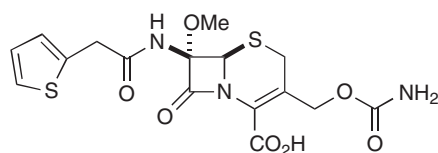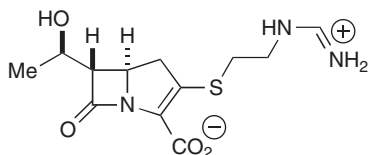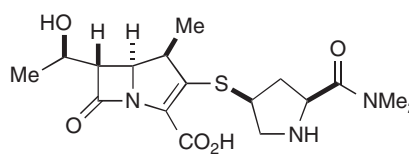
Penicillins
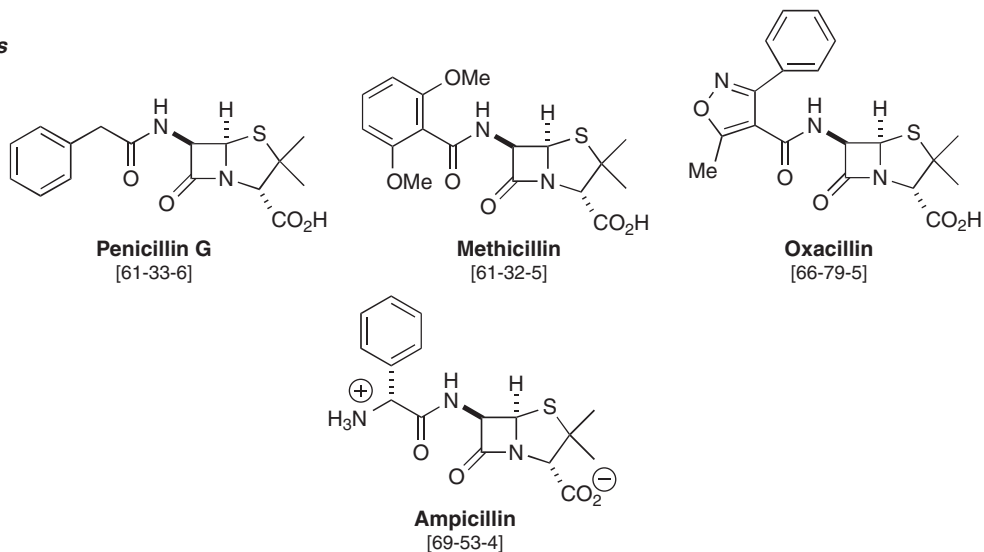
Penicillin G
[61-33-6]

Methicillin
[61-32-5]

Oxacillin
[66-79-5]

Ampicillin
[69-53-4]

*β-Lactamase
Inactivators*

Clavulanate
[61177-45-5]

Sulbactam
[68373-14-8]

Tazobactam
[89786-04-9]

*Third-generation
Cephalosporins*

Cefotaxime
[63527-52-6]

Ceftazidine
[72558-82-8]

Ceftriaxone
[73384-59-5]

Cefipime
[88040-23-7]

Figure 1   (Continued)
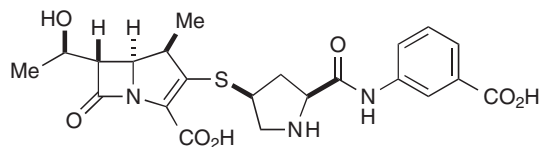
Figure 1 The structures of representative β-lactams including penicillins, inactivators of β-lactamases, third-generation cephalosporins, a cephamycin, and carbapenems.

domains are spatially separated.[81,82] While some PBPs modify the peptide stem structure by hydrolysis, and others are important during cell division, the inhibition of the assembling PBPs in particular by the β-lactams is lethal (by mechanisms still poorly understood) to the bacterium.

The particular event in PBP catalysis that is undone by the β-lactams is the transpeptidation reaction. As shown in **Figure 2**, the Lipid II structure consists of a disaccharide, one of which has a bifurcated peptide stem. The terminus of one of the bifurcations is the dipeptide -D-Ala-D-Ala. In the first half-reaction of transpeptidation, the -D-Ala-D-Ala dipeptide is cleaved. The end -D-Ala is released, and the energy content of the amide bond, which had joined the -D-Ala-D-Ala is preserved by acyl-transfer to an active site serine of the PBP transglycosylase domain. The resulting serine acyl-enzyme is intercepted by the amine terminus of the *other* bifurcation of the peptide stem, presented by the *adjacent* peptidoglycan strand. Owing to this reaction the two strands are conjoined. This catalysis is stopped by encounter of the PBP with a β-lactam. The β-lactam is recognized by the PBP as a -D-Ala-D-Ala surrogate, and the acylation reaction of the serine by the β-lactam proceeds robustly. The resulting PBP acyl-enzyme, however, is nearly completely incapable of transfer. It is stable. Although for many PBPs acyl transfer to water, and perhaps to other nucleophiles, can occur eventually, the time scale for this transfer is much, much longer than the time scale that defines bacterial viability. The mechanism of action of the β-lactams is therefore the essentially irreversible inactivation of the transpeptidase PBP domain, resulting in the inability to the bacterium to either grow or to maintain its peptidoglycan cytoskeleton.

In order to protect its PBPs from β-lactam irreversible inactivation, one can conceptualize several strategies that a bacterium might use. The PBPs are membrane-bound enzymes (they are located in the outer leaflet of the cell membranes of both the Gram-positive and Gram-negative bacteria). For the Gram-positive bacterium,
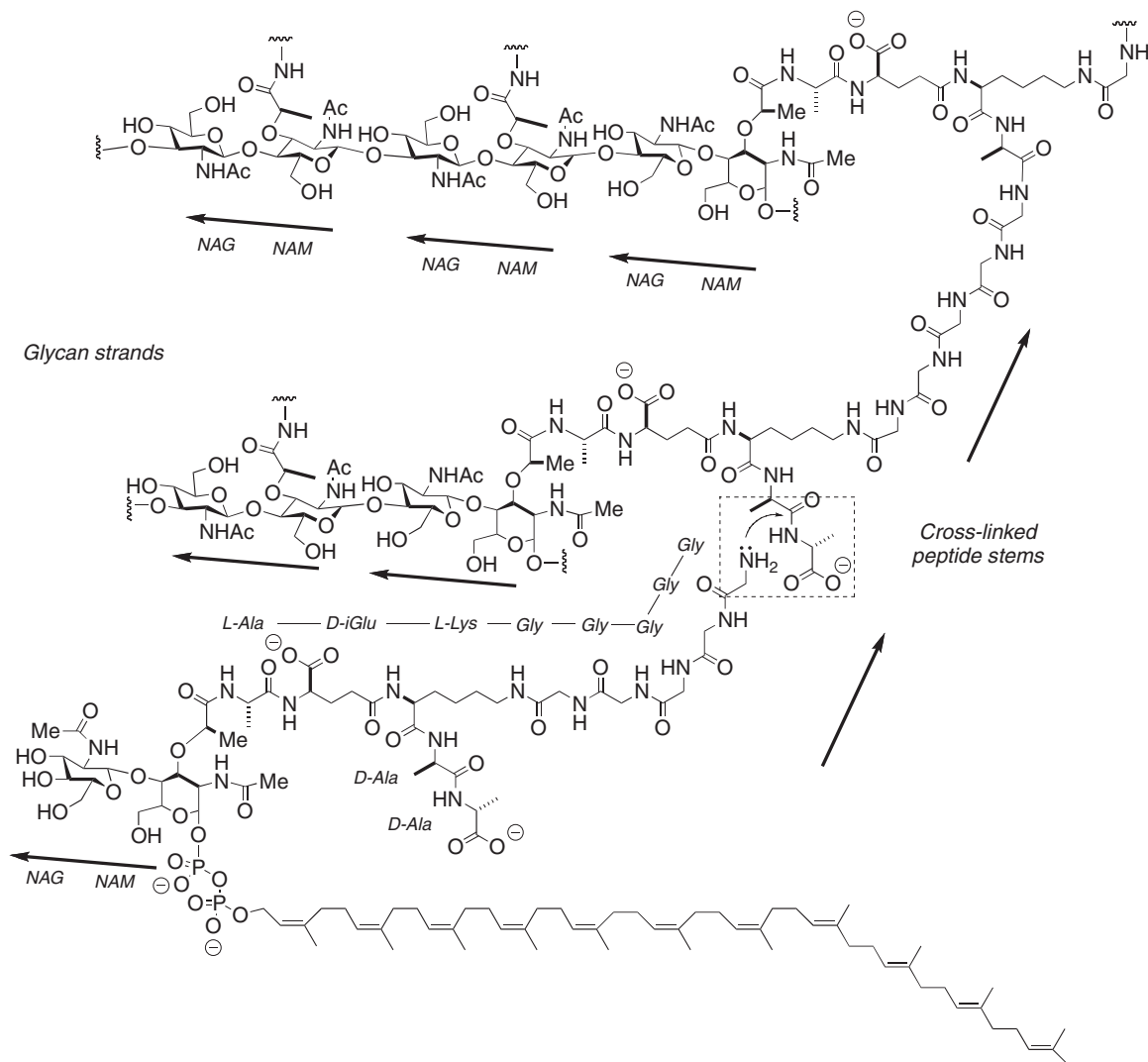
Figure 2 The PBP transpeptidation reaction, exemplified with the peptidoglycan of the Gram-positive bacterium *Staphylococcus aureus*.

the β-lactam must traverse the exoskeleton to reach the PBP that is its target. As there is no simple way for the Gram-positive bacterium to control small molecule diffusion, an effective resistance strategy would be the acquisition of a PBP that more effectively distinguishes in favor of recognition of the -D-Ala-D-Ala stem and against the β-lactam. Alternatively, the Gram-positive bacterium may devise a transpeptidase cross-linking strategy that does not depend on a -D-Ala-D-Ala stem structure. For the Gram-negative bacterium, the β-lactam must traverse the porins of the outer membrane to enter the periplasmic space. Effective strategies here are the selection of porins less suited for β-lactam entry, the acquisition of efflux transporters, and expression of a sentinel enzyme in the periplasm to intercept, and chemically destroy, the β-lactam before it can engage the PBP. It is not simply that all of these events may happen. Rather, all of these events *are* happening, and with such breathtaking speed as to call into question our ability to manipulate β-lactam structure toward new generations of β-lactam antibacterials, before pathogenic bacteria have perfected these resistance mechanisms against the breadth of β-lactam structure.[83–85]

All aspects relating to the bacterial cell wall structure have been reviewed recently. Among the topics covered are the cytoplasmic steps of peptidoglycan synthesis,[86] Lipid II as biosynthetic intermediate[87,88] and antibacterial target,[89] PBP structure and catalysis,[90] peptidoglycan hydrolytic remodeling,[91] structural

variation of the peptide stems,[92] the role of PBPs in rod-shaped bacterial division,[93] the role of PBPs in determining bacterial shape,[94] the role of PBPs in β-lactam resistance,[95] and alterations in peptidoglycan structure in resistant Gram-positive bacteria.[96]

8.13.2.2   Enzymatic Basis of β-Lactam Resistance in Gram-Positive Bacteria

The identification of a resistance mechanism for the β-lactams was virtually coincident with their discovery.[97] The molecular basis of this mechanism was a hydrolytic enzyme, initially called a penicillinase but now universally referred to as the β-lactamase family (reflecting their much greater breadth than penicillins for substrate recognition) that is present in both Gram-positive and Gram-negative bacteria. As a consequence of the very high catalytic efficiency of many of these enzymes, the therapeutic value of the simple penicillins was severely compromised hardly more than a decade after the widespread use of the penicillins. Indeed, one of the best studied of the β-lactamases is the Gram-negative derived TEM enzyme (named after the patient from which the β-lactam-resistant bacterium was isolated), which exhibits catalytic perfection wherein the rate-limiting step in its catalysis of simple penicillin hydrolysis is substrate diffusion.[98] It is now well understood that the β-lactamase evolutionary lineage is ancient, and that the β-lactamases evolved from PBP ancestors.[99] Accordingly, the two have mechanistic similarity in that both proceed via an acyl-enzyme intermediate, and mechanistic divergence by the acquired ability of the β-lactamase to accomplish the efficient hydrolytic deacylation of this intermediate. (A second and equally important sub-class of the β-lactamases that uses metal catalysis, without an acyl-enzyme intermediate, is discussed in the section on β-lactam resistance in Gram-negative bacteria.)

The therapeutic advantage of the β-lactams against *S. aureus* was restored – momentarily – by the chemical synthesis of new generations of the penicillins, represented by methicillin and oxacillin (**Figure 1**). These penicillins were particularly effective against Gram-positive Staphylococci resistant to first-generation penicillins due to penicillinase expression. The advantage did not last long.[69,100,101] Within a short period of time, *S. aureus* resistant to methicillin was encountered (MRSA, methicillin-resistant *S. aureus*), and over the past five decades, MRSA has grown from observation, to localized nosocomial epidemic, and now to a very serious health concern, as this pathogen is also increasingly encountered in the community.[43,63,102] An excellent historical summary of the emergence of Gram-positive and Gram-negative β-lactam resistance (in Gram-negatives, β-lactamase adaptation remains the major resistant determinant) is given by Hawkey.[9]

Transformation of *S. aureus* from a relatively β-lactam-resistant bacterium by virtue of β-lactamase expression, to the highly β-lactam-resistant MRSA occurred primarily as a result of its acquisition of a new enzyme.[103] This new enzyme was not a β-lactamase, but an *accessory* PBP enzyme. The mechanistic distinction of this new PBP (in the literature it is termed PBP 2′, or more commonly PBP 2a) is its ability to accomplish the transpeptidation reaction of cell wall biosynthesis (it does not have a transglycosylase domain to carry out this preceding reaction), due to its superior ability to distinguish the D-Ala-D-Ala stem peptide terminus from a β-lactam antibacterial.[104,105] PBP 2a is susceptible to acylation by β-lactams, but at concentrations and rates that are not meaningful. While the genetic origin of the PBP 2a is not known for certain, the evidence at hand is consistent with its acquisition from a related Staphylococcal animal pathogen, such as *Staphylococcus sciuri*.[106–108] The mechanism for the genetic transfer involves the staphylococcal cassette chromosome *mec* (SCCmec). This gene contains recombinase genes allowing for its chromosomal incorporation, the resistance gene *mecA* encoding the PBP 2a, and the regulatory genes *mecI* (encoding a protein repressor of the *mecA* gene) and *mecR1*. The mecR1 protein is transmembrane, and presents an extracellular β-lactam binding domain, that upon β-lactam acylation transduces a signal to the mecR1 cytosolic domain, resulting in derepression of the *mecA* gene. The mechanism for the derepression is unknown, but may involve proteolysis of the mecI repressor. Hence, exposure of the MRSA bacterium to β-lactam antibacterials results in the simultaneous loss of the transpeptidase activity of the endogenous PBP enzymes (by acylation of the active-site serine of the transpeptidase domain) *and* induction of PBP 2a expression through mecR1 signaling. Successful cell wall synthesis, now in the presence of the β-lactam antibacterial, occurs by the cooperative catalysis of the transglycosylase domains of the endogenous PBPs, and the transpeptidase domain of PBP 2a.[105,108,109] There is a fitness cost for this cooperative catalysis in some SCCmec variants[110] but not others.[111]

From the vantage of the design of improved β-lactam antibacterials, *the* critical question is the structural difference between PBP 2a and the endogenous PBPs that renders PBP2a less susceptible to β-lactam acylation.[112] The high resolution (1.8 Å) structure of PBP 2a determined by Lim and Strynadka[113] has enabled speculation, but not a definitive answer. Zapun *et al.*[95] present a concise summary of the challenge presented by the PBP 2a structure from which some of this discussion is summarized. Several observations concerning the PBPs provide an important context for this discussion. Cell wall synthesis (during growth) and degradation (during the septation of cell division) must necessarily be a highly regulated process involving multienzyme complexes (or hyperstructures).[114–116] The biosynthetic PBPs are recruited *in vivo* to the division site by specific structural determinants of the cell wall structure,[117] and they (including PBP 2a) undergo a conformational change *in vitro* in response to the presence of substrate.[118–120] Moreover, this conformational change correlates to enhanced acylation activity of cephalosporin structures with improved MRSA activity.[121–123] With reference to these conformations, the existence of a less β-lactam reactive 'closed' state and a more β-lactam reactive 'open' state is supported by PBP crystallography[124,125] and the pH dependence of PBP 2a acylation (PBP 2a is fivefold more reactive to acylation by an assay β-lactam at pH 5.5 compared to pH 7.0).[126] An identical pH correlation is seen for the β-lactam susceptibility of the MRSA bacterium.[127] Examination of the native PBP 2a crystal structure, compared to the β-lactam acylated PBP 2a structures, is suggestive of what the closed and opened structural states may resemble. The structural differences are subtle, and focus on the conformational presentation of the nucleophilic serine to the scissile carbonyl of the β-lactam. In the native PBP 2a structure, this serine is poorly positioned, but in the β-lactam acyl-enzyme complex the serine has obviously engaged and opened the β-lactam, corresponding to a rearranged active site achieved by a rotational motion of the structural domains comprising the PBP active site. This motion is inferred to position the serine for catalysis, and also to open the active site for substrate binding.[125] As noted by Zapun *et al.*,[95] the most desirable structure from the viewpoint of β-lactam structure design is neither the native nor acylated PBP 2a structures, but the structure of the pre-acylation β-lactam-PBP 2a complex. Such a structure would more clearly identify the locations permissive for β-lactam substitution. Alternatively, if indeed the conformation of the PBP 2a active site is under allosteric control, the dual use of an allosteric effector and the β-lactam might render PBP 2a highly β-lactam susceptible. The realization of this possibility would require that the allosteric effector does not resemble the peptidoglycan, as synthetic peptidoglycan segments are both extremely difficult to synthesize and do not have acceptable drug likeness. Finally, the sense of this discussion implies that the PBP 2a structure is immutable, without capability for mutational adjustment. This assumption is demonstrably unreasonable.[69,128]

While the transformation of *S. aureus* to a highly β-lactam-resistant state coincides with acquisition and genetic incorporation of SCCmec resulting in PBP 2a expression, realization of high-level β-lactam resistance is now recognized to involve other accommodations in cell wall biosynthesis. The implicit ambiguity of the preceding sentence emphasizes how little is known as to the molecular mechanism of these accommodations. As many as 30 auxiliary *fem* (for factor *e*ssential for *m*ethicillin resistance) genes are identified.[129,130] Among these, attention has focussed on several of the *fem* genes identified as coding for the nonribosomal aminoacyl-tRNA transferases[96] involved in the biosynthesis of the peptide stem of the Lipid II precursor of peptidoglycan biosynthesis, and later used as the acyl-enzyme acceptor in PBP transpeptidase-catalyzed cross-linking.[131–134] For cross-linking, *S. aureus* uses a pentaglycyl terminus that is synthesized by *fem*-encoded aminoacyl transferases (**Figure 2** shows the use of the pentaglycine stem acceptor in transpeptidation). Other Gram-positive bacteria use different acceptor structures: *Enterococcus faecalis* uses an L-Ala-L-Ala terminus and *Enterococcus faecium* a single D-iAsx residue. Similarly, the acceptor structure among Gram-negative bacteria also have characteristic variation.[92] In a remarkable experiment, Arbeloa *et al.*[135] demonstrated that incorporation of the *fem* genes, responsible for the synthesis of the pentaglycine stem of the *S. aureus* peptidoglycan, into *E. faecalis* resulted in the synthesis of a 'mosaic' peptidoglycan using these new stems by the endogenous *E. faecalis* PBPs. The complementary experiment, expression of PBP 2a in *E. faecalis*, provided to the new host full resistance to β-lactams. These experiments dramatically emphasize the adaptive ability of bacteria in cell wall synthesis. However, while genetic experiments indicate that the pentaglycine acceptor stem is a critical aspect in *S. aureus* β-lactam resistance, we are presently at a loss to explain why.[95]

A similar mechanism also involving PBP target modification is used by another Gram-positive pathogen, *Streptococcus pneumoniae*, to attain β-lactam resistance. β-Lactam antibiotic challenge, like many other

antibiotics, induces the SOS response in those bacteria that have this response.[136,137] As a result, transcription becomes mutation-prone. Under β-lactam challenge, three of the six PBP enzymes of *Streptococcus pneumoniae* undergo extensive mutation, resulting in β-lactam resistance.[138,139] The mutations that transform two of these essential PBPs of *S. pneumoniae*, PBP 1b and PBP 2x, from essential enzymes to enzymes that preserve their essentiality while acquiring β-lactam resistance, have been evaluated by *in vitro* β-lactam challenge and the resulting mutants compared to those seen clinically.[140,141]

The *in vitro* evolution of resistance under β-lactam challenge has been examined with a third Gram-positive pathogen, *Enterococcus faecium*. Both Gram-positive and Gram-negative bacteria possess the ability to covalently attach proteins to their cell walls, using a transpeptidation reaction that is conceptually identical to the one used in peptidoglycan biosynthesis.[142] However, the donor peptide in these reactions comprises of L-amino acids, and the amine at the N-terminus of the acceptor peptide may be either an L- or D-amino acid. The enzymes that catalyze this transpeptidation are mechanistically distinct from PBPs. As just one point of difference, the catalytic amino acid at the active site that they use is cysteine, and not serine. Gram-positive bacteria possess an L,D-transpeptidase (abbreviated Ltd$_{fm}$) having the ability to catalyze L,D-cross-linking of its peptidoglycan strands, wherein the L-Lys-D-Ala bond of the Gram-positive peptide stem (and not the D-Ala-D-Ala peptide bond that is used by the PBPs) acts as the peptide donor and the D-iAsx (in the case of *E. faecium*) is the amine acceptor. The resulting cross-link in the peptidoglycan is termed a 3–3 cross-link, which is distinct from the 4–3 cross-link formed by the PBPs using the -D-Ala-D-Ala terminus. As 3–3 cross-linking does not require a -D-Ala-D-Ala peptide stem terminus, and the β-lactam antibacterials are D-Ala-D-Ala mimetics, should the bacterium have the ability to use Ltd$_{fm}$ transpeptidases for cell wall biosynthesis, this altered biosynthetic route would be expected to be impervious to the β-lactam (and to the glycopeptides, discussed subsequently, which also take advantage of D-Ala-D-Ala recognition). This expectation was addressed by an experiment by Mainardi *et al.*,[143] and with a resoundingly positive answer. Under conditions of β-lactam (ampicillin) challenge, the extent of 3–3 cross-linking increased from 3% in the starting β-lactam-susceptible *E. faecium* to the exclusive cross-linking mechanism in the highly ampicillin-resistant *E. faecium* obtained from the successive ampicillin challenge. Additional enzymes of the Gram-positive Ltd transpeptidase family now are characterized.[144] Unexpectedly, the Ltd$_{fm}$ transpeptidase was found to be inactivated by carbapenem β-lactams, with initial experimental data fully consistent with an acylation reaction (with concomitant β-lactam ring opening) of the active site cysteine, to form a stable acyl-enzyme.[145] Although these β-lactam challenge experiments were done *in vitro*, it now must be understood as probable that a resistance mechanism encountered *in vitro* will subsequently be seen in the clinic. Assuming that this occurs, Arthur and co-workers note that combination therapy with a penicillin/cephalosporin and carbapenem should be effective.[145]

8.13.2.3   Enzymatic Basis of β-Lactam Resistance in Gram-Negative Bacteria

The mechanisms used for β-lactam resistance in the Gram-negative bacteria are fundamentally different. Rather than target adaptation, Gram-negative bacteria combine porin selection, active efflux, and optimization of the β-lactamase detoxifying enzyme (localized in their periplasm) for the hydrolytic destruction of the β-lactam before it can engage and inactivate their PBP enzymes. Optimization of porin selection against antibacterial ingress, and the acquisition of active transporters for effective antibacterial efflux, are important factors but are incremental. The acquisition and optimization of the catalytic ability of the β-lactamase is decisive. As noted previously, the β-lactamase enzyme is an evolutionarily ancient adaptation of bacteria, from a no less ancient PBP, for the purpose of defense. At the dawn of the antibiotic era, a relatively small number of β-lactamase antecedents are recognized. However, as a result of the intensive use of β-lactams as antibacterials, the β-lactamase family has diversified into a remarkably large enzyme family, the newest members of which possess the ability to efficiently hydrolyze the newest generations of β-lactam structure. The rapidity of this diversification arguably exceeds our ability to comprehend the hydrolytic power of new β-lactamase variants, and to devise structures capable of evading their catalysis. In evidence of the diversity of their structure, the β-lactamase family encompasses a seemingly bewildering array of acronyms and abbreviations. A summary of the definitions used to describe the β-lactamase family, including those based on biochemical properties (primarily protein sequence), and a newer system that distinguishes the β-lactamases in terms of functional recognition of substrates and inhibitors, is essential for further discussion.

The recent literature offers excellent summaries of the challenges on efforts to systematically characterize the β-lactamases.[146–150] The classical classification system for the β-lactamases was devised by Ambler on the basis of biochemical criteria (Classes A, B, C, and D), and is now complemented by the Bush–Jacoby–Medeiros system that incorporates functional and mechanistic criteria (Groups 1, 2, 3, and 4). In the Ambler classification, the β-lactamases are divided among four classes: Classes A, C, and D enzymes operate via a serine-active site nucleophile with catalysis involving an acyl-enzyme intermediate, and Class D enzymes operate via a metal-containing active site and with catalysis not involving an acyl-enzyme intermediate. Historically, Class A encompassed serine-dependent enzymes that preferentially hydrolyzed penicillins, but this limitation on the Class A substrate spectrum assuredly no longer exists today. The most common Class A prefixes are TEM and SHV, and less frequently IMI, KPC, NMC, and SME. Historically, the Class D enzymes were those serine β-lactamases capable of hydrolysis of the oxacillin and cloxacillin penicillins (both poor substrates of the early Class A 'penicillinases'). Accordingly, most Class D enzymes are named using the 'OXA' prefix. The Class D substrate spectrum now encompasses the carbapenems, and as a result, these OXA enzymes confer carbapenem resistance in important Gram-negative pathogens (such as *Acinetobacter* spp.).[151] It is now known that the mechanism used by the class D enzymes for serine activation for β-lactam acylation has striking analogy to the acylation-detection mechanism used by the sensor domain of the mecR protein in MRSA to initiate derepression of the *mec* gene, thus allowing PBP 2a expression.[101,152] A very similar sensor mechanism is used also by Gram-positive bacteria to control β-lactamase expression.[153–156] Nonetheless, there is strong similarity in the organization of the active site amino acids between Classes A and D. Many of the Class A, B, and D enzymes are now disseminated via transposon, integron, or plasmid genetic exchange. Historically, the Class C enzymes were chromosomal-encoded β-lactamases having the ability to hydrolyze cephalosporins (again, historically poor to fair substrates of the early Class A penicillinases). Moreover, these enzymes were generally refractory toward inhibition by clavulanate, a β-lactamase inactivator (*vide infra*). The primary nomenclature prefix used for the Class C enzymes is AmpC. The newer generation Class C enzymes are capable of hydrolyzing new generation cephalosporins and carbapenems, and are now also genetically mobile (plasmid-borne). As a consequence there is concern that further movement of these enzymes into new Gram-negative niches (such as *E. coli*) to create new β-lactam-resistant pathogens is probable.[157,158] In the Bush–Jacoby–Medeiros classification, Group 1 coincides with the Ambler Class C. Group 2 are serine-dependent β-lactamases susceptible to clavulanate (encompassing both Class A and Class D enzymes). Group 3 are the metallo-β-lactamases (often abbreviated as MBLs), and class 4 is reserved for miscellaneous (not yet characterized) β-lactamases.

Furthermore there are abbreviations describing the β-lactamase family that also require definition. The basis for these additional abbreviations is strongly correlated to the temporal development by medicinal chemists of β-lactam structure (see **Figure 1**). The first β-lactams were the penicillins. The penicillins were replaced by the cephalosporins, a β-lactam class with improved safety, oral availability, and resistance to early β-lactamases. With increasing β-lactam resistance, the early generation cephalosporins were succeeded by the oxyimino class (as are shown in **Figure 1**) having improved β-lactamase resistance, by the cephamycin class exemplified by cefoxitin (also with improved β-lactamase resistance), and by the carbapenems. The carbapenems today represent the most potent β-lactam antibiotics, but increasingly susceptible to β-lactamase hydrolysis. Finally, the introduction in the mid-1970s of β-lactamase inactivators (exemplified by clavulanate, sulbactam, and tazobactam) extended the utility of early generation β-lactams by the ability of these inhibitors to inactivate many β-lactamases.[159] The use of β-lactam combination therapy (co-administration of a β-lactam and a β-lactamase inactivator) is a major method for β-lactam anti-infective therapy.[160] With the introduction of all of these β-lactams, and the clinical development of resistance, the catalytic power of the β-lactamases has stratified. Those β-lactamases that were once inhibited by cavulanate, and are no longer, are referred to as inhibitor-resistant TEM (IRT) enzymes. Those β-lactamases that were poorly capable of oxyimino cephalosporin hydrolysis, but have now acquired this ability, are referred to as extended-spectrum β-lactamases (ESBLs).[161,162] The definition promulgated for the ESBL β-lactamases is a β-lactamase 'generally acquired rather than inherent to a species, that is able to confer resistance to the oxyimino cephalosporins (but not carbapenems).'[150] At this time, the major ESBL sub-categories[163] are the TEM-ESBLs and SHV-ESBLs (from the Ambler serine Class A), the CTX-M ESBLs (also serine Class A),[164] the AmpC-ESBLs (from the Ambler serine class C), and the OXA-ESBLs (from the Ambler serine Class D). The final stratification (at this time) is

the carbapenemases.[149] By virtue of their ability to hydrolyze carbapenems, these enzymes are distinguished from the ESBLs even though many carbapenemases encompass the oxyiminocephalosporins as substrates. The carbapenemases include Class A (IMI, KPC, GES, NMC, and SME),[165] Class D (OXA),[166] and Class B MBL (VIM, IMP, GIM, and SPM) enzymes. As fully discussed by Livermore,[150] these classifications are consensus and pragmatic, and close attention to the nuances of these terms is necessary as the β-lactamases continue to evolve. It is evident from the complexity of this presentation that the β-lactam class of antibacterials is truly in peril from β-lactamase resistance mechanisms. The magnitude of this peril cannot be overemphasized.[85]

The logical criterion to separate further discussion on the enzymology of Gram-negative resistance to the β-lactams is resistance due to the presence of a Class B metallo-β-lactamase (MBL) and resistance due to the presence of a serine β-lactamase (again Class A, C, or D). The MBLs are an important aspect of this peril.[167–169] They are a major sub-class of the carbapenemases, and on a worldwide basis they account for a substantial portion of β-lactam-resistant *P. aeruginosa* and *A. baumannii* (and increasingly also *Enterobacteriaceae*) infections.[149] The most common Class B β-lactamase prefixes are VIM and IMP.[170,171] Aggressive study of the MBLs (exemplified by the following recent publications) has focused on the protein structure of the active site,[172–174] metal-binding,[175–183] catalytic mechanism,[184–187] and the basis for β-lactam substrate recognition.[188,189] The possible molecular events for MBL hydrolysis of β-lactams are reviewed by Crowder *et al.*[190] An unexpected clinical benefit of recombinant MBLs is suggested by a recent study, where the use of an orally administered MBL preserved colonization in β-lactam-treated mice.[191] Such strategies would have value in preventing the proliferation of β-lactam-resistant bacteria. A major research objective is the discovery of effective MBL inhibitors, and while there is recent progress[192–197] the prospect of a clinically useful MBL inhibitor is distant.

The pivotal concept for the understanding of the enzymatic basis for serine β-lactamase-derived β-lactam resistance is the acyl-enzyme intermediate. The fate of three different acyl-enzymes is captured in schematic summary by **Figure 3**. The top structure is the PBP acyl-enzyme derived from acylation of the serine by the penultimate D-Ala residue of the -D-Ala-D-Ala terminus of the peptide stem of the peptidoglycan. As discussed previously, this acyl-enzyme is primarily used biosynthetically for transpeptidation (transfer to the amine terminus of an adjacent peptide stem). The PBP acyl-enzyme undergoes hydrolysis only within a PBP sub-class that controls the extent of cell-wall cross-linking. The lower structure of this **Figure 4** is the acyl-enzyme derived from a β-lactam antibiotic. This acyl-enzyme is destined to undergo hydrolysis. The β-amino acid product of β-lactamase hydrolysis is biologically inactive: the β-lactam has been destroyed, and is no longer a threat (by its ability to irreversibly acylate PBP serines) to the bacterium. The third acyl-enzyme shown is derived from a β-lactam that is a β-lactamase inactivator. The β-lactamase inactivators shown in **Figure 1** are clavulanate, sulbactam, and tazobactam. Owing to the functional group complexity of these inactivators, the five-membered heterocyclic ring resulting from β-lactam opening has the ability for further reaction. In the case of clavulanate, the five-membered 5-alkylideneoxazolidine ring opens, at the instigation of the nonbonding electron pair on the nitrogen, allowing departure of an enolate anion as a leaving group. An analogous reaction (here, with a sulfinate leaving group) occurs for sulbactam and tazobactam. When these five-membered rings are opened, the β-lactamase loses its ability to efficiently complete deacylation, the second half-reaction of hydrolysis. Its serine is now irreversibly acylated.[198] Clearly, the efficacy of these compounds depends on the facility of these ring fragmentations (indicated by the three branching arrows), and also on the ability of the active site to intercept the new acyl-enzymes with proximal active site nucleophiles[199] and to otherwise stabilize the new acyl-enzyme.[200–207] As noted previously, active site mutations that slow the fragmentation, or are more permissive for hydrolysis of the acyl-enzymes, correspond to the IRT class of inhibitor-resistant β-lactamases.[208] New classes of β-lactamase inactivators, many of which also depend on acyl-enzyme fragmentation or rearrangement pathways, remain a medicinal chemistry focus as exemplified by several recent publications.[209–212]

The importance to the continuing study of the molecular mechanism of the β-lactamases, for the purpose of assessing new β-lactam structure and for the conceptual development of ideas to preserve the clinical efficacy of the β-lactams, is certain. Two recent observations underscore this assertion. The basis for β-lactam resistance by *Mycobacterium tuberculosis* is the possession of a chromosomally encoded Class A ESBL β-lactamase, BlaC. This β-lactamase is, however, inactivated by clavulanate, suggesting the utility of combination β-lactam–clavulanate therapy for tuberculosis.[213,214] A second example involves the site-directed mutagenesis
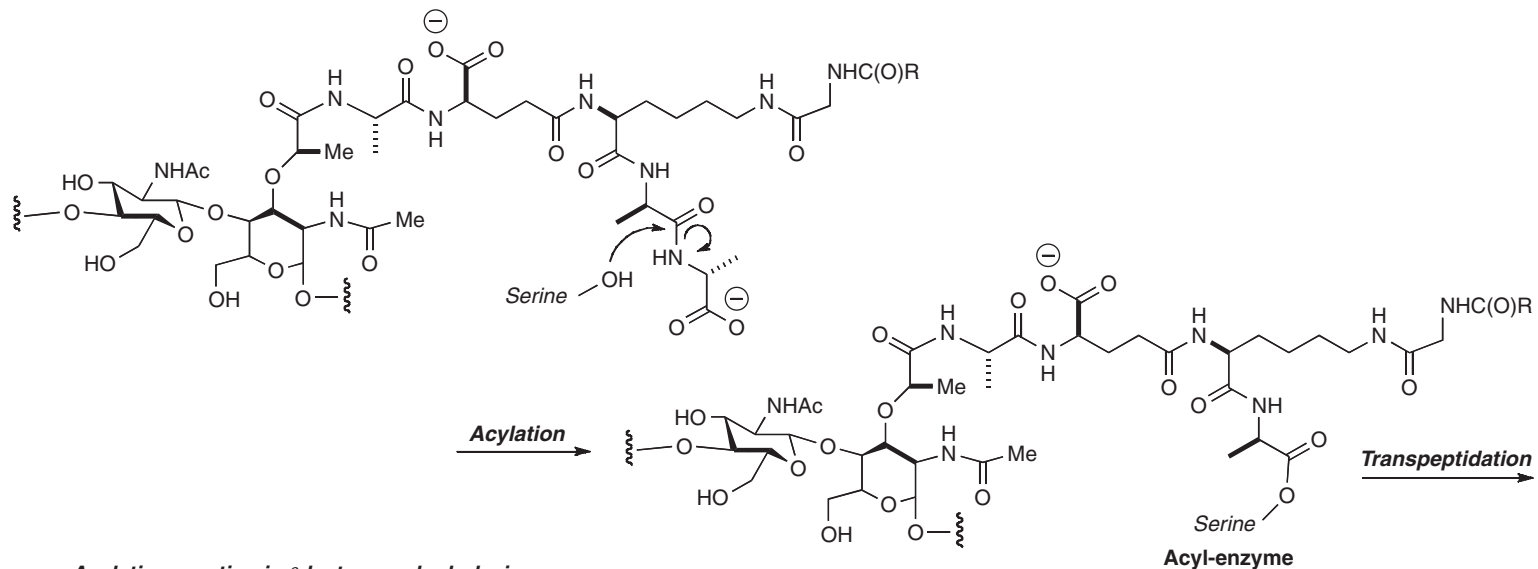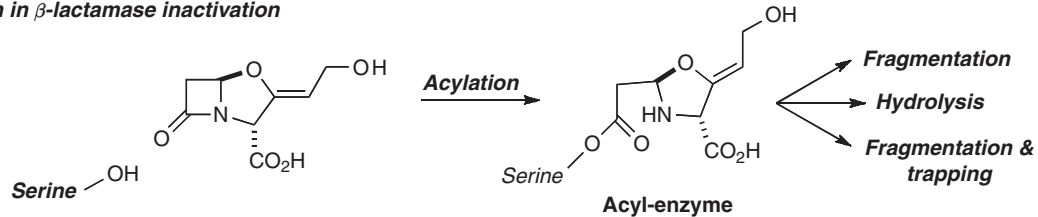
Figure 3 The different fates of three acyl enzymes: the -D-Ala-D-Ala derived PBP D-Ala acyl-enzyme (transpeptidation); the penicillin-derived β-lactamase acyl-enzyme (hydrolysis); the β-lactamase inactivator-derived β-lactamase acyl-enzyme (hydrolysis and fragmentation to stable acyl-enzymes).
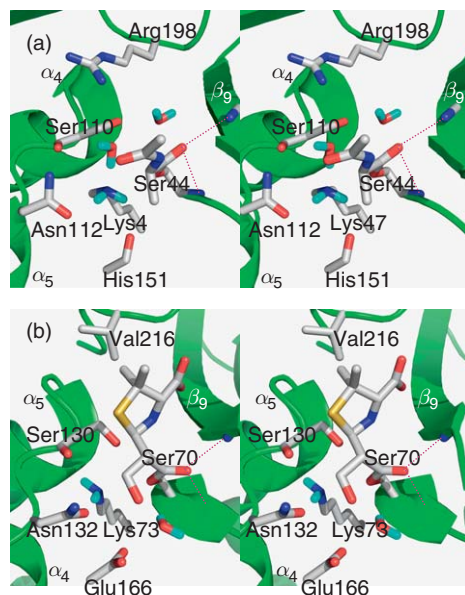
Figure 4 Structural comparison (in stereo representation) of the acyl-enzymes of a PBP (top structure, obtained from computational modeling) and a Class A β-lactamase (bottom structure, from the X-ray structure PDB Code 1TEM). The PBP shown is the *Escherichia coli* PBP 5, a hydrolytic PBP that controls peptidoglycan cross-linking by removal of the terminal D-Ala from peptide stems. The general base used in the acylation reaction is Lys47. Its D-Alanyl acyl-enzyme is derived from acyl transfer to Ser44 of the active site. The carbonyl oxygen of the acyl-enzyme ester is identified by the two dashed lines, indicating the location of the hydrogen bonds to this carbonyl oxygen from the amides of the oxyanion hole. The β-lactamase is the TEM Class A β-lactamase, acylated at Ser70 with a β-lactam inactivator. The mechanistic basis for the inactivation is a 6α-hydroxymethyl substituent on the β-lactam ring of the inactivator. The alcohol of this side chain displaces the hydrolytic water molecule from the active site, thus stabilizing the acyl enzyme. Lys73 is the cognate general base of Lys47 of the PBP, abetted by the carboxylate of Glu166. The dashed lines again identify the hydrogen bonds from the amide NHs of the oxyanion hole to the carbonyl of the acyl enzyme.

of a Class C β-lactamase. This study confirms that dramatic alteration in the substrate profile of an enzyme can result from mutation of otherwise highly conserved active site amino acids.[215,216] Comparison of the k_{cat}, K_M, and k_{cat}/K_M values for β-lactam substrates of the mutant β-lactamases against a predictor of antibacterial efficacy, showed that the strongest correlation was with k_{cat}.[216] This correlation supports the hypothesis advanced by Frère[217,218] that the efficacy of a resistance enzyme, under conditions where the concentration of the antibacterial that it sees *in vivo* exceeds the K_M value for that antibacterial, should more closely correlate with k_{cat} than k_{cat}/K_M.

Given the fundamental importance of the PBPs and the serine β-lactamases to the understanding of β-lactam reactivity, the existence of quality structural data for both enzymes, and their evolutionary related-ness, one might expect that a detailed understanding of the catalytic mechanisms for acylation (for both enzymes) and for deacylation (whether by transpeptidation or hydrolysis), would now be in hand. This understanding has proven elusive. The points of similarity for the active sites of both enzymes encompass several signature motifs. The most important is a Ser–Xaa–Xaa–Lys tetrad (where S is the serine that undergoes acylation, and the side chain amine of the lysine of the tetrad is in contact with the nucleophilic hydroxyl of the serine). A second conserved lysine-containing motif places into the active site a second lysine side chain, the amine of which is also capable of interacting with the serine and lysine of the first motif. The likelihood that the lysine of the SXXK tetrad of the PBPs is the general base in serine activation for acylation is strongly implicated by recent computational study.[219] Comparison of the PBP acyl-enzyme generated by this computational study with crystallographic PBP acyl-enzymes shows very strong similarity.[220,221] When this acyl-enzyme is com-pared to that of a Class A acyl-enzyme (obtained with a penicillin-based inhibitor),[222] the key point of difference between the PBP and the serine β-lactamase active site is identified: the presence in the latter of a glutamate, directly interacting with the side chains of the serine and lysine in the SXXK tetrad. This glutamate

enters the β-lactamase active site by helix addition to the protein. While the role of the serine as the acyl acceptor in both enzymes is beyond dispute, the spatial interplay of the serine with two lysines *and* a glutamate in the Class A β-lactamases has resulted in several mechanistic conjectures for the acylation mechanism (especially whether it is the lysine of the SXXK tetrad or the new glutamate, acting through a conserved water, that is the primary general base catalyst for serine acylation).[223–226] A reasonable possibility is that the serine β-lactamase active site uses the interplay of these three residues to modulate the pK_a of the SXXK lysine such that this lysine and the glutamate have matched basicities. Since the β-lactamase must accommodate a diversity of β-lactam structure, and the angular presentation of the bicyclic β-lactam structure varies considerably among cephalosporins (nearly planar), penicillins (substantially angled), and carbapenems (acutely angled), this basicity accommodation would allow either the glutamate of lysine to effectively function as the general base for acylation, depending on the substrate.[227] Once the acyl-enzyme is achieved, the necessity to accommodate the different angularity of the bicyclic β-lactam is no longer required, and a single catalytic base suffices for the activation of water addition to the acyl-enzyme. There is a broad consensus that the glutamate fulfills the mechanistic role of general base in deacylation.

The sophistication that is implicit in this mechanism is a measure of the magnitude of the problem presented by the resistance mechanisms that exist against the β-lactams, and that continue to evolve. A complementary perspective on the relationship between the β-lactamases and β-lactam resistance is presented in Chapter 8.03. The β-lactam structure imbues nearly unparalleled safety, mechanistic specificity, and efficacy in the clinic and community. The possibility that these qualities will erode from this structure, quite literally before our eyes, is frightening.

8.13.3   Enzymatic Basis for Glycopeptide Resistance

The glycopeptides are bacteriostatic Gram-positive antibiotics,[228–230] discovered over 50 years ago, whose clinical importance has progressively increased over the past two decades with the emergence of methicillin-resistant *S. aureus*. The parent structures of the glycopeptide class (**Figure 5**) are vancomycin (used clinically in the United States) and teicoplanin (used elsewhere in the world).[231–233] As a result of the former use of glycopeptide derivatives in animal husbandry, and the more recent dramatic increase in the use of vancomycin to control human infections (for the treatment of postoperative colitis caused by *Clostridium difficile* and *S. aureus*, and more recently for MRSA), clinically significant resistance has emerged. Vancomycin resistance first appeared in the enterococci (VRE, vancomycin-resistant enterococci particularly *E. faecium* and *E. faecalis*),[234–236] and subsequently – by horizontal gene transfer – from the enterococci to *S. aureus*.[61,100,237,238] Two levels of resistance by *S. aureus* to the glycopeptides are encountered: intermediate level resistance (VISA), observed 10 years ago and continuing as a clinical concern in both the enterococci and cocci Gram-positive bacteria; and full resistance (VRSA), at this time a rare clinical event but nonetheless an event with potential future clinical ramifications.[66,237–241] The molecular mechanism for glycopeptide resistance is target modification.[242,243] This target, both for recognition and for modification leading to resistance development, is a specific sub-structure: the -D-Ala-D-Ala dipeptide terminus of the peptide stem of the nascent peptidoglycan.[244,245] Two types of target modifications to the Gram-positive bacterial cell wall are observed that confer glycopeptide resistance. The first modification is nongenetic and morphological (increased cell wall thickness and increased uncrosslinked stems with -D-Ala-D-Ala termini), while the second modification is genetically acquired, more effective at imparting resistance, and enzyme-dependent.[233,246,247] This second enzymatic method for glycopeptide resistance has proven to be deeply unsettling from our naive perspective on microbiological adaptation, as it establishes a general mechanism whereby (seemingly wholesale) remodeling of the Gram-positive cell wall can be accomplished for the purpose of antibiotic evasion.[96]

Although glycopeptide resistance as a result of morphological adaptation is only indirectly related to enzymatic mechanism (and thus peripheral to the focus of this review) it merits summary, as relevant introduction to the basis for the second, and enzymatic, mechanism of vancomycin resistance. Vancomycin and teicoplanin interfere with cell wall biosynthesis as a result of their ability to form a noncovalent complex with the -D-Ala-D-Ala dipeptide terminus of the peptide stem of peptidoglycan precursors (**Figure 6**).[248] As described previously (with reference to β-lactam resistance), the key event in early stage of peptidoglycan
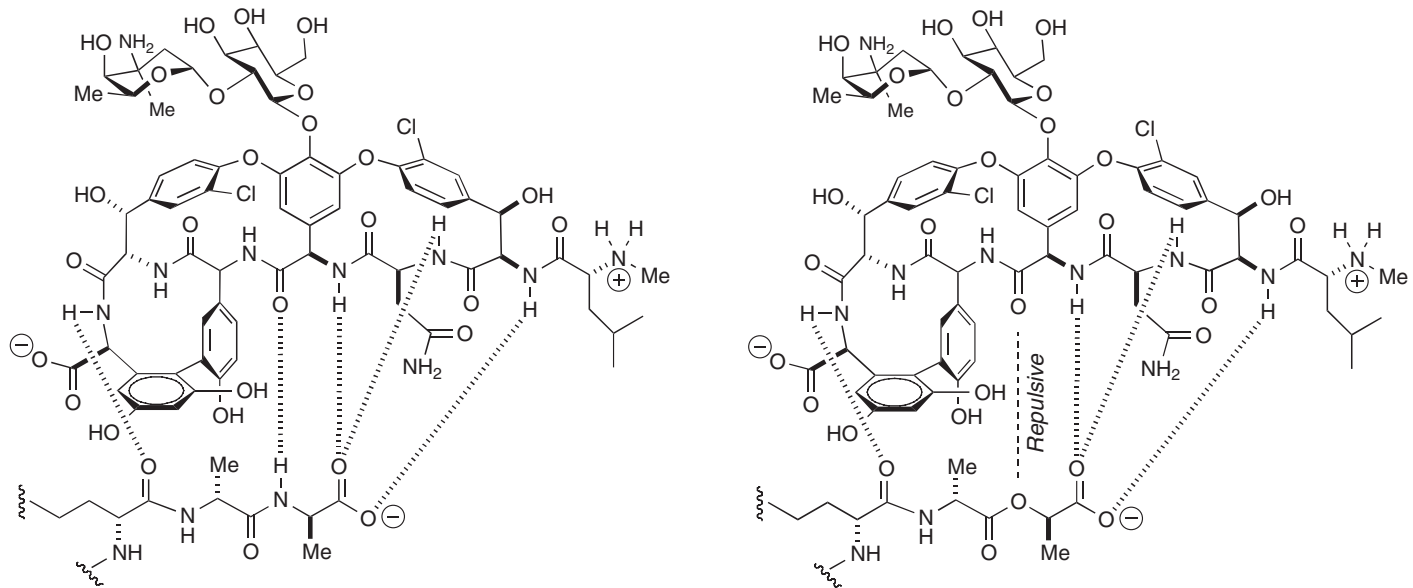
Figure 5 Structure of the vancomycin complex with the D-Ala-D-Ala terminus of the peptidoglycan peptide stem (left) and the less stable (by ~1000-fold) complex with the D-Ala-D-Ala terminus found in glycopeptide-resistant Gram-positive bacteria. The replacement of the amide link by that of an ester results in the loss of a hydrogen bond and repulsive nonbonding interactions between the carbonyl and the ester oxygen.
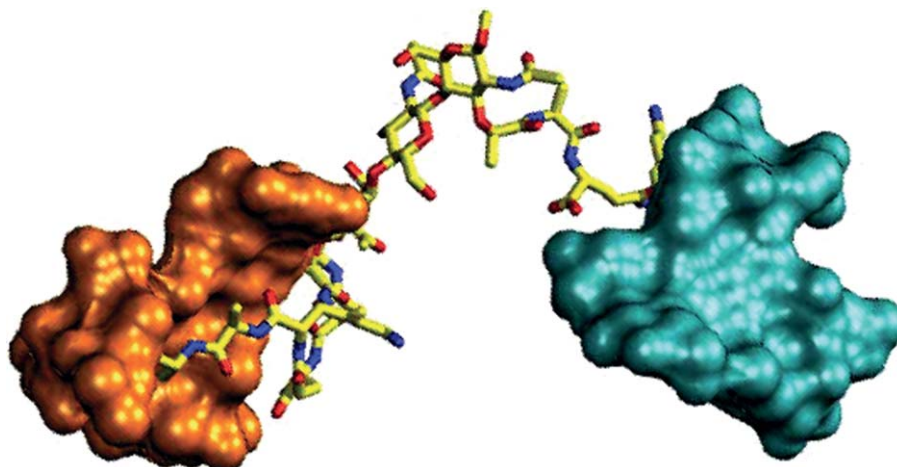
Figure 6 Computational structure of a peptidoglycan segment (shown in capped-sticks representation colored by atom type: C, yellow; N, blue; O, red), containing two peptide stems. Each stem has a D-Ala-D-Ala terminus, and each terminus is complexed by a single vancomycin molecule (each shown as a Connolly solvent-accessible surface, with one vancomycin colored orange and the other colored cyan). The association constant for complexation of the first vancomycin is $1.1 \times 10^6 \, \text{mol}^{-1}\text{l}$ ($\Delta G°$, $-34.5 \, \text{kJ} \, \text{mol}^{-1}$; $\Delta H°$, $-41.2 \, \text{kJ} \, \text{mol}^{-1}$; $\Delta S°$, $-6.7 \, \text{kJ} \, \text{mol}^{-1}$) and the association constant for the second vancomycin is $4.0 \times 10^5 \, \text{mol}^{-1}\text{l}$. Data are taken from M. Rekharsky; D. Hesek; M. Lee; S. O. Meroueh; Y. Inoue; S. Mobashery, *J. Am. Chem. Soc.* **2006**, *128*, 7736.

synthesis is the PBP transglycosylase elongation of the glycan strands, using the -D-Ala-D-Ala containing peptidoglycan precursor, Lipid II. The key events in the final steps of peptidoglycan polymerization is removal, catalyzed by the transpeptidase domain of the PBP, of the D-Ala terminus with concomitant formation of an acyl-enzyme. In the second step of transpeptidase catalysis, the acyl-enzyme is transferred to the terminal amino acid on the peptide stem of an adjacent peptidoglycan strand. Complexation of the -D-Ala-D-Ala strand terminus by the glycopeptide can be seen as presenting a structural (steric) basis for interference with either the transglycosylase reaction, or with the transpeptidase cross-linking. The experimental distinction between the two pathways is not easy. Detailed kinetic studies implies glycopeptide complexation with Lipid II, resulting in inhibition of the transglycosylase reaction, as the dominant inhibition pathway for several important glycopeptide structures.[249,250] This conclusion is consistent with the recognition that there is a small steady-state Lipid II pool during cell wall synthesis that requires extremely efficient recycling.[89] Nonetheless, the generalization that the PBP transglycosylation reaction is the single mechanism of the glycopeptides is not warranted (*vide infra*). The realization that the glycopeptide must engage its targets (Lipid II, PBPs) at the nexus of peptidoglycan biosynthesis explains the Gram-positive specificity of the glycopeptide antibacterial spectrum: the Gram-positive bacterium has a surface exposed cell wall, whereas access to Lipid II and the PBPs in Gram-negative cell wall synthesis is prevented by the outer membrane of the Gram-negative bacterium. Moreover, this realization implies direct value to the use of (appropriately modified) glycopeptides to the purpose of identifying the location of cell wall synthesis by the Gram-positive bacterium. Two studies, using fluorescent-tagged glycopeptides, have undertaken this study, and with dramatic observation. Using the rod-shaped *B. subtilis* bacterium, Tiyanont *et al.*[251] found that sub-MIC concentrations of fluorescent glycopeptides visualized a broad helical for the cell wall across the cylindrical outer surface of this bacterium. This strongly suggests a supramolecular pattern for cell wall biosynthesis, accomplished by enzymatic hyperstructures coordinated with the other enzymes involved in bacterial growth and division.[115,116] A second study compared the cocci of susceptible and VISA-resistant *S. aureus*. In both, the subcellular site of cell wall synthesis was the division septum (**Figure 7**), and with lower diffusion of the glycopeptide to this site in the resistant *S. aureus*.[252] The sub-cellular morphology of this division site is complex,[253] but is fully consistent with this interpretation. This observation is interpreted to implicate the division septum as the point for glycopeptide encounter of its targets. If access of the glycopeptide to this division septum is impeded, then the likelihood of the glycopeptide stalling (say, as a result of encounter with -D-Ala-D-Ala termini left on the mature cell wall) on its path to the
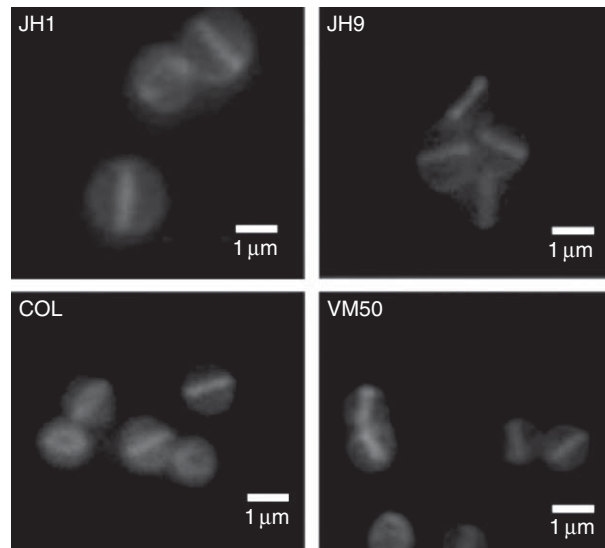
Figure 7 Localization of cell wall synthesis in isogenic pairs of vancomycin-susceptible and vancomycin-resistant *Staphylococcus aureus* strains. Images show labeling of JH1/JH9 (susceptible) and COL/VM50 (resistant) cells with a fluorescent vancomycin derivative, after growth with an excess of D-serine (synthesis of older peptidoglycan with D-Ala-D-ser termini, to which the labeled vancomycin cannot bind) followed by transient incubation with D-Alanine, which results in D-Ala incorporation into new peptidoglycan, to which the vancomycin can bind. In all cases, the main location for cell wall synthesis is at the division septum. Data reproduced from P. M. Pereira; S. R. Filipe; A. Tomasz; M. G. Pinho, *Antimicrob. Agents Chemother.* **2007**, *51* (10), 3627, with permission from the American Society for Microbiology.

septal tip is increased (**Figure 8**). For this reason, a correlation of increased glycopeptide resistance with increased cell wall thickness is anticipated,[254,255] and it is hardly a surprise that there is indisputable independent evidence of this anticipation.[256–261] A final (and as yet not fully understood) observation is the loss of the cell wall biosynthetic PBP enzyme from the *S. aureus* division site as a result of glycopeptide complexation of lipid II,[117] indicating the presence of substrate-competent Lipid II as a requirement for the recruiting of this key enzyme to the biosynthetic hyperstructure.

   These observations broadly define the circumstance permissive for enzymatic resistance to the glycopeptides. As the key substructure for glycopeptide recognition – the -D-Ala-D-Ala of the peptide stem of Lipid II – contains an entity (the terminal D-Ala residue) that is removed in the course of PBP transpeptidase-dependent cross-linking, if this entity can be substituted by an alternative structure inconsistent with glycopeptide binding but functional in transpeptidation, then glycopeptide resistance will result. This event has now happened. Moreover, it has happened by several different molecular events[96] notwithstanding the requirement that for this to happen, significant remodeling of the biosynthetic pathways for cell wall biosynthesis is necessary. The genetic basis for these events derive from the mechanisms for antibiotic immunity devised by the producing organism.[73,262–264] The primary molecular mechanism for these events is the substitution of the D-Ala terminus by D-lactate (thus forming a Lipid II D-Ala-D-Lac depsipeptide structure), wherein the hydrogen-bonding capacity of the ester is significantly reduced relative to the D-Ala-D-Ala amide linkage that it replaces. That the energetic cost of this loss in hydrogen bonding combined with lone pair electrostatic repulsions correlate directly to the observed 10^3-fold loss in glycopeptide affinity for the D-Ala-D-Lac compared to the D-Ala-D-Ala terminus,[265] was established by Boger and colleagues,[266] and confirmed by the synthesis of a glycopeptide analogue that recovers much of this lost affinity.[267] The D-Ala-D-Lac resistance phenotypes *vanA*, *vanB*, and *vanD* correspond to high, variable, and moderate vancomycin/teicoplanin resistance levels in the enterococci.[61] A second structural replacement for the D-Ala terminus is that of D-ser, as seen in the *vanC*, *vanF*, and *vanG* resistance phenotypes.[268] The presence of D-ser in the Lipid II stem results in steric interference with glycopeptide complexation, achieving relatively low levels of glycopeptide resistance. A newer and quite different mechanism[96] for glycopeptide resistance by the enterococci, seen thus far *in vitro* (not yet clinically), is cell wall synthesis using an L,D-dipeptide terminus and L,D-transpeptidase catalysis of cross-linking.[144,269]
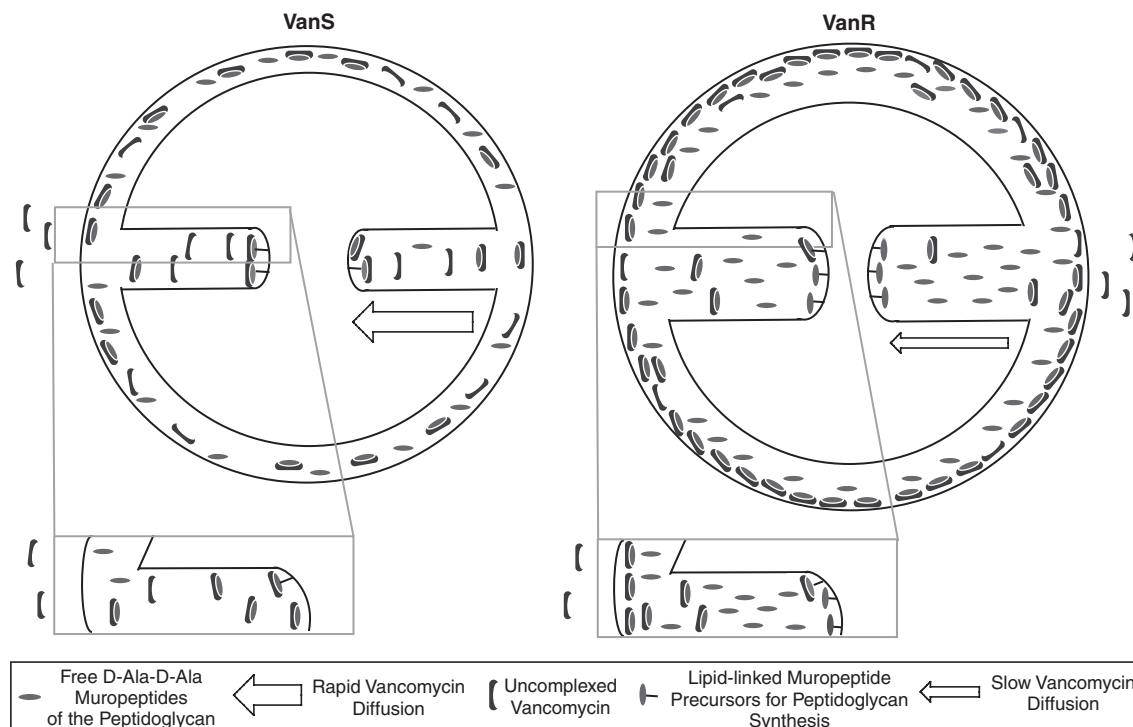
Figure 8   Model for vancomycin resistance in VISA strains. The path of vancomycin to its lethal target (lipid II) should be through the division septum. In resistant cells (VanR), the diffusion rate of vancomycin molecules to the septal tip is decreased, lowering the effective concentration of antibiotic that reaches the lipid-linked peptidoglycan precursor (lipid II) at the site of cell wall synthesis, per unit time, and therefore tilting the balance in favor of continued cell wall synthesis. This model implies that vancomycin efficiency varies during the cell cycle, as the path from the outside of the cell to the lethal targets is shorter when the septum starts to be formed and longer when septum synthesis approaches completion. Scheme reproduced from P. M. Pereira; S. R. Filipe; A. Tomasz; M. G. Pinho, *Antimicrob. Agents Chemother.* **2007**, *51* (10), 3627, with permission from the American Society for Microbiology.

Resistance as a result of *vanA* acquisition is effective, encompasses resistance to both vancomycin and teicoplanin, is the resistance phenotype most frequently encountered in the enterococci, and is the resistance mechanism that has now moved into *S. aureus*. Whereas the adaptation of a thicker Gram-positive cell wall, resulting in intermediate vancomycin resistance (VISA *S. aureus*, MIC = 8 mg l^{-1}), is believed to be primarily nongenetic, in the (thus far, rare) clinical isolates that show high-level vancomycin resistance (VRSA *S. aureus*, MIC $\geq$ 32 mg l^{-1}), do so as a result of *vanA* gene expression. As summarized by Courvalin,[61] the prototype *vanA* resistance element is Tn1546, first isolated from a clinical *Enterococcus faecium* as a plasmid-borne 11-kb transposon. The *vanA* gene may be located either on the chromosome or on a plasmid. On the basis of the preceding discussion on the molecular mechanism of the β-lactams, and the above discussion of the molecular basis for *vanA*-dependent resistance, the replacement of the D-Ala-D-Ala terminus by a D-Ala-D-Lac depsipeptide terminus is a nontrivial biosynthetic undertaking.[270] Not only must the replacement be biosynthesized, but also the requisite enzymatic machinery must be introduced to incorporate the replacement into Lipid II, and then to properly use the modified Lipid II for cell wall synthesis. This is accomplished in the *vanA* phenotype by seven genes operating under two promoters. One promoter activates two of these seven as regulatory genes. These two are VanR, the cytoplasmic response regulator protein, and VanS, the sensor protein.[271] The second promoter activates the remaining five, divided between three resistance proteins and two accessory proteins. The three resistance proteins are the VanH dehydrogenase, the VanA ligase, and the VanX D,D-dipeptidase. The two accessory proteins are the VanY D,D-carboxypeptidase and VanZ, of unknown function. While the

regulation and function of the *vanB*, *vanD*, *vanC*, *vanE*, and *vanG* genes are similar to *vanA*, the organization of genes (and the relative sequence homology of the proteins of the gene) have substantial differences. These aspects, as well as an overview of the two-component regulation of the *vanA* cluster via the membrane-bound histidine kinase VanS, are reviewed by Depardieu *et al.*[73] While the molecular basis for activation of the VanS sensor of *vanA* is not known, some circumstantial evidence suggests that activation may occur as a result of Lipid II accumulation.[96] In contrast, the *vanB* cluster in enterococci appears to be directly activated by the glycopeptide (or the glycopeptide-Lipid II complex).[272] The origin of these genes is adaptation of the autoimmunity mechanism used by the Gram-positive *Streptomyces* that biosynthesize these glycopeptides,[263,273–276] and these genes are widespread within nonpathogenic soil bacteria.[7,277] The third resistance enzyme, VanX,[278–285] is a zinc-dependent dipeptidase that imparts high-level vancomycin resistance by hydrolytic cleavage of the D-Ala-D-Ala dipeptide, diminishing its biosynthetic pool. VanX is abetted by the complementary[270,286] D,D-carboxypeptidase VanY.[282,287] This carboxypeptidase ensures the absence of the D-Ala-D-Ala recognition motif for the glycopeptides, by the hydrolytic deletion of the terminal D-Ala residues of any D-Ala-D-Ala-containing cell wall biosynthetic precursors (such as UDP-MurNAc-L-Ala-γ-D-Glu-L-Lys-D-Ala-D-Ala). The *vanC*, *vanE*, and *vanG* gene clusters encode a bifunctional enzyme, VanXY, to accomplish these two tasks.[288,289] As a result of the correlation between VanX activity and high-level vancomycin resistance, the possibility that VanX inhibition would synergize the activity of vancomycin against resistant pathogens (in the same way that β-lactamase inhibitors synergize β-lactamase-susceptible β-lactams) has received attention.[290–293]

Expression of *vanA* resistance is regulated at the level of transcriptional initiation at the promoters.[294] The VanH dehydrogenase catalyzes the efficient NADPH (and also NADH, depending on enzyme source)-dependent reduction of pyruvate to D-lactate.[295–297] The VanA ligase catalyzes the ATP-dependent synthesis of the D-Ala-D-Lac depsipeptide.[247,265,298–301] From this point, glycopeptide-resistant cell wall synthesis is dependent on the ability of endogenous MurF ligase to incorporate the D-Ala-D-Lac depsipeptide into Lipid II, and then the endogenous PBPs to use this Lipid II for transpeptidation-dependent stem cross-linking (where D-Lac functions as the leaving group). In order for this to occur optimization of the acceptor stem structure, accomplished by the Fem transferase enzymes,[135,302] may be necessary.[96,303] The increasing appearance of vancomycin-resistant microorganisms is indisputable evidence that the adaptation of these endogenous enzymes is hardly a barrier. Nonetheless, these enzymes are indispensible to resistance, and for this reason the murF enzyme[275,276,304–306] is regarded as an opportune target for new antibacterial discovery.[307–312]

This circumstance clearly refocuses attention on the role of the PBPs in abetting vancomycin resistance, and as well on the directly related issue of the fitness cost of use of the D-Ala-D-Lac stem structure in peptidoglycan biosynthesis by vancomycin-resistant bacteria. The particular circumstance that has galvanized efforts toward answers to these two questions is the emergence of vancomycin-resistant *S. aureus* from the vancomycin-resistant enterococci (VRE). Experiments to evaluate the fitness cost exerted by the *vanA* gene in VRE indicate the cost as minimal.[234,313,314] The dominant VRE lineage (Clonal Complex-17, CC17) in the environment[315] and in the clinic[316] typically confers multi-drug resistance (ampicillin, quinolone) and is genetically durable.[317] Evaluations of the fitness of vancomycin-resistant *S. aureus*, in the presence of β-lactam antibiotics, provide some small measure of optimism. As discussed previously, intermediate-level vancomycin resistance in *S. aureus* results primarily as a result of the biosynthesis of a thicker cell wall. This phenotype appears transiently following vancomycin challenge.[318] MRSA (methicillin-resistant *S. aureus*) achieves high-level β-lactam resistance as a result of expression of an additional PBP (PBP 2a) that is less sensitive to the β-lactam inactivation of its transpeptidase catalytic activity, relative to the endogenous PBPs, but remains vancomycin-sensitive since the transpeptidation linkage catalyzed by the PBP 2a enzyme uses a D-Ala-D-Ala donor stem as substrate. Can the PBP 2a enzyme use a D-Ala-D-Lac donor stem as substrate? If the answer to this question is positive, and if this alternative stem may be used without fitness cost, then the prospect would exist of an *S. aureus* pathogen with concurrent high-level β-lactam and glycopeptide resistance. This possibility has been evaluated independently by several experimental approaches, and with the initial conclusion that this is not a likely circumstance. In particular, the ability to use the D-Ala-D-Lac donor stem as a substrate is clearly reserved to the endogenous, and β-lactam sensitive, *S. aureus* PBP 2 enzyme and not to the *mecA*-encoded PBP 2a enzyme.[319–321] Moreover, there is an emerging consensus that in *S. aureus*, *vanA* alone exerts a measurable fitness cost.[322] Likewise, there is an appreciable fitness cost to the concurrent presence of *mecA* and *vanA* in laboratory *S. aureus* strains,[323,324] with

spontaneous deletion of *mecA* under conditions of vancomycin challenge.[325] As explicitly noted by Noto *et al.*[325] this loss of β-lactam resistance provides a rationale for the simultaneous treatment of VISA isolates with both β-lactams and glycopeptides,[326] as demonstrated by Fox *et al.*[327] in an experimental model of VRSA endocarditis.

Thus, vancomycin may be viewed correctly as a molecular 'clamp' that renders a biosynthetically essential substrate (Lipid II) unrecognizable for the PBP enzyme catalyst of cell wall biosynthesis,[328] and *vanA* gene expression may be viewed correctly as leading to the successful biosynthetic replacement of the key structural feature that stabilizes the clamp interaction. These observations might be taken to imply that an appropriate focus of medicinal chemistry efforts to synthesize new glycopeptide derivatives with improved Gram-positive activity, would be improved D-Ala-D-Ala (or D-Ala-D-Lac) affinity. The knowledge of this as an objective has been raised as a question by several observations. The ability to analyze the molecular events that confer the antibacterial properties of the glycopeptides, and impart resistance, is recent experimental accomplishment. Given the therapeutic importance of the glycopeptides, and the pressing need for better Gram-positive anti-infective agents, synthetic optimization of the glycopeptide class has proceeded using antibacterial potency and efficacy as endpoints. These efforts have yielded the semi-synthetic 'lipo' glycopeptide class,[241,329–332] exemplified by the vancomycin/teicoplanin derivatives oritavancin,[333,334] telavancin,[335] and dalbavancin;[336–338] and the lipoglycodepsipeptide ramoplanin.[250,339–342] Subsequent mechanistic evaluations of these compounds establish the lipoglycopeptides to possess antibacterial properties *beyond* their ability to recognize the D-Ala-D-Ala substructure.[233]

The existence of these new mechanisms has been demonstrated, while the molecular basis – and molecular target(s) – remain speculative. Nonetheless, since the existence of new mechanisms may profoundly influence the future development of the glycopeptide class, especially in relation to evading *vanA* resistance pathways, a summary of these observations is appropriate. Following functionalization of the saccharides of vancomycin, Kahne and co-workers observed direct interaction with the PBP hyperstructure,[343,344] consistent with the retention of membrane proteins (including PBP 1b) by affinity chromatography using a vancomycin derivative as the ligand.[345] Using structurally modified glycopeptides (abolishing their ability to bind -D-Ala-D-Ala), Leimkuhler *et al.*[249] observed significant antibacterial activity by using damaged chlorobiphenylvancomycin, modest antibacterial with damaged dalbavancin, and loss of antibacterial activity with the damaged teicoplanin structure. The retention of significant antibacterial activity by the altered chlorobiphenylvancomycin structure, despite loss of its ability to bind -D-Ala-D-Ala, was explained by detailed kinetic study in terms of direct binding to (and inhibition of the transglycosylase activity of) PBPs. Observations interpreted as involving a direct interaction of telavanicin with the bacterial cell membrane were made.[346] A final possible mechanism is peptidoglycan binding at structural motifs other than that of -D-Ala-D-Ala. A solitary D-Ala is incorporated into the teichoic acids of the Gram-positive cell wall,[347,348] and the extent of this modification correlates to a measurable increase in vancomycin resistance.[349] A second possible recognition motif was identified by Schaefer and colleagues using $^{13}C,^{19}F$ echo double NMR of various glycopeptides bound to isolated cell walls from *S. aureus* as well as whole *S. aureus* cells.[350,351] A -D-Ala-D-Ala-binding impaired oritavancin derivative was bound, whereas a -D-Ala-D-Ala-binding impaired vancomycin derivative was not bound. The NMR data imply the presence of an adjoining -Gly$_5$ stem binding site in the oritavancin structure (**Figure 9**) that enables oritavancin to bind to the peptidoglycan to inhibit the transpeptidase PBP enzymatic reaction, in addition to the transglycosylase reaction.[352,353]

To what extent these dual modes represent new opportunities for structure-based glycopeptide design, or presage yet unrecognized resistance mechanisms already devised but yet to be selected by a pathogen, is uncertain. There is no doubt, however, that the mechanistic complexity of the interference by the glycopeptides in cell wall biosynthesis, seen in terms of the apparent single focus of the *van* resistance loci (to replace -D-Ala-D-Ala with -D-Ala-D-Lac), identifies strategies that to further exploit this mechanistic diversity toward new generation glycopeptides.

8.13.4   Enzymatic Basis for Aminoglycoside Resistance

The discovery of streptomycin by Waksman[354] introduced the aminoglycoside class of antibiotics, having as a mechanism of action the inhibition of protein synthesis.[355] Among the commonly used aminoglycosides are neomycin (used topically), gentamycin, tobramycin, and amikacin. Although the clinical appearance of
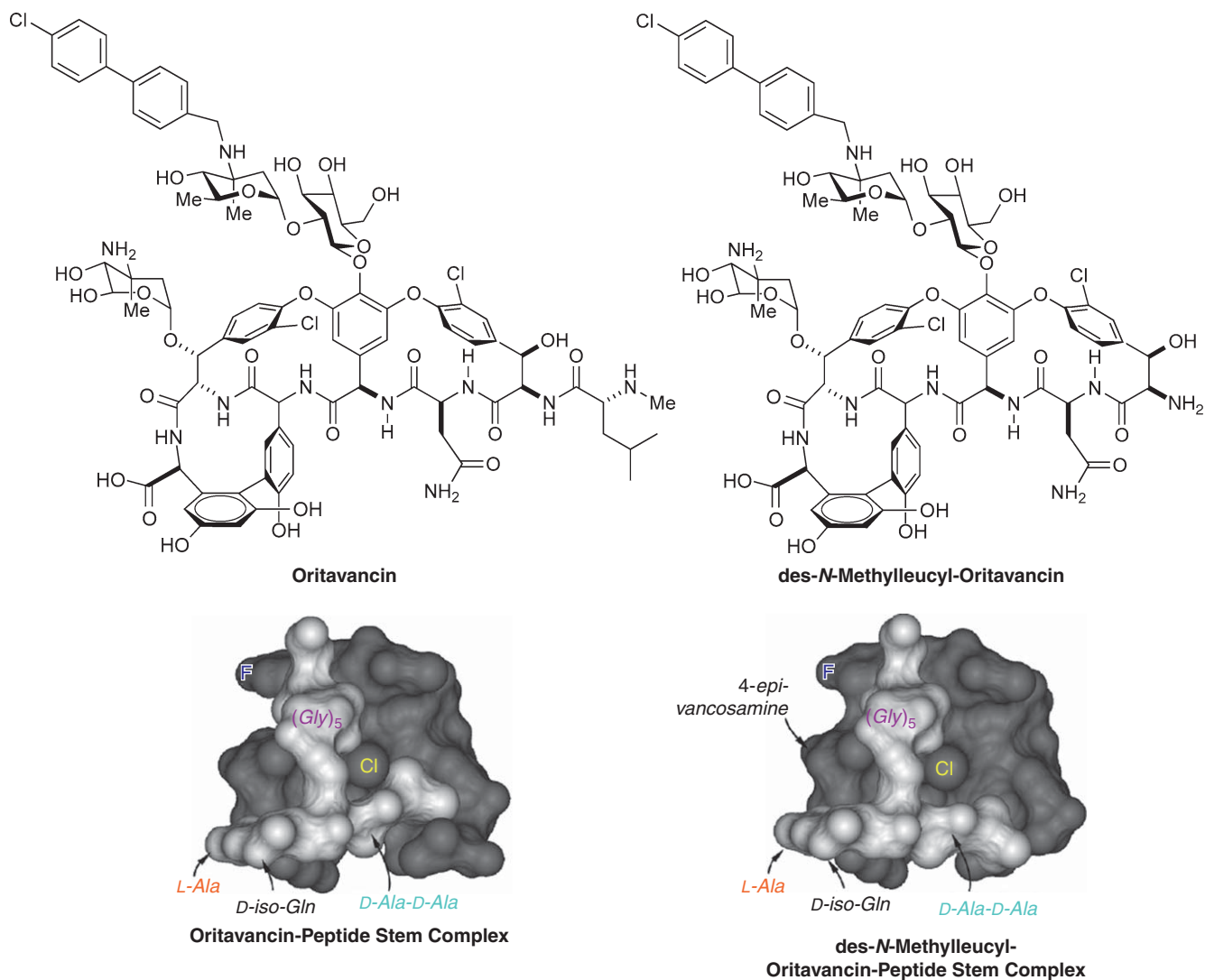
Figure 9 Space-filling model of the oritavancin-peptide stem complex (left) and des-*N*-methylleucyl–oritavancin peptide stem complex (right) assigned by NMR analysis of ^{19}F-labeled oritavanicin bound to *Staphylococcus aureus* peptidoglycan. The oritavancin complex is more than 10-fold more stable than the des-*N*-methylleucyl oritavancin peptide stem complex. The two oritavancin structures are in dark gray and the peptidoglycan stem is in light gray. In the oritavancin complex, slightly altered positions of the D-Ala-D-Ala stem terminus and the fluorobiphenyl moiety are seen. Data are reproduced from S. J. Kim; L. Cegelski; D. Stueber; M. Singh; E. Dietrich; K. S. Tanaka; T. R. J. Parr; A. R. Far; J. Schaefer, *J. Mol. Biol*. **2008**, *377* (1), 281, with permission from Elsevier.

resistance to the aminoglycosides followed several years after their introduction into clinical use, the amino-glycosides have endured as chemotherapeutic agents, especially for the treatment of infections by nosocomial Gram-negative bacteria, mycobacteria (especially multi-drug-resistant tuberculosis), and parasites.[356] Although the intrinsic toxicity of the aminoglycoside class remains as a therapeutic disadvantage – they inhibit both prokaryotic and eukaryotic peptide synthesis – medicinal interest in the aminoglycoside class has surged as the molecular mechanisms of action of this class have been revealed. Reflecting this increased interest, all aspects of the aminoglycosides have been reviewed recently, including their recognition by the RNA of the ribosome[357–361] and the structure-based design of aminoglycoside derivatives with improved therapeutic activity.[362–370] Moreover, the affinity of the aminoglycosides for polynucleotides (both DNA and RNA) has stimulated their experimental evaluation for DNA transfection,[371–373] for epigenetic therapy,[370,374–376] and as a scaffold for drug presentation to nucleic acids.[377–381] In recognition of this extraordinary diversity of mechanism for the aminoglycosides, and of their continuing chemotherapeutic value as antibiotics, the mechanisms that confer resistance to the aminoglycosides have received no less intense scrutiny.[355,382,383] Similar to the β-lactam antibiotics, enzymatic transformations are the key events in aminoglycoside resistance. These events include covalent modification of the aminoglycoside by group transfer reactions, resulting in derivatives with attenuated RNA binding, and increasingly by the direct enzymatic transformation of the RNA binding site on the ribosome to likewise affect attenuated binding affinity.

As is evident from this terse overview, mechanistically relevant biological recognition of the aminoglycoside structure occurs at the oligonucleotide (mechanism of action) and protein (resistance enzymes) level. The breadth of this recognition (encompassing both the prokaryote and eukaryote ribosome as target), when seen from the perspective of the surprising diversity of biologically active aminoglycoside structure, is astonishing. This has led to efforts to decipher the basis of polynucleotide recognition of the aminoglycosides by extensive semi-synthetic modification, as well as the evaluation of new aminoglycoside structure (and evaluation of new aminoglycoside binding sites) by array methodology.[384–388] It is now evident from these efforts to understand aminoglycoside recognition by its biological targets that structural windows of opportunity for design remain (see Hanessian *et al.*[389] for an example of current structure-based design). While there are core structural similarities within this class, there are also key differences. Particular points of difference are now found between the way aminoglycosides are recognized by the RNA of the prokaryotic and eukaryotic ribosomes,[390–392] and between recognition by the prokaryotic ribosome and (by at least some) of the prokaryotic resistance enzymes.[393] The development of these concepts concerning structural recognition, with particular reference to the enzymes involved in aminoglycoside resistance, will be developed first by presentation of aminoglycoside structures, by brief summary of the mechanism of aminoglycoside inhibition of protein synthesis, and finally by the enzymology of aminoglycoside resistance.
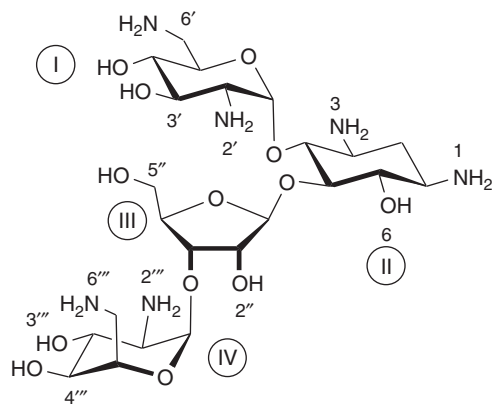
Representative structures of typical aminoglycosides are shown in **Figure 10**. The core scaffold of most aminoglycosides is 2-deoxystreptamine (the diaminocyclohexanetriol labeled as ring II in **Figure 10**), glyco-sylated (by aminodeoxysaccharides) at the 4- and 5- (or 6)-positions. By convention, the aminodeoxysaccharide linked to position 4- of 2-deoxystreptamine is ring I, and the aminodeoxysaccharide linked to position 5- (or 6) is ring III. Additional rings (IV, V) may also be present. Atypical aminoglycoside structures such as hygromycin (used as an antiparasitic in animal husbandry) and apramycin (having both Gram-negative and Gram-positive activity) are also known, and can possess interesting and nuanced biological properties. Hygromycin, for example, shares the ability to interfere with ribosome assembly with several typical aminoglycosides,[394,395] and has been used in the semi-synthesis of new highly active aminoglycoside structure.[396] **Figure 10** also summarizes the pK_a values of the amino groups of paromomycin (determined by Barbieri and Pilch[397] by ^{15}N NMR) and neomycin B (determined by Freire *et al.*)[401] as typical aminoglycosides. Interest in aminoglycoside pK_a determination reflects increasing awareness that spatial recognition of these amines (as their ammonium cations) is a critical aspect for both ribosomal recognition[398,399] and resistance enzyme recognition,[393,400] and may be an important factor in determining aminoglycoside cytoxicity (e.g., possibly as a result of the increased amine pK_a that results from deoxygenative removal of adjacent hydroxyls).[370] As is evident from the data in **Figure 10**, the differences among the aminoglycoside amines in ammonium acidity in excess of 10^3 are found. Nonetheless, the aminoglycoside structure is recognized in its nearly (if not fully) protonated state, certainly by the 'decoding' site on the 16S subunit of the ribosome[399,401–403] and by several of the resistance enzymes for which this aspect has been examined.
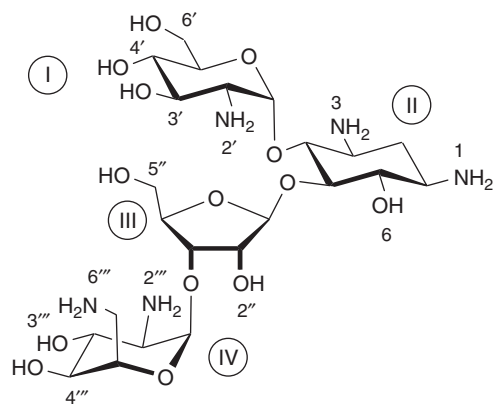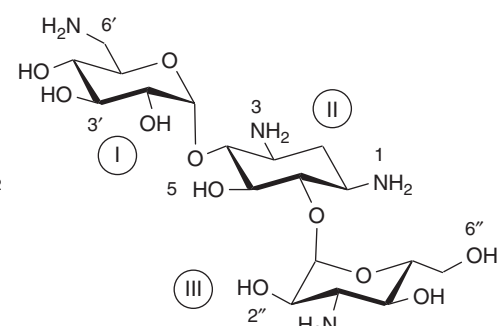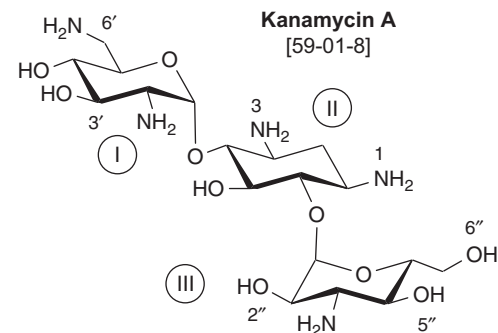
Figure 10 Structures of paromomycin, neomycin B, kanamycin A, and kanamycin B as representative aminoglycosides. The CAS Registry Number for each structure is given underneath its name. In each structure, the 2-deoxystreptamine is identified as Ring II. Paromomycin and neomycin are 4,5-diglycosylated (Rings I and III, respectively) on the 2-deoxystreptamine, while the kanamycins are 4,6-diglycosylated. The position numbers on Ring I are labeled with a prime superscript; on the 2-deoxystreptamine Ring II are unlabeled; on Ring III are labeled with a double prime superscript; and on Ring IV are labeled with a triple prime superscript. The substantial variation in basicity among the amines of the paromomycin and neomycin glycosides is emphasized by the pK_a values of the conjugate acid of each amine.
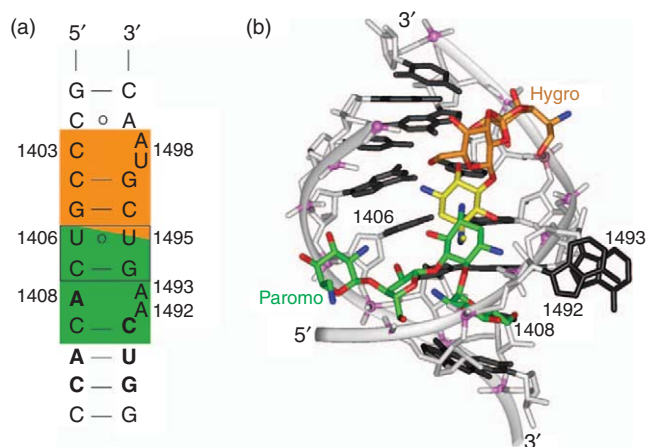
Figure 11 Aminoglycoside binding to the bacterial decoding site in 16S rRNA. (a) Secondary structure of the bacterial decoding site of helix 44 in 16S rRNA. Residues specific to the bacterial sequence are shown in bold. Nucleotides involved in the hygromycin B binding site are colored as orange, and those involved in paromomycin/neomycin B binding site are colored in green. The box outlined in black denotes the recognition site for the 2-deoxystreptamine core of the aminoglycosides (dash, classical Watson–Crick base pair; circle, non-Watson–Crick base pair). (b) Three-dimensional structure of the bacterial decoding-site rRNA in complex with paromomycin (green) and with hygromycin B (overlaid in orange and yellow). The 2-deoxystreptamine portion of hygromycin B is shown in yellow; the phosphate groups in lavender; the RNA bases in black; and the deoxyribose–phosphate backbone in light gray. Reproduced from J. A. Sutcliffe, *Curr. Opin. Microbiol.* **2005**, *8* (5), 534, with permission from Elsevier.

The antibiotic mechanism of the aminoglycosides is inhibition of protein synthesis, as a result of their recognition at the decoding A-site of the 16S rRNA site of the 30S ribosome (**Figure 11**). This summary of the consequences of this recognition is adapted from the excellent recent reviews of Sutcliffe,[359] Hainrichson *et al.*,[370] and Ogle and Ramakrishnan.[404] A critical step in aminoacyl-tRNA selection is the formation of a mini-helix between the mRNA codon and the anticodon of the matching aminoacyl-tRNA. Upon codon–anticodon pairing, the conformation of the A-site alters. In this conformational alteration, the nucleobases of the conserved A1492 and A1493 nucleotides change from an off-state where the adenines are folded within the helix, to an on-state where these adenines are flipped out and are interacting with the codon–anticodon mini-helix. Only the 'on' state is permissive for continued protein translation. When aminoglycosides are bound at this decoding site, the on-state conformation is stabilized even in the absence of bound aminoacyl-tRNA. As a result, misreading occurs, and truncated proteins are formed or incorrectly folded protein structures accumulate. Fluorescence study indicates that the reduction of mobility of the adenine of A1492 more strongly correlates with anti-bacterial activity than binding affinity.[405,406] Computational simulation of aminoglycoside induction of the transition from the off- to on-state has been presented.[407]

Intensive efforts to understand the molecular basis of aminoglycoside recognition by the decoding site, accomplished both by NMR evaluation of solution structures and crystallographic evaluation of solid state structures, have now defined the differences between recognition by the prokaryotic and eukaryotic ribosome.[390,392,408] In prokaryotic A-site recognition, the aminodeoxysaccharide I inserts into the A-site helix and stacks onto the G1491 ribonucleotide, as the deoxystreptamine ring II forms four conserved hydrogen bonds engaging the A1493, G-1494, and U1495 ribonucleotides. The saccharide rings are each in the expected chair conformations, as is also seen in the solid state.[409] The Ring I–II motif comprises of the primary structural determinant for ribosome binding, verified by crystallographic comparison of different bound aminoglycosides where these rings are seen to superimpose.[393] The role of the remaining rings are secondary (although not unimportant) interactions with the RNA in order to position, and stabilize, the presentation of rings I and II. All data concerning the biological recognition of the aminoglycosides, whether by polynucleotide or by protein, underscore the role of ammonium and hydroxyl electrostatic interactions as fundamental to the recognition event. While this realization has defined opportunity – as we shall see, the conformations used by the polynucleotide of the ribosome and the protein of the resistance enzymes are different – it has also emphasized

the difficulty in exploiting this interaction. To begin with, the dominant importance of electrostatics explains concisely how aminoglycosides are promiscuous binders of oligonucleotide structure, including ribozymes,[385] to the Gram-positive T box antiterminator RNA,[381] to tRNA[410] and to the HIV RNA.[380,402,411] Moreover, structural change may also indicate mechanistic change; the atypical aminoglycoside apramycin binds to the ribosome decoding site[412] yet has a different mechanism (inhibition of the elongation step of protein synthesis).[390,413] The core structural difficulties in understanding this diversity of structure and mechanisms include the intrinsic conformational mobility of the aminoglycoside, the dependence of this conformational mobility on the ionization state of the aminoglycoside (the fully protonated aminoglycoside is more conformationally mobile than the partially protonated aminoglycoside),[393,403] the effect of counterion on the aminoglycoside pK_a values,[397] the effect of structural change (such as is done routinely during medicinal chemistry structure–activity development) on the pK_a,[402] and last but not the least is the ability of the RNA structure itself to conformationally adjust in response to the electrostatics of ligand binding.[401,402,414] Hence, knowledge of the preferred Ring I–II conformation recognized at the ribosome decoding site, termed the *syn*-Ψ conformation wherein Ring I is proximal to Ring III (similar to the perspective shown in **Figure 10**), does not easily lend itself to the creation of conformationally biased (such as by intramolecular tethering[402,415–418]) aminoglycosides that retain biological activity. At this time, the more attractive strategy for aminoglycoside structure optimization would appear to be hypothesis generation driven by recognition of possible sub-site binding interactions.[366,389,419] An outstanding example of the potential benefits to this strategy is provided by the 4-amino-2-hydroxybutanoyl (AHB) substitution that is found on the N-2 of the deoxystreptamine rings of amikacin and butirosin.[420]

Although the introduction of such substitutions will necessarily alter the aminoglycoside pK_a values, and simultaneously introduce conformational perturbations, the latter effect may not be at all deleterious. As noted previously, the aminoglycoside conformation recognized at the decoding site and by aminoglycoside resistance enzymes, are different (**Figure 12**).[415] Bacteria possess extremely effective resistance mechanisms to withstand the aminoglycosides, and while some of these resistance mechanisms have yet to move into pathogens, the aminoglycosides are under the same level of compromise as human anti-infectives as are all other antibacterials. There are three basic mechanisms for aminoglycoside resistance. The first resistance mechanism is the customary adaptation by bacteria to limit ingress and to facilitate egress of the antibacterial, especially in the latter case by the expression of transporters.[356,421–425] Given their very hydrophilic nature, aminoglycoside uptake by bacteria is active. The transporter(s) used are not well characterized (in *E. coli,* the oligopeptide permease OppA is not used).[426] Modification of the anionic lipopolysaccharide surface of the Gram-negative bacterium can significantly reduce, or sensitize, the bacterium to aminoglycosides.[427–429] The molecular mechanisms responsible for this effect are not known. Further aspects related to aminoglycoside transport (both intake and efflux) are discussed by Magnet and Blanchard.[355]

The second resistance mechanism for the aminoglycosides is enzyme-catalyzed transformation, by group transfer, to the amino and hydroxyl groups of the aminoglycoside in order to suppress their ability to bind to the target site. The third, and newest, resistance mechanism is the modification of the RNA of the decoding site by methylation. An identical resistance mechanism operates for macrolide antibacterial binding, at another site on the ribosome, as is discussed in a following section. Because of this RNA methylation, catalyzed by methyltransferase enzymes, a steric barrier is introduced that precludes aminoglycoside binding. Since these latter two resistance mechanisms are enzyme-catalyzed, they comprise the remaining discussion on the aminoglycosides.

A large number of aminoglycoside-modifying enzymes (AME) is known. These enzymes fall into four mechanistic classes. The first enzyme class (the *N*-acetyltransferase or AAC class) catalyzes *N*-acetylation of the aminoglycoside amino groups. The second enzyme class (the *O*-nucleotidyltransferase or ANT class) catalyzes ATP-dependent nucleotidyl transfer of an AMP moiety, with release of the pyrophosphate, to hydroxyl functional groups of the aminoglycoside. The third enzyme class (the *O*-phosphotransferase or APH class) catalyzes ATP (or GTP)-dependent hydroxyl *O*-phosphorylation. Lastly, these separate catalytic entities are also found combined into single proteins that have bifunctional abilities to modify the aminoglycoside. AME nomenclature uses the three-letter prefix (AAC, ANT, or APH), followed by sub-type designation using the position number and location of the modification, and often followed by a large Roman numeral and small Roman letter designation to indicate the particular substrate(s) or resistance profiles. Thus, the most common phosphotransferase is APH(3′), where the 3′ designation indicates that it catalyzes phosphorylation of the
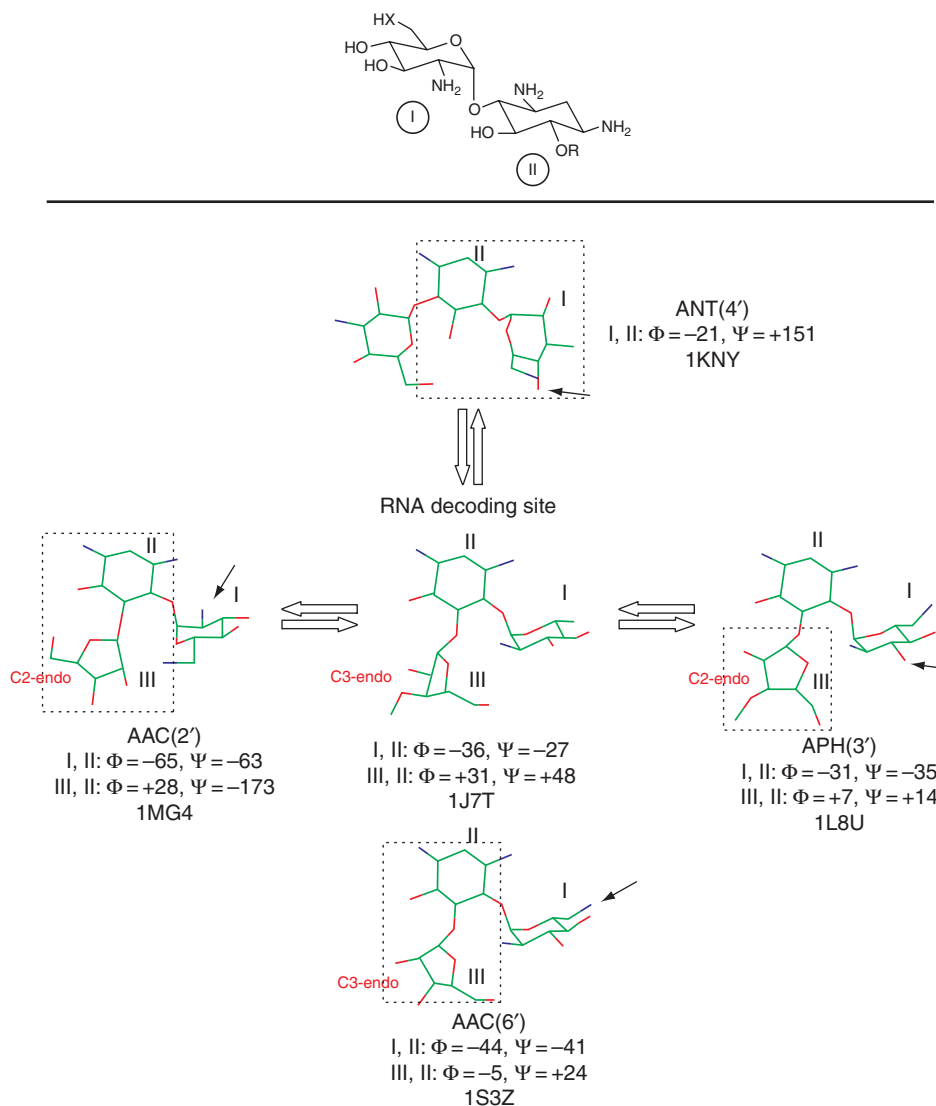
Figure 12 Schematic representation of the conformational differences observed for the bound aminoglycoside at the rRNA decoding site compared to several resistance enzymes. The aminodeoxyglucose saccharide is identified as Ring I and the 2-deoxystreptamine as Ring II. The values of the glycosidic torsion angles Φ(H1–C1–Ox–Cx) and Ψ(Hx–Cx–O1–C1) are given for each glycosidic linkage. The four enzyme structures are identified by AME type and by PDB code number. The black arrows indicate the position of group transfer to the aminoglycoside catalyzed by each enzyme. The figure is adapted from A. Bastida; A. Hidalgo; J. L. Chiara; M. Torrado; F. Corzana; J. M. Perez-Canadillas; P. Groves; E. Garcia-Junceda; C. Gonzalez; J. Jimenez-Barbero; J. L. Asensio, *J. Am. Chem. Soc.* **2006**, *128*, 100 with permission from ACS.

3-hydroxyl group found on the ring I saccharide of the aminoglycoside. By convention, the substituents of ring I are designated by a single prime superscript; the substituents of the 2-deoxystreptamine ring II by no superscript; the substituents of ring III by a double prime superscript; and the substituents of ring IV by a triple prime superscript (see **Figure 10**). Further examples include the AAC(3) family, which comprises four major types (termed AAC(3)-I to AAC(3)-IV) reflecting substrate recognition patterns, and the widespread Gram-positive plasmid-encoded bifunctional enzyme AAC(6′)-Ie-APH(2′). **Figure 13** summarizes the most common patterns for AAC, ANT, and APH aminoglycoside modification.

The AAC class contains more than 48 members, and as a class falls within the GNAT *N*-acetyl transferase superfamily.[302] The AAC class is prominently represented by enzymes capable of acetylating the N-1 and
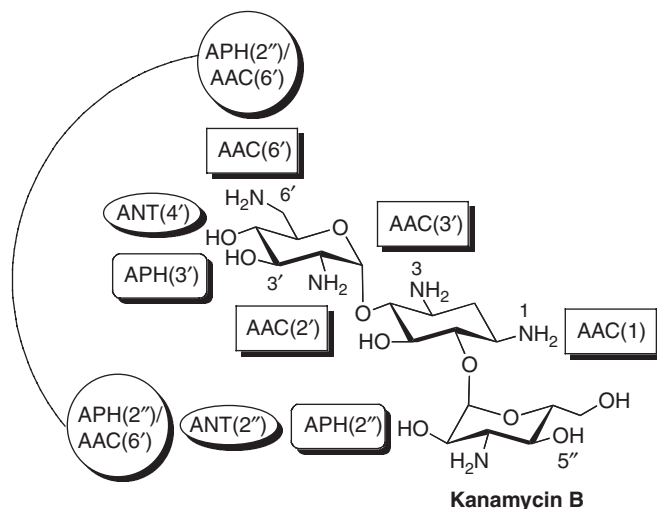
Figure 13   Target sites of enzymatic modification of the aminoglycosides that confer resistance. The three classes of enzymes are the AAC (aminoglycoside *N*-acetyltransferase), the APH (aminoglycoside *O*-phosphotransferase), and ANT (aminoglycoside *O*-nucleotidyltransferase) classes. The number in parentheses is the position on the aminoglycoside that is modified by the group (acetyl, phospho, nucleotidyl) transfer. The bifunctional enzyme is a single enzyme having separate domains capable of aminoglycoside modification by either acetyl transfer or by phosphoryl transfer. The figure is redrawn from M. Hainrichson; I. Nudelman; T. Baasov, *Org. Biol. Chem.* **2008**, *6*, 227.

N-3 positions of the Ring II deoxystreptamine and N-2′ and N-6′ (the sub-class with the largest genetic variation) of Ring I. The AAC(2′) enzymes are chromosomally encoded, although often with expression levels insufficient to provide aminoglycoside resistance. Genetic evidence[383] suggests that the acceptance of aminoglycosides by these enzymes as substrates may be adventitious (the likely endogenous substrate is the peptidoglycan),[430,431] presaging the recent adaptation of an AAC(6′) enzyme for fluoroquinolone resistance,[432] as discussed in a later section. In addition, there is strong evidence for a dual ability of *Salmonella enterica* AAC(6′)-Iy to acetylate *in vitro* eukaryotic histones, suggesting the AAC(6′) enzyme as the evolutionary progenitor of the eukaryotic histone acetyltransferases.[433] Given the molecular mechanism of the aminoglycosides, one would anticipate that the localization of these enzymes would be throughout the bacterial cytoplasm, and this has been confirmed with an AAC(6′)-Ib cyan fluorescent protein fusion. Recent progress in this class includes the isolation of new representatives, exemplified by AAC(2′)-Ia,[431] AAC(2′)-Ic from *Mycobacterium tuberculosis*,[434] the chromosomal AAC(6′)-Iy from *Salmonella enterica*,[435] the *P. aeruginosa* AAC(3) for the NMR determination of bound substrate conformation,[436] and the *Streptomyces albulus* AAC(6′).[437] Magnet *et al.*[438] observe tight aminoglycoside binding to, but not catalysis of acetyl transfer, to a protein isolated from *P. aeruginosa* that was previously annotated as an AAC(6′). Detailed kinetic and thermochemical evaluation of substrate binding was done with the *S. enterica* AAC(6′)-Iy enzyme.[439] The broad spectrum *E. coli* AAC(3)-IV follows a sequential, random bi–bi kinetic mechanism.[440] The AAC enzymes have been examined for inhibition by bisubstrate mimetics as inhibitors (indeed, they provide the example for the elucidation of the kinetic equations for sequential bireactant enzymes)[441] with encouraging success using AAC(6′)-Ii as a representative target.[442,443] Yan *et al.*[444] with this same enzyme establish the importance of the N-6 amine as a nucleophile in this acetylation reaction, and as well for 16S rRNA binding. Structural manipulation of the arbekacin, especially by N6-methylation, results in its evasion of N6-acetylation by AAC(6′)-Ib.[425,445] Last, the AAC domain is an important component of the bifunctional AME class, as discussed below.

An equally large AME class are the APH phosphotransferases.[446] Most APHs belong to the APH(3′) family that is relevant to enterococci and staphylococci clinical resistance.[424,447,448] The best studied enzymes of this class are APH(2″),[449,450] APH(3′)-IIa,[451–453] and APH(3′)-IIIa.[400,454–457] This latter enzyme (APH(3′)-IIIa) has broad abilities for aminoglycoside recognition and is capable of phosphorylation of either the 3′ or 5″-hydroxyl of neomycin and paromomycin (4,5-deoxystreptamine disubstituted aminoglycosides) and the 3′-hydroxyl of

the kanamycins (4,6-disubstituted). Continuing mechanistic study of this enzyme affirms the importance of electrostatic interactions in substrate recognition,[458] including detailed active site mutation and evaluations of the thermodynamics of substrate binding.[459] Appropriate fluoro-substitution of the aminoglycoside prevents APH phosphorylation, by loss of hydroxyl nucleophilicity, with much smaller attenuation (5- to 10-fold) of antibacterial potency.[460] A low level but intrinsic ATPase activity for two representative plasmid-encoded APHs exerts a measurable fitness cost.[453] A particularly interesting recent development in the APH field is the recognition that the previously annotated APH(2″)-Ic preferentially uses GTP rather than ATP as the phosphoryl donor (and hence is renamed as an APH(2″)-IIIa).[461] The role for the conserved APH Asp190 as the general base in hydroxyl activation for phosphoryl transfer is highly probable based on experimental and computational evaluation.[449,459,461] Crystallization of APH(2″)-IIIa has been achieved recently (but not yet with a structure determination),[462] as well as other members of the APH class APH(2″)-Ib,[463] APH(9)-Ia,[464] and APH(hygromycin-7″).[465]

The smallest of the AME enzyme families are the chromosomal and plasmid-encoded ANT nucleotidyl transferases, encountered in both Gram-positive and Gram-negative pathogens. The ANT enzymes are nonetheless well studied, especially the Gram-negative ANT(2″)[466–468] and the Gram-positive ANT(6)[469] and ANT(4′) enzymes.[393,470] A detailed study by Bastida and colleagues[393] emphasizes the conformational mobility of the aminoglycoside in response to the electrostatic contacts a particular aminoglycoside encounters during binding to the enzyme.[468] Kanamycin A and B, two similar structures but having different spatial arrays of ammonium groups, bind to this enzyme in different conformations (kanamycin A is bound in a Ring I, Ring II anti-Ψ glycosidic conformation, while kanamycin B is bound in a syn-Ψ glycosidic conformation, similar to that found on the ribosome). When the kanamycins are fully protonated, as the evidence indicates that they are when enzyme-bound, the energy difference between these two conformations is negligible. Hence, while the conformation of some aminoglycosides recognized by some AME enzymes is similar to the conformation when the aminoglycoside is bound to the ribosome, may be the exception rather than the rule (**Figure 12** shows representative aminoglycoside conformations that are bound to AME enzymes).

The final class of AME enzymes are the bifunctional enzymes, single polypeptides combining two catalytic domains.[449,471] Bifunctional enzymes, and the bifunctional enzyme AAC(6′)-Ie–APH(2″)-Ia, are encountered with greater frequency especially in the Gram-positive enterococci and staphylococci (including MRSA).[447,448] Although the nomenclature is identical, the sequence and catalytic mechanism of the AAC(6′) domain of the bifunctional enzyme is different from the single domain AAC(6′) enzyme mentioned previously. Arbekacin, which is used extensively for MRSA therapy in Japan is a substrate for AAC(6′)-Ie–APH(2″)-Ia enzyme and it is difficult to argue that there is no direct correlation between the two events.[472] Optimization of aminoglycoside structure in order to evade recognition by this enzyme is a priority in medicinal chemistry.[425,473] Newly characterized bifunctional enzymes include an ANT(3″)-Ia–AAC(6″)-IId enzyme[474] from *Serratia marcescens* and an AAC(3)-Ib–AAC(6′)-Ib′ enzyme[475] from *P. aeruginosa*. The advantage of the gene fusion leading to the creation of the bifunctional enzyme is clearly the ability of the new AME enzyme to engage and inactivate a larger breadth of aminoglycoside structure.

The breadth of enzymatic methods that establish aminoglycoside resistance, by group transfer reactions that abolish ribosome affinity, is remarkable. The utility of the aminoglycosides as antibacterials is now under even further threat by the very recent proliferation of a new resistance mechanism. The new resistance mechanism is enzyme-catalyzed target modification. RNA methylation is used by bacteria for gene regulation, and by aminoglycoside-producing actinomycetes bacteria for auto-resistance.[476] Within the past 6 years,[477] plasmid-encoded[478–480] RNA methylase genes have emerged in Gram-negative pathogens.[481,482] An example is nosocomial *Acinetobacter baumanii* isolates with high-level amikacin and high-level β-lactam (including carbapenem) resistance.[483] This high level (and structurally broad-based) aminoglycoside resistance is achieved by RNA methylation (either N-7 of G1405 or N-1 of A1408). The nomenclature for these 16S methyltransferases is evolving.[484,485] At this time these methyltransferases encompass Agr/Arm/Sgm, Kam, Kmr, Ksg, and Rmt families, some of which have relatively low similarity to the actinomycetes enzymes.[476,486] Continuing efforts to characterize these enzymes are in progress (for recent progress see Vlahovicek *et al.*[487]).

8.13.5   Enzymatic Basis for Macrolide and Ketolide Resistance

The macrolides, exemplified by the 14-membered macrolactone erythromycin and the 16-membered macrolactone tylosin (**Figure 14**), are an antibacterial class comprising both natural products, and in subsequent generations, derivatives of natural products (termed semi-synthetic structures). The core macrolide structure consists of a large ring cyclic ester (macrolactone), derived by polyketide biosynthetic assembly, which is glycosylated. The size of the ring, the substituents on the ring, and the nature of the glycosylation (saccharide identity and position) all contribute significantly to defining the biological activity of the macrolides. Although the macrolides are only bacteriostatic, they continue to represent an appreciable portion of the antibacterial market (20% by sales in 2006) due to their efficacy in the treatment of community-acquired respiratory infection. A primary basis for this success is the very favorable *in vivo* distribution of these structures to respiratory tissues.

The molecular target of the macrolides is the peptidyl-transferase site of the bacterial 50S ribosome.[488–494] Remarkably, this one ribosomal rRNA site is the target of diverse antibacterial structures other than the macrolides, including the ketolide semi-synthetic macrolide sub-class (exemplified by telithromycin), phenicols (exemplified by chloramphenicol), the streptogramins (exemplified by the pristinamycins and quinuprustin), the lincosamides (exemplified by lincomycin and clindamycin), and the oxazolidinones (exemplified by linezolid). Not surprisingly, emphatic evidence of commonality among the rRNA binding sites is provided by shared resistance determinants. For this reason, abbreviations are encountered frequently to denote overlapping binding sites and common resistance mechanisms. Among these abbreviations are MLS_B (for macrolide, lincosamide, streptogramin B phenotype), MLS_BK (for macrolide, lincosamide, streptogramin B, ketolide phenotype), and PhLOPSA (for phenicol, lincosamide, oxazolidinone, pleuromutilin, and streptogramin phenotype).[73,495]

Although all of these different classes bind to the peptidyl-transferase site, their specific location within the peptidyl-transferase site (and thus mode of interaction with the ribosome)[496] is exceedingly dependent on the
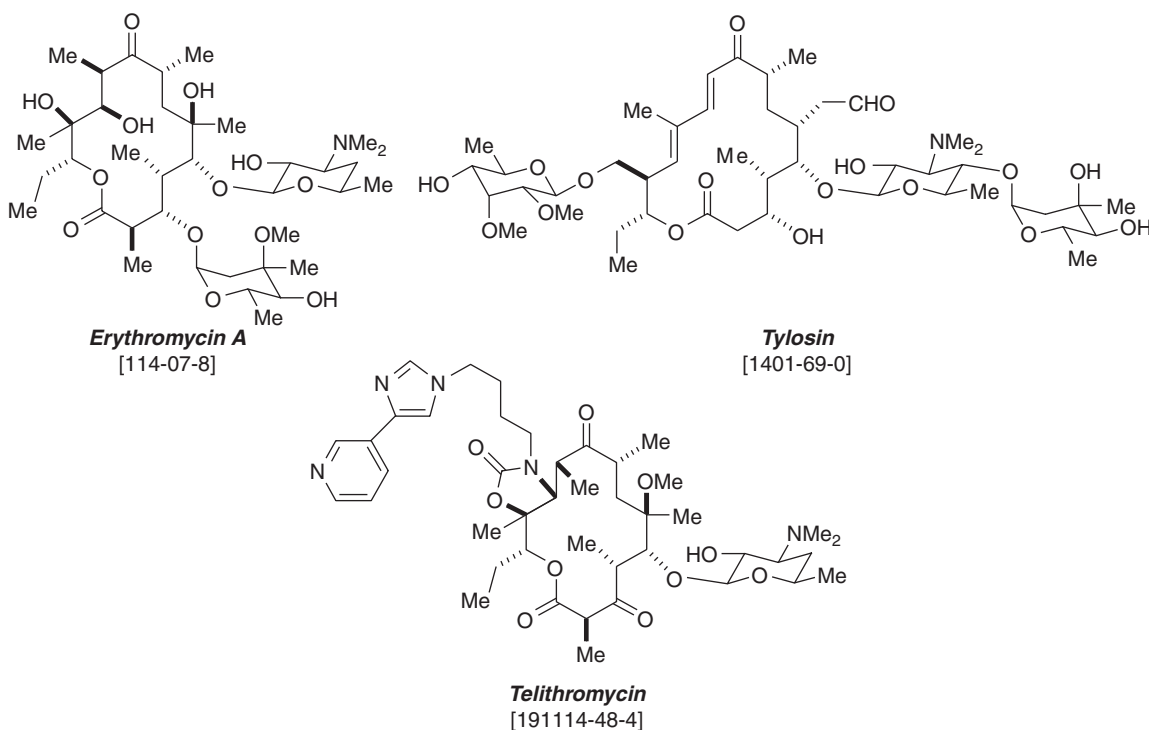
Figure 14   Representative structures of the macrolide (erythromycin and tylosin) and ketolide subclass of the macrolides (telithromycin).

structure of the particular antibacterial. Owing to the intensive effort by numerous labs, key aspects relating to the molecular mechanism of these antibacterials at the peptidyl-transferase site are understood. The following summary is adapted from a review by Poehlsgaard and Douthwaite.[492] Newly synthesized peptide passes through the 50S subunit tunnel, emerging on the back of the ribosome. The narrowest position of the tunnel, approximately one-third of its length, is a hydrophobic crevice framed by the β-hairpin elongations of the L4 and L22 ribosomal proteins (**Figure 15**). This constriction point is both a sensor and control point for peptide synthesis. Binding of the MLS$_B$K antibacterials occurs in the region of this constriction point, and with close contact of these antibacterials to the A2058 (*E. coli* numbering) 23S rRNA nucleotide.[488,497–507] Macrolides bind by a two-step kinetic mechanism. They bind first to a low-affinity site at the entrance of the exit tunnel, followed by Mg(II)-dependent slower conformational transitions to the higher affinity site (dissociation constant of 10^{-7}–10^{-8} mol l^{-1}) deeper within the tunnel.[508] As a result of macrolide binding, peptide progression through the ribosomal exit tunnel is hindered, and growth of the nascent peptide may terminate by dissociation of the short, nascent peptidyl-tRNA.[509] For erythromycin, the length of the peptide in the dissociated peptidyl-tRNA is largely six, seven, or eight amino acids.[488] In addition, as macrolide binding may occur prior to final ribosome assembly at the 23S stage, progression to the mature 50S ribosome particle is disrupted.[510] It is, however, important to retain the concept of macrolide binding as an impediment, rather than as a full blockage, of peptide synthesis. The structural studies of Steitz and co-workers indicate an ability of the nascent peptide to exit, notwithstanding the presence of the macrolide.[504] The concept of impediment, rather than barrier, explains in part, for example, the ability of polyamines to attenuate macrolide efficacy (alter binding site)[508,511] and to perturb the accuracy of translation.[512] Moreover, it is consistent with the differential
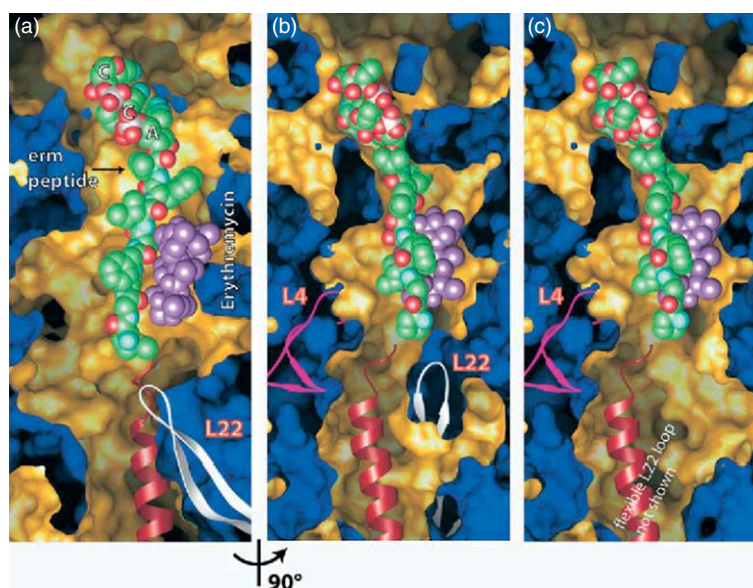
Figure 15 A model for the eight N-terminal amino acids of the *ermC* operon leader peptide bound at the PTC end of the peptide exit tunnel as proposed by Tu *et al*. The molecular surface of the interior of the nascent peptide exit tunnel is shown for the wild-type large ribosomal subunit (a and b) and the L22 deletion mutant ribosome (c). In all three views, the exit tunnel is cut in half along its length to show the diameter of the tunnel. In its lower part, polyalanine in α-helical conformation (red) is modeled to mark the general trajectory of the tunnel. Wild-type L22 (white) and L4 (magenta) are shown as ribbons. In (a), the erm peptide (with a sequence of Met–Gly–Ile–Phe–Ser–Ile–Phe–Val) is depicted in a position that would enable it to pass over bound erythromycin (purple) without interfering sterically with either the erythromycin or the surrounding ribosome. (b) The model in (a) is rotated by 90° to show that by the eighth residue, the peptide has essentially cleared the drug and its N-terminal residue is close to the β-hairpin of L22. (c) In the L22 mutant ribosome, the L4/L22 gate becomes almost twice as wide as it is in the wild-type ribosome as a result of the movement of the β-loop of L22. This gives the elongating peptide more room to get past this part of the tunnel, although the wild-type L4/L22 gate does not seem so constricted as to prevent peptide passage even when erythromycin is bound. Reproduced from D. Tu; G. Blaha; P. B. Moore; T. A. Steitz, *Cell* **2005**, *121*, 257 with permission from Elsevier.

ability of the earlier generation macrolides to induce a key macrolide resistance mechanism, whereas structurally modified macrolides – the ketolides – do not. The basis for this phenomenon is explained below.

Bacterial resistance to the macrolides invariably corresponds to the synergistic operation of drug exporters[513–519] with enzymatic alteration of the ribosomal binding site (target modification).[520] The recent survey of macrolide resistance in *Streptococcus pneumoniae* by Wierzbowski *et al.*[521] gives representative data. Of 865 macrolide-resistant isolates, 47% were positive for the *mef*(A) transporter; 43% were positive for target modification (by rRNA methylation, *vide infra*), 5% were positive for both, and 5% were negative for both (resistance due to other ribosome target modifications). Direct enzyme-catalyzed structural modification of the macrolides, similar to what is observed for the aminoglycosides, is not yet a major resistance mechanism. Nonetheless, bacteria exhibiting macrolide resistance as a result of covalent modification are found,[491] including modification by enzymatic hydrolysis, phosphorylation,[522,523] and glycosylation.[524] Enzymatic glycosylation likely has origins in host cell immunity. Target modification of the ribosome is the most important enzymatic resistance mechanism, and probably also is a host cell antibiotic immunity mechanism. Several types of ribosome modifications are known, and the repertoire of modifications that provide macrolide resistance is increasing. The primary resistance mechanism remains – it was first observed short after the clinical introduction of the macrolides – S-adenosylmethionine-dependent methylation and dimethylation, of the N^6 nitrogen of the A2058 ribonucleotide. The enzyme catalysts for these methylations are the *erm* (erythromycin resistance methylase) methylases. As noted previously, the A2058 ribonucleotide is found at the macrolide binding site of the ribosome. In the *Haloarcula* ribosome–macrolide complex, the ribonucleotide that is the cognate of the *E. coli* A2058 ribonucleotide engages in noncovalent contact with the desosamine saccharide (including a hydrogen bond between the 2′-OH of desosamine with N1 of adenine).[504] Monomethylation of A2058 (to form m^6A) results in MLS_B type I resistance phenotype (low-to-moderate macrolide and streptogramin resistance, high lincosamine resistance) while dimethylation (to form m_2^6A) gives the type II MLS_B resistance phenotype (high resistance to all MLS_B antibiotics, including the ketolide telithromycin).[492,525] The structural basis for resistance as a result of A2058 (di)methylation is simple steric interference of MLS_B antibiotic binding.

The implementation of this resistance mechanism, however, depends on two factors: the ability of the antibiotic to induce *erm* methylase expression, and the substrate specificity of the methylase(s) of the bacterium.[526,527] Ketolides are semi-synthetic derivatives of erythromycin, characterized by a ketone rather than the secondary alcohol substituent at the β-position to the lactone carbonyl and the absence of the cladinose saccharide, as both are found in erythromycin. Telithromycin is shown as a representative ketolide in **Figure 14**. The improved clinical efficacy of the ketolides against macrolide-resistant pathogens was believed to correspond to an inability of the ketolides to induce the *erm* methylases. Although the correlation of specific macrolide structure to relative *erm* expression is firmly established,[525] it is now recognized that the ketolides are indeed capable of inducing *erm* methylase expression, but much less effectively compared to early generation macrolides such as erythromycin.[528] For this reason, the progressive appearance of ketolide resistance exhibits alternative ribosome modifications,[529] such as ribonucleotide sequence optimization within the 2057–2611 ribonucleotide domain V of the 23S rRNA[521,530] and mutation of A2058 to G (or U2609 to C) coupled with alterations within the L22 protein.[531–533] Finally, the synthesis of specific small pentapeptides represents another emerging mechanism for macrolide resistance, as a result of sequence-specific interactions between the nascent peptide and the macrolide–ribosome complex.[534] The mechanism for this resistance pathway proposed by Lovmar *et al.*,[507] on the basis of detailed cell-free kinetics, is concurrent ejection of the pentapeptidyl-tRNA *and* the macrolide.

The implication of this mechanism is the operation of sophisticated mechanisms to monitor, evaluate, and respond to stalled ribosomal peptide synthesis. A very similar process occurs for *erm* methylase induction by macrolide binding. Although far from all of the details of methylase induction in response to stalled peptide synthesis, key aspects of this resistance mechanism are known. An excellent summary of these aspects, as well as an overview of inducible and constitutive *erm* resistance phenotypes, is provided by Depardieu *et al.*[73] As shown by Weisblum and Dubnau, induction arises post-transcriptionally. The *erm* mRNA is synthesized in an inactive conformation, wherein four inverted repeats at the 5′ end sequester the ribosome binding site and initiation codon. Stalled translation is believed to provoke a conformational rearrangement of the mRNA, resulting in unmasking of the initiation sequences with concurrent protection of the RNA from RNAase degradation (resulting in a dramatic increase in the lifetime of the mRNA). The tetrapeptide IFVI leader sequence of the *E. coli erm*(C) methylase (capable of m_2^6A synthesis) specifically correlates to successful induction,[535,536] again emphasizing the subtle structural interplay at the macrolide ribosomal binding site.[537] Vazquez-Laslop *et al.*[538]

summarize in detail the present understanding of the molecular mechanism of erythromycin-dependent ribosome stalling during translation of the *ermC* regulatory leader peptide cistron. Stalling occurs when expression of the *C*-terminal sequence of the nascent peptide of this protein reaches nine amino acids in length (MGIFSIFVI$_9$...). The structural basis for stalling is steric interference of the cladinose saccharide of the bound erythromycin with the IFVI$_9$ tetrapeptide segment, resulting in reorientation of the A2062 nucleotide of the 23S rRNA as a 'trigger' (**Figure 16**). Owing to this reorientation, the peptidyl transferase center is rendered incapable of further peptide bond formation, enabling release of the *ermC* ribosome and activation of the expression of the methylase gene. A comparison of the erythromycin-induced stalling mechanism to other stalling mechanisms led Vazquez-Laslop *et al.*[538] to suggest that stalling may represent a general phenomenon for adjustment of the rate of polypeptide synthesis in response to the specific requirements of expression, folding, and targeting of proteins.

Although RNA methylation is a general mechanism for bacterial gene regulation[539] and a specific mechanism for aminoglycoside and macrolide resistance, the implementation of ribosomal methylation as a macrolide resistance mechanism depends on the methylase type and characteristic within the bacterial genome. For
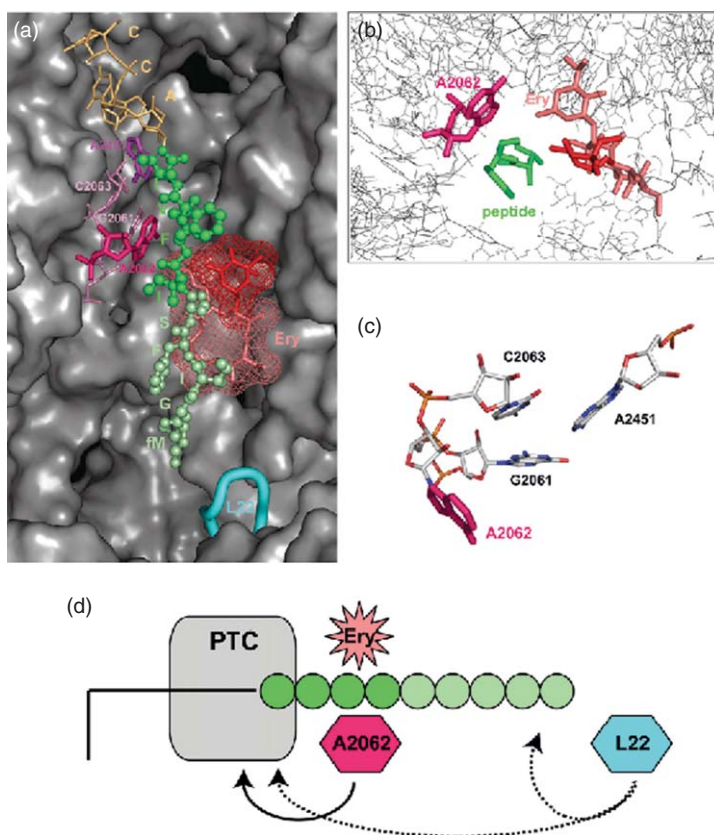

Figure 16   The model for the mechanism of macrolide-dependent ribosome stalling proposed by Vazquez-Laslop *et al*. (a) A model of the nine amino acid-long ErmCL nascent peptide (shown in ball-and-stick) in the exit tunnel of the erythromycin-bound ribosome. The five N-terminal amino acid residues are colored pale green, and the four C-terminal residues, critical for stalling, are shown in bright green. Peptide sequence is indicated. The CCA 30 end of the P site-bound peptidyl-tRNA is colored wheat. Erythromycin is shown as salmon-colored sticks with the cladinose sugar highlighted in red. The van der Waals surface of the drug is represented by a mesh. A2062 of 23S rRNA is shown in hot pink, C2063 and G2061 are pale pink, and A2451 is magenta. The β loop of protein L22 is colored cyan. (b) Relative location of the nascent peptide, erythromycin, and A2062 in the exit tunnel (viewed from the PTC down the tunnel). (c) Relative location of A2062, exposed in the tunnel, linked to nucleotides in the PTC active site of the *Escherichia coli* ribosome. (d) A general model of the drug- and nascent peptide-dependent stalling. Solid arrow marks communication of a signal from A2062 to the PTC active site. Dashed arrows indicate possible contribution of L22 (or other tunnel elements) to establishing the inactive conformation of the PTC or stabilizing peptidyl-tRNA in the ribosome. Reproduced from N. Vazquez-Laslop; C. Thum; A. S. Mankin, *Mol. Cell* **2008**, *30* (2), 190, with permission from Elsevier.

example, tylosin (a widely used veterinary macrolide) resistance in a Gram-positive *Streptomyces* results from the synergistic monomethylation of *two* 23S ribonucleotides (G748 and A2058).[506] While this dual methylation imparts tylosin (and mycinamycin) resistance, it is unsuccessful at imparting erythromycin resistance. Among bacterial pathogens, it is observed that *erm* gene expression is both inducible and constitutive. Widespread inducible MLSB resistance is found in the staphylococci (including MRSA) and streptococci, and transform to constitutive expression at relatively high frequency.[73] A succinct review of the genetics of MLSKO (now adding the oxazolidinone antibacterials to the MLSK abbreviation) resistance has been presented by Roberts.[540]

   The structures of representative Erm methylases are known.[541–543] Efforts to characterize their enzymatic properties,[544] and particularly to better understand the basis for erm-type methylase recognition of the nascent ribosome (erm methylation occurs primarily, but not exclusively, at the stage of the immature 23S particle)[545] continue.[526,546–550] A new objective for the extension of the clinical efficacy of the macrolides is the identification of competitive inhibitors of Erm methylase recognition of its rRNA binding site (as opposed to inhibitors of *S*-adenosylmethionine binding as substrate).[544,546,551] Computational evaluation of the predicted Erm(C) methyltransferase, by structure-based virtual screening, has yielded micromolar potency inhibitors with *in vitro* activity (the best compound reducing the MIC for erthyromycin, against a macrolide-resistant *E. coli* due to ErmC-methylase expression, from >100 to 6.25 mg l^{-1}).[552] Traditional SAR manipulation of the macrolide structure to improve antibacterial activity continue, as evidenced by regular publications on this topic (recent representative examples).[553–561]

8.13.6   Enzymatic Basis for Resistance to the Quinolone Antibacterials

The quinolones, a synthetic class of antibacterials, are active *in vitro* against both Gram-negative bacteria (where their molecular target is DNA gyrase) and Gram-positive bacteria (where their molecular target is Topoisomerase IV).[562,563] The quinolones have substantial clinical value against several Gram-negative pathogens (including *P. aeruginosa*, *N. gonorrheae*, and *S. pneumoniae*) and against *Mycobacteria* ssp. In the five decades since their discovery, the chemical structure of the quinolone core has advanced through several generations (**Figure 17**). The incorporation of a 6-fluoro substituent in the second-generation quinolone structures (exemplified by Ciprofloxacin) resulted in the eponymous fluoroquinolone sub-class of the quinolones, and achieving a significant increase in the potency of the quinolones both against their molecular target
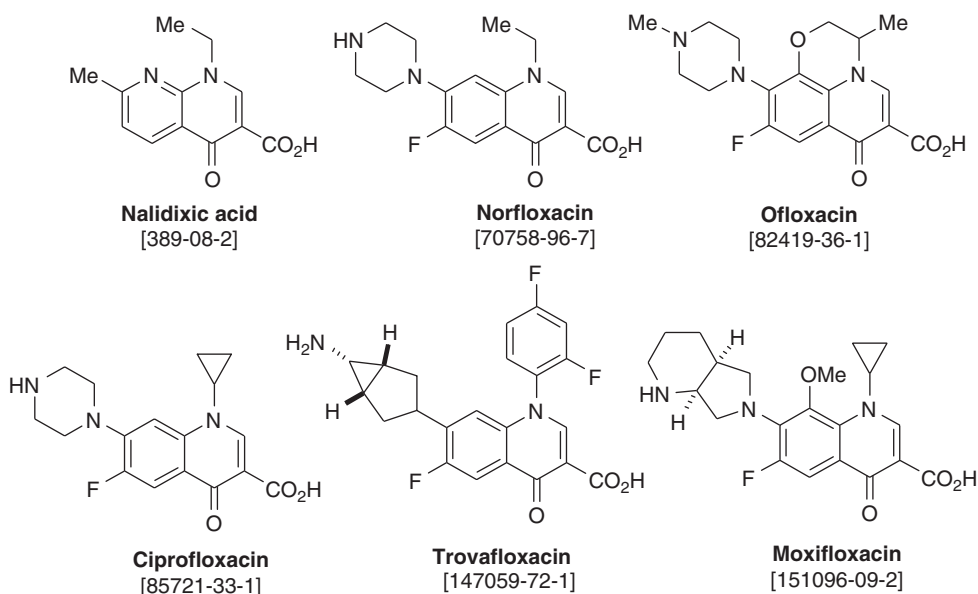

Figure 17   Representative structures of the quinolone (nalidixic acid, top left structure, as parent structure) and the fluoroquinolone (the remaining five structures) antibacterial class.

and against these pathogens.[564] The mechanism of the quinolones is a reversible binding to the enzyme–DNA binary complex, at the stage of the covalent tyrosine-phosphoester intermediate involved in transient cleavage of the DNA strand. Quinolone resistance develops by several different processes. The primary resistance mechanism for the quinolones is acquisition of transporters for the active efflux of the quinolone,[565–568] coupled with spontaneous mutation of the targets.[569–573] Following compensatory mutation, the fitness cost of the gyrase mutations can be negligible.[574]

Given that the two processes (active efflux and spontaneous target mutation) that are the primary mechanisms (to this point in time) for quinolone resistance do not involve either transformation of the anti-infective or the anti-infective target, the quinolones would not have previously been discussed in connection with enzymatic development of resistance. Their inclusion in this discussion is justified. The basis is the recent appearance of multidrug-resistance plasmids containing an enzymatic resistance factor specific to the quinolones.[575,576] While the benefit (as measured by MIC) of this quinolone resistance factor is incremental, and therefore with uncertain clinical advantage, the history of bacterial drug resistance coincides with adaptive inclusion of incremental mechanisms to attain high resistance. Accordingly, the phenomenon of plasmid-mediated quinolone resistance is being closely evaluated.[577–581] Two separate quinolone resistance factors (QRDR, quinolone resistance-determining regions) are found on these plasmids. The first factor is a gyrase (and also topoisomerase)-binding protein, termed Qnr, that are members of the larger family of pentapeptide repeat proteins (PRP). These proteins are characterized by a right-handed quadrilateral β-helix fold, which acts as a B-DNA mimetic to bind to, and thus inhibit *in vitro*, gyrase activity.[582–584] The possible mechanisms whereby the Qnr-gyrase complex[585] (and Qnr-topoisomerase complex[586]) evade the inhibitory activity of the quinolone with respect to these enzymes are suggested. The second resistance factor is indeed enzyme-catalyzed covalent modification of the quinolone. The enzyme that accomplishes covalent quinolone reaction is, remarkably, one of the aminoglycoside acetyltransferases (the AAC(6′)-Ib enzyme) that also represents a major resistance determinate for the aminoglycoside antibacterials.[587,588] Following quinolone recognition and binding by this acetyltransferase, acetyl transfer (from acetyl-CoA) occurs to the secondary amine of the piperazine-substituent found on many of the quinolones (shown in **Figure 18** using ciprofloxacin as a substrate). As not all quinolones have this piperazine substituent, this reaction is therefore not a general mechanism for quinolone resistance. Maurice *et al.*[432] and Vetting *et al.*[589] have successfully modeled ciprofloxacin into the active site of this acetyltransferase, in a position suitable for acetyl transfer to the piperazine ring.

Owing to the diversity of mechanisms used for quinolone resistance, examination of the genetics and proteomics of quinolone resistance have been undertaken, yielding interesting results. Both the quinolones and aminoglycosides induce SOS-type responses (error-prone-derived mutations) at sub-inhibitory concentrations.[590–593] The intrinsic resistance of swarming *P. aeruginosa* to quinolones (and as well as resistance to polymyxin and to the aminoglycosides) was greater than their planktonic counterparts, and correlated with increased virulence.[594] Arguably, one of the most interesting observations concerns the bacteriocidal molecular mechanism of the quinolones. Until recently, the mechanism was presumed to be the creation of DNA strand lesions and of blocked replication, directly resulting from gyrase (topoisomerase) inhibition by the quinolone. In opposition to this presumption, Dwyer *et al.*[595] observe that the causative mechanism for bacterial death is downstream from gyrase (topoisomerase) inhibition, and is the result of initial oxidative damage to the iron–sulfur centers and ensuing widespread intracellular oxidative damage. The operation of the identical
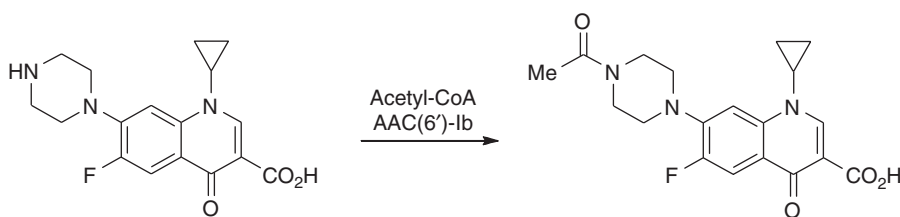
Figure 18 Enzymatic inactivation of ciprofloxacin by N-acetylation catalyzed by the aminoglycoside acetyltransferase AAC(6′)-Ib enzyme.

oxidative stress bacteriocidal mechanism is seen for two additional classes of antibacterials (the β-lactams and the aminoglycosides) having distinctively different molecular targets.[596] The existence of a single bacteriocidal mechanism, held in common with three different antibacterial classes, each class with its own molecular targets, is a forceful reminder of just how little we know, at the cellular level, of how these pathways interconnect.[597–599]

8.13.7 Conclusion

The oxymoron, that death by infection was a fact of life, was understood for centuries. During our lifetimes, the short era of the Age of Antibiotics, this cause-and-effect relationship has been largely abolished. The complacency that followed this abolition is now gone, and is replaced by sober realizations that no antibiotic is omnipotent, and that even the seemingly most innocuous bacterium has the ability to acquire and devise powerful resistance mechanisms against valued antibiotics. While many key aspects of these resistance mechanisms are known, the dynamic between antibiotic use and ensuing resistance development plays out at levels of biological depth and subtlety, of which we still know little. The answer is, as always, continued research focus, scientific perspicacity, and medical vigilance. It is premature to suggest an answer to the question posed at the start of this essay: is this the end of the Age of Antibiotics, or the start of its renaissance? It is probable, however, that the current levels of focus, perspicacity, and vigilance may only be just enough. This choice may well be one we come to regret.

Abbreviations

AAC	aminoglycoside *N*-acetyltransferase
AME	aminoglycoside modifying enzymes
ANT	aminoglycoside *O*-nucleotidyl transferase
APH	aminoglycoside phosphotransferase
ESBL	extended spectrum β-lactamase
MRSA	methicillin-resistant *Staphylococcus aureus*
PBP	penicillin binding protein
VRE	vancomycin-resistant enterococci

References

1. C. T. Walsh, *Nat. Rev. Microbiol.* **2003**, *1*, 65.
2. S. B. Levy; B. Marshall, *Nat. Med.* **2004**, *10*, S122.
3. C. Nathan, *Nature* **2004**, *431*, 899.
4. D. L. Heymann, *Cell* **2006**, *124*, 671.
5. W. Wolfson, *Chem. Biol.* **2006**, *13*, 1.
6. J. S. Bradley; R. Guidos; S. Baragona; J. G. Bartlett; E. Rubinstein; G. G. Zhanel; M. D. Tino; D. L. Pompliano; F. Tally; P. Tipirneni; G. S. Tillotson; J. H. Powers; G. S. Tillotson, *Lancet Infect. Dis.* **2007**, *7*, 68.
7. G. D. Wright, *Nat. Rev. Microbiol.* **2007**, *5*, 175.
8. P. M. Bennett, *Br. J. Pharmacol.* **2008**, *153*, S347.
9. P. M. Hawkey, *Br. J. Pharmacol.* **2008**, *153*, S406.
10. C. Walsh, *Antibiotics: Actions, Origins, Resistance*; ASM Press: Washington, DC, 2003; p 336, ISBN 1-55581-254-6.
11. D. M. Livermore, *Lancet Infect. Dis.* **2005**, *5*, 450.
12. D. J. Payne; A. MacGowan, *Curr. Opin. Pharmacol.* **2005**, *5*, 449.
13. G. L. Drusano; A. Louie; M. Deziel; T. Gumbo, *Clin. Infect. Dis.* **2006**, *42*, 525.
14. A. Sandiumenge; E. Diaz; A. Rodriguez; L. Vidaur; L. Canadell; M. Olona; M. Rue; J. Rello, *J. Antimicrob. Chemother.* **2006**, *57*, 1197.
15. J. Davies, *EMBO Rep.* **2007**, *8*, 616.
16. P. G. Davey; C. Marwick, *Clin. Microb. Infect.* **2008**, *14*, 15.
17. F. Allerberger; H. Mittermayer, *Clin. Microbiol. Infect.* **2008**, *14*, 197.
18. S. K. Olofsson; O. Cars, *Clin. Infect. Dis.* **2007**, *45* (Suppl. 2), S129.
19. S. J. Dancer, *J. Antimicrob. Chemother.* **2008**, *61*, 246.

20. J. Clardy; M. A. Fischbach; C. T. Walsh, *Nat. Biotechnol.* **2006**, *24*, 1541.
21. P. Fernandes, *Nat. Biotechnol.* **2006**, *24*, 1497.
22. G. D. Wright; A. D. Sutherland, *Trends Mol. Med.* **2007**, *13*, 260.
23. L. A. Mitscher, *J. Nat. Prod.* **2008**, *71*, 497.
24. A. F. Chalker; R. D. Lunsford, *Pharmacol. Ther.* **2002**, *95*, 1.
25. S. J. Projan, *Curr. Opin. Pharmacol.* **2002**, *2*, 513.
26. M. B. Schmid, *Nat. Biotechnol.* **2006**, *24*, 419.
27. L. L. Silver, *Nat. Rev. Drug Discov.* **2007**, *6*, 41.
28. C. Nathan; F. M. Goldberg, *Nat. Rev. Drug Discov.* **2005**, *4*, 887.
29. S. R. Norrby; C. E. Nord; R. Finch, *Lancet Infect. Dis.* **2005**, *5*, 115.
30. R. E. Christoffersen, *Nat. Biotechnol.* **2006**, *24*, 1512.
31. J. L. Fox, *Nat. Biotechnol.* **2006**, *24*, 1521.
32. S. J. Projan, *Drug Discov. Today* **2008**, *13*, 279.
33. R. E. Hancock, *Lancet Infect. Dis.* **2005**, *5*, 209.
34. M. N. Alekshun; S. B. Levy, *Biochem. Pharmacol.* **2006**, *71*, 893.
35. F. C. Tenover, *Am. J. Infect. Control* **2006**, *34*, S3.
36. M. N. Alekshun; S. B. Levy, *Cell* **2007**, *128*, 1037.
37. P. Nordmann; T. Naas; N. Fortineau; L. Poirel, *Curr. Opin. Microbiol.* **2007**, *10*, 436.
38. F. D. Lowy, *Nat. Med.* **2007**, *13*, 1418.
39. J. Chastre, *Clin. Microb. Infect.* **2008**, *14*, 3.
40. E. A. Groisman; J. Casadesus, *Mol. Microbiol.* **2005**, *56*, 1.
41. L. Cegelski; G. R. Marshall; G. R. Eldridge; S. J. Hultgren, *Nat. Rev. Microbiol.* **2008**, *6*, 17.
42. H. J. Wu; A. H. Wang; M. P. Jennings, *Curr. Opin. Chem. Biol.* **2008**, *12*, 93.
43. H. F. Chambers, *N. Engl. J. Med.* **2005**, *352*, 1485.
44. R. Wang; K. R. Braughton; D. Kretschmer; T. H. Bach; S. Y. Queck; M. Li; A. D. Kennedy; D. W. Dorward; S. J. Klebanoff; A. Peschel; F. R. Deleo; M. Otto, *Nat. Med.* **2007**, *13*, 1510.
45. J. B. Kaper; J. P. Nataro; H. L. Mobley, *Nat. Rev. Microbiol.* **2004**, *2*, 123.
46. J. B. Kaper; M. A. Karmali, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4535.
47. S. B. Levy, *Adv. Drug Deliver. Rev.* **2005**, *57*, 1446.
48. M. I. Gomez; A. Prince, *Curr. Opin. Pharmacol.* **2007**, *7*, 244.
49. C. G. Giske; L. Buaro; A. Sundsfjord; B. Wretlind, *Microb. Drug Resist.* **2008**, *14*, 23.
50. C. G. Giske; D. L. Monnet; O. Cars; Y. Carmeli, *Antimicrob. Agents Chemother.* **2008**, *52*, 813.
51. G. N. Schroeder; H. Hilbi, *Clin. Microbiol. Rev.* **2008**, *21*, 134.
52. D. A. Alfredson; V. Korolik, *FEMS Microbiol. Lett.* **2007**, *277*, 123.
53. S. Tristram; M. R. Jacobs; P. C. Appelbaum, *Clin. Microbiol. Rev.* **2007**, *20*, 368.
54. H. M. Wexler, *Clin. Microbiol. Rev.* **2007**, *20*, 593.
55. L. Dijkshoorn; A. Nemec; H. Seifert, *Nat. Rev. Microbiol.* **2007**, *5*, 939.
56. F. Perez; A. M. Hujer; K. M. Hujer; B. K. Decker; P. N. Rather; R. A. Bonomo, *Antimicrob. Agents Chemother.* **2007**, *51*, 3471.
57. A. C. Fluit, *FEMS Immunol. Med. Microbiol.* **2005**, *43*, 1.
58. E. Y. Furuya; F. D. Lowy, *Nat. Rev. Microbiol.* **2006**, *4*, 36.
59. A. Balagopal; C. L. Sears, *Curr. Opin. Pharmacol.* **2007**, *7*, 455.
60. E. J. Kuipers; M. J. Janssen; W. A. de Boer, *Curr. Opin. Pharmacol.* **2003**, *3*, 480.
61. P. Courvalin, *Clin. Infect. Dis.* **2006**, *42* (Suppl. 1), S25.
62. S. J. Brickner; S. Mobashery, *Curr. Opin. Microbiol.* **2007**, *10*, 425.
63. N. Zetola; J. S. Francis; E. L. Nuermberger; W. R. Bishai, *Lancet Infect. Dis.* **2005**, *5*, 275.
64. H. Grundmann; M. Aires-de-Sousa; J. Boyce; E. Tiemersma, *Lancet* **2006**, *368*, 874.
65. U. Seybold; E. V. Kourbatova; J. G. Johnson; S. J. Halvosa; Y. F. Wang; M. D. King; S. M. Ray; H. M. Blumberg, *Clin. Infect. Dis.* **2006**, *42*, 647.
66. D. M. Livermore; A. Pearson, *Clin. Microbiol. Infect.* **2007**, *13* (Suppl. 2), 7.
67. M. M. Mwangi; S. W. Wu; Y. Zhou; K. Sieradzki; H. de Lencastre; P. Richardson; D. Bruce; E. Rubin; E. Myers; E. D. Siggia; A. Tomasz, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9451.
68. A. Pantosti; A. Sanchini; M. Monaco, *Future Microbiol.* **2007**, *2*, 323.
69. H. de Lencastre; D. Oliveira; A. Tomasz, *Curr. Opin. Microbiol.* **2007**, *10*, 428.
70. V. M. D'Costa; K. M. McGrann; D. W. Hughes; G. D. Wright, *Science* **2006**, *311*, 374.
71. A. Tomasz, *Science* **2006**, *311*, 342.
72. B. R. Levin; D. E. Rozen, *Nat. Rev. Microbiol.* **2006**, *4*, 556.
73. F. Depardieu; I. Podglajen; R. Leclercq; E. Collatz; P. Courvalin, *Clin. Microbiol. Rev.* **2007**, *20*, 79.
74. R. I. Aminov; R. I. Mackie, *FEMS Microbiol. Lett.* **2007**, *271*, 147.
75. S. Demaneche; H. Sanguin; J. Pote; E. Navarro; D. Bernillon; P. Mavingui; W. Wildi; T. M. Vogel; P. Simonet, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3957.
76. G. D. Wright, *Adv. Drug Deliv. Rev.* **2005**, *57*, 1451.
77. H. Kresse; M. J. Belsey; H. Rovini, *Nat. Rev. Drug Discov.* **2007**, *6*, 19.
78. C. Goffin; J. M. Ghuysen, *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1079.
79. P. Macheboeuf; C. Contreras-Martel; V. Job; O. Dideberg; A. Dessen, *FEMS Microbiol. Rev.* **2006**, *30*, 673.
80. D. L. Perlstein; Y. Zhang; T. S. Wang; D. E. Kahne; S. Walker, *J. Am. Chem. Soc.* **2007**, *129*, 12674.
81. Y. Yuan; D. Barrett; Y. Zhang; D. Kahne; P. Sliz; S. Walker, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5348.
82. A. L. Lovering; L. H. de Castro; D. Lim; N. C. Strynadka, *Science* **2007**, *315*, 1402.
83. J. F. Fisher; S. O. Meroueh; S. Mobashery, *Chem. Rev.* **2005**, *105*, 395.
84. K. Poole; A. D. Russell; P. A. Lambert, *Adv. Drug Deliv. Rev.* **2005**, *57*, 1443.

85. J. M. Thomson; R. A. Bonomo, *Curr. Opin. Microbiol.* **2005**, *8*, 518.
86. H. Barreteau; A. Kovac; A. Boniface; M. Sova; S. Gobec; D. Blanot, *FEMS Microbiol. Rev.* **2008**, *32*, 168.
87. J. van Heijenoort, *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 620.
88. A. Bouhss; A. E. Trunkfield; T. D. Bugg; D. Mengin-Lecreulx, *FEMS Microbiol. Rev.* **2008**, *32*, 208.
89. E. Breukink; B. de Kruijff, *Nat. Rev. Drug Discov.* **2006**, *5*, 321.
90. E. Sauvage; F. Kerff; M. Terrak; J. A. Ayala; P. Charlier, *FEMS Microbiol. Rev.* **2008**, *32*, 234.
91. W. Vollmer; B. Joris; P. Charlier; S. Foster, *FEMS Microbiol. Rev.* **2008**, *32*, 259.
92. W. Vollmer, *FEMS Microbiol. Rev.* **2008**, *32*, 287.
93. T. den Blaauwen; M. A. de Pedro; M. Nguyen-Distèche; J. A. Ayala, *FEMS Microbiol. Rev.* **2008**, *32*, 321.
94. A. Zapun; T. Vernet; M. G. Pinho, *FEMS Microbiol. Rev.* **2008**, *32*, 345.
95. A. Zapun; C. Contreras-Martel; T. Vernet, *FEMS Microbiol. Rev.* **2008**, *32*, 361.
96. J. L. Mainardi; R. Villet; T. D. Bugg; C. Mayer; M. Arthur, *FEMS Microbiol. Rev.* **2008**, *386*, *408* 32.
97. E. P. Abraham, *J. Chemother.* **1991**, *3*, 67.
98. H. Christensen; M. T. Martin; S. G. Waley, *Biochem. J.* **1990**, *266*, 853.
99. I. Massova; S. Mobashery, *Antimicrob. Agents Chemother.* **1998**, *42*, 1.
100. N. Woodford, *Clin. Microbiol. Infect.* **2005**, *11* (Suppl. 3), 2.
101. C. C. Fuda; J. F. Fisher; S. Mobashery, *Cell Mol. Life Sci.* **2005**, *62*, 2617.
102. Y. Feng; C. J. Chen; L. H. Su; S. Hu; J. Yu; C. H. Chiu, *FEMS Microbiol. Rev.* **2008**, *32*, 23.
103. B. Guignard; J. M. Entenza; P. Moreillon, *Curr. Opin. Pharmacol.* **2005**, *5*, 479.
104. B. Berger-Bachi, *Cell Mol. Life Sci.* **1999**, *56*, 764.
105. M. G. Pinho; S. R. Filipe; H. de Lencastre; A. Tomasz, *J. Bacteriol.* **2001**, *183*, 6525.
106. S. W. Wu; H. D. Lencastre; A. Tomasz, *Microb. Drug Resist.* **2005**, *11*, 215.
107. M. T. Rahman; N. Kobayashi; M. M. Alam; M. Ishino, *Microb. Drug Resist.* **2005**, *11*, 205.
108. Y. Zhou; A. Antignac; S. W. Wu; A. Tomasz, *J. Bacteriol.* **2008**, *190*, 508.
109. S. F. Pereira; A. O. Henriques; M. G. Pinho; H. de Lencastre; A. Tomasz, *J. Bacteriol.* **2007**, *189*, 3525.
110. M. Ender; N. McCallum; R. Adhikari; B. Berger-Bachi, *Antimicrob. Agents Chemother.* **2004**, *48*, 2295.
111. S. M. Lee; M. Ender; R. Adhikari; J. M. Smith; B. Berger-Bachi; G. M. Cook, *Antimicrob. Agents Chemother.* **2007**, *51*, 1497.
112. A. Tomasz, *Lancet* **2003**, *361*, 795.
113. D. Lim; N. C. Strynadka, *Nat. Struct. Biol.* **2002**, *9*, 870.
114. T. A. Leski; A. Tomasz, *J. Bacteriol.* **2005**, *187*, 1815.
115. V. Norris; T. den Blaauwen; A. Cabin-Flaman; R. H. Doi; R. Harshey; L. Janniere; A. Jimenez-Sanchez; D. J. Jin; P. A. Levin; E. Mileykovskaya; A. Minsky; M. J. Saier; K. Skarstad, *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 230.
116. V. Norris; T. den Blaauwen; R. H. Doi; R. M. Harshey; L. Janniere; A. Jimenez-Sanchez; D. J. Jin; P. A. Levin; E. Mileykovskaya; A. Minsky; G. Misevic; C. Ripoll; M. J. Saier; K. Skarstad; M. Thellier, *Annu. Rev. Microbiol.* **2007**, *61*, 309.
117. M. G. Pinho; J. Errington, *Mol. Microbiol.* **2005**, *55*, 799.
118. C. Fuda; M. Suvorov; S. B. Vakulenko; S. Mobashery, *J. Biol. Chem.* **2004**, *279*, 40802.
119. C. Fuda; D. Hesek; M. Lee; K. Morio; T. Nowak; S. Mobashery, *J. Am. Chem. Soc.* **2005**, *127*, 2056.
120. C. Fuda; M. Suvorov; Q. Shi; D. Hesek; M. Lee; S. Mobashery, *Biochemistry* **2007**, *46*, 8050.
121. W. P. Lu; Y. Sun; M. D. Bauer; S. Paule; P. M. Koenigs; W. G. Kraft, *Biochemistry* **1999**, *38*, 6537.
122. C. Fuda; D. Hesek; M. Lee; W. Heilmayer; R. Novak; S. B. Vakulenko; S. Mobashery, *J. Biol. Chem.* **2006**, *281*, 10035.
123. T. A. Davies; M. G. Page; W. Shang; T. Andrew; M. Kania; K. Bush, *Antimicrob. Agents Chemother.* **2007**, *51*, 2621.
124. P. Macheboeuf; A. M. Di Guilmi; V. Job; T. Vernet; O. Dideberg; A. Dessen, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 577.
125. A. L. Lovering; L. De Castro; D. Lim; N. C. Strynadka, *Protein Sci.* **2006**, *15*, 1701.
126. S. Lemaire; C. Fuda; F. Van Bambeke; P. M. Tulkens; S. Mobashery, *J. Biol. Chem.* **2008**, *283*, 12769.
127. S. Lemaire; F. Van Bambeke; M. P. Mingeot-Leclercq; Y. Glupczynski; P. M. Tulkens, *Antimicrob. Agents Chemother.* **2007**, *51*, 1627.
128. R. Banerjee; M. Gretes; L. Basuino; N. Strynadka; H. F. Chambers, *Antimicrob. Agents Chemother.* **2008**, *52*, 2089.
129. B. Berger-Bachi; M. Tschierske, *Drug Resist. Updat.* **1998**, *1*, 325.
130. H. De Lencastre; S. W. Wu; M. G. Pinho; A. M. Ludovice; S. Filipe; S. Gardete; R. Sobral; S. Gill; M. Chung; A. Tomasz, *Microb. Drug Resist.* **1999**, *5*, 163.
131. M. Tschierske; C. Mori; S. Rohrer; K. Ehlert; K. J. Shaw; B. Berger-Bachi, *FEMS Microbiol. Lett.* **1999**, *171*, 97.
132. S. S. Hegde; T. E. Shrader, *J. Biol. Chem.* **2001**, *276*, 6998.
133. R. Villet; M. Fonvielle; P. Busca; M. Chemama; A. P. Maillard; J. E. Hugonnet; L. Dubost; A. Marie; N. Josseaume; S. Mesnage; C. Mayer; J. M. Valery; M. Etheve-Quelquejeu; M. Arthur, *Nucleic Acids Res.* **2007**, *35*, 6870.
134. J. Hubscher; A. Jansen; O. Kotte; J. Schafer; P. A. Majcherczyk; L. G. Harris; G. Bierbaum; M. Heinemann; B. Berger-Bachi, *BMC Genomics* **2007**, *8*, 307.
135. A. Arbeloa; J. E. Hugonnet; A. C. Sentilhes; N. Josseaume; L. Dubost; C. Monsempes; D. Blanot; J. P. Brouard; M. Arthur, *J. Biol. Chem.* **2004**, *279*, 41546.
136. A. Fajardo; J. L. Martinez, *Curr. Opin. Microbiol.* **2008**, *11*, 161.
137. I. Erill; S. Campoy; J. Barbe, *FEMS Microbiol. Rev.* **2007**, *31*, 637.
138. C. Contreras-Martel; V. Job; A. M. Di Guilmi; T. Vernet; O. Dideberg; A. Dessen, *J. Mol. Biol.* **2006**, *355*, 684.
139. R. Carapito; L. Chesnel; T. Vernet; A. Zapun, *J. Biol. Chem.* **2006**, *281*, 1771.
140. V. Job; R. Carapito; T. Vernet; A. Dessen; A. Zapun, *J. Biol. Chem.* **2008**, *283*, 4886.
141. P. Maurer; B. Koch; I. Zerfass; J. Krauss; M. van der Linden; J. M. Frere; C. Contreras-Martel; R. Hakenbeck, *J. Mol. Biol.* **2008**, *376*, 1403.
142. S. Dramsi; S. Magnet; S. Davison; M. Arthur, *FEMS Microbiol. Rev.* **2008**, *32*, 307.
143. J. L. Mainardi; M. Fourgeaud; J. E. Hugonnet; L. Dubost; J. P. Brouard; J. Ouazzani; L. B. Rice; L. Gutmann; M. Arthur, *J. Biol. Chem.* **2005**, *280*, 38146.

144. S. Magnet; A. Arbeloa; J. L. Mainardi; J. E. Hugonnet; M. Fourgeaud; L. Dubost; A. Marie; V. Delfosse; C. Mayer; L. B. Rice; M. Arthur, *J. Biol. Chem.* **2007**, *282*, 13151.
145. J. L. Mainardi; J. E. Hugonnet; F. Rusconi; M. Fourgeaud; L. Dubost; A. N. Moumi; V. Delfosse; C. Mayer; L. Gutmann; L. B. Rice; M. Arthur, *J. Biol. Chem.* **2007**, *282*, 30414.
146. B. G. Hall; M. Barlow, *J. Antimicrob. Chemother.* **2005**, *55*, 1050.
147. J. M. Frere; M. Galleni; K. Bush; O. Dideberg, *J. Antimicrob. Chemother.* **2005**, *55*, 1051.
148. F. Perez; A. Endimiani; K. M. Hujer; R. A. Bonomo, *Curr. Opin. Pharmacol.* **2007**, *7*, 459.
149. A. M. Queenan; K. Bush, *Clin. Microbiol. Rev.* **2007**, *20*, 440.
150. D. M. Livermore, *Clin. Microbiol. Infect.* **2008**, *14* (Suppl. 1), 3.
151. L. Poirel; S. Figueiredo; V. Cattoir; A. Carattoli; P. Nordmann, *Antimicrob. Agents Chemother.* **2008**, *52*, 1252.
152. A. Marrero; G. Mallorqui-Fernandez; T. Guevara; R. Garcia-Castellanos; F. X. Gomis-Ruth, *J. Mol. Biol.* **2006**, *361*, 506.
153. F. Kerff; P. Charlier; M. L. Colombo; E. Sauvage; A. Brans; J. M. Frere; B. Joris; E. Fonze, *Biochemistry* **2003**, *42*, 12835.
154. D. Golemi-Kotra; J. Y. Cha; S. O. Meroueh; S. B. Vakulenko; S. Mobashery, *J. Biol. Chem.* **2003**, *278*, 18419.
155. M. S. Wilke; T. L. Hills; H. Z. Zhang; H. F. Chambers; N. C. Strynadka, *J. Biol. Chem.* **2004**, *279*, 47278.
156. K. Thumanu; J. Cha; J. F. Fisher; R. Perrins; S. Mobashery; C. Wharton, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10630.
157. H. Mammeri; L. Poirel; P. Nordmann, *J. Antimicrob. Chemother.* **2007**, *60*, 490.
158. H. Mammeri; P. Nordmann; A. Berkani; F. Eb, *FEMS Microbiol. Lett.* **2008**, *282*, 238.
159. D. M. Livermore; R. Hope; S. Mushtaq; M. Warner, *Clin. Microbiol. Infect.* **2008**, *14* (Suppl. 1), 189.
160. M. Akova, *Clin. Microbiol. Infect.* **2008**, *14* (Suppl. 1), 185.
161. D. L. Paterson; R. A. Bonomo, *Clin. Microbiol. Rev.* **2005**, *18*, 657.
162. M. Gniadkowski, *Clin. Microbiol. Infect.* **2008**, *14* (Suppl. 1), 11.
163. D. M. Livermore; N. Woodford, *Trends Microbiol.* **2006**, *14*, 413.
164. D. M. Livermore; P. M. Hawkey, *J. Antimicrob. Chemother.* **2005**, *56*, 451.
165. J. Walther-Rasmussen; N. Hoiby, *J. Antimicrob. Chemother.* **2007**, *60*, 470.
166. J. Walther-Rasmussen; N. Hoiby, *J. Antimicrob. Chemother.* **2006**, *57*, 373.
167. T. R. Walsh; M. A. Toleman; L. Poirel; P. Nordmann, *Clin. Microbiol. Rev.* **2005**, *18*, 306.
168. T. R. Walsh, *Clin. Microbiol. Infect.* **2005**, *11*, 2.
169. M. S. Helfand; R. A. Bonomo, *Curr. Opin. Pharmacol.* **2005**, *5*, 452.
170. G. Garau; A. M. Di Guilmi; B. G. Hall, *Antimicrob. Agents Chemother.* **2005**, *49*, 2778.
171. C. Bebrone, *Biochem. Pharmacol.* **2007**, *74*, 1686.
172. Y. Yamaguchi; T. Kuroki; H. Yasuzawa; T. Higashi; W. Jin; A. Kawanami; Y. Yamagata; Y. Arakawa; M. Goto; H. Kurosaki, *J. Biol. Chem.* **2005**, *280*, 20824.
173. J. Spencer; J. Read; R. B. Sessions; S. Howell; G. M. Blackburn; S. J. Gamblin, *J. Am. Chem. Soc.* **2005**, *127*, 14439.
174. I. Garcia-Saez; J. D. Docquier; G. M. Rossolini; O. Dideberg, *J. Mol. Biol.* **2008**, *375*, 604.
175. K. De Vriendt; G. Van Driessche; B. Devreese; C. Bebrone; C. Anne; J. M. Frere; M. Galleni; J. Van Beeumen, *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 180.
176. G. R. Periyannan; A. L. Costello; D. L. Tierney; K. W. Yang; B. Bennett; M. W. Crowder, *Biochemistry* **2006**, *45*, 1313.
177. A. L. Costello; N. P. Sharma; K. W. Yang; M. W. Crowder; D. L. Tierney, *Biochemistry* **2006**, *45*, 13650.
178. A. Badarau; M. I. Page, *Biochemistry* **2006**, *45*, 11012.
179. A. Badarau; M. I. Page, *Biochemistry* **2006**, *45*, 10654.
180. A. Badarau; C. Damblon; M. I. Page, *Biochem. J.* **2007**, *401*, 197.
181. M. Dal Peraro; A. J. Vila; P. Carloni; M. L. Klein, *J. Am. Chem. Soc.* **2007**, *129*, 2808.
182. J. M. Gonzalez; F. J. Medrano Martin; A. L. Costello; D. L. Tierney; A. J. Vila, *J. Mol. Biol.* **2007**, *373*, 1141.
183. L. I. Llarrull; M. F. Tioni; J. Kowalski; B. Bennett; A. J. Vila, *J. Biol. Chem.* **2007**, *282*, 30586.
184. N. P. Sharma; C. Hajdin; S. Chandrasekar; B. Bennett; K. W. Yang; M. W. Crowder, *Biochemistry* **2006**, *45*, 10729.
185. P. Oelschlaeger; J. Pleiss, *J. Mol. Biol.* **2007**, *366*, 316.
186. F. Simona; A. Magistrato; D. M. Vera; G. Garau; A. J. Vila; P. Carloni, *Proteins* **2007**, *69*, 595.
187. J. Crisp; R. Conners; J. D. Garrity; A. L. Carenbauer; M. W. Crowder; J. Spencer, *Biochemistry* **2007**, *46*, 10664.
188. C. Bebrone; C. Anne; K. De Vriendt; B. Devreese; G. M. Rossolini; J. Van Beeumen; J. M. Frere; M. Galleni, *J. Biol. Chem.* **2005**, *280*, 28195.
189. M. Kontou; S. Pournaras; I. Kristo; A. Ikonomidis; A. N. Maniatis; C. Stathopoulos, *Biochemistry* **2007**, *46*, 13170.
190. M. W. Crowder; J. Spencer; A. J. Vila, *Acc. Chem. Res.* **2006**, *39*, 721.
191. U. Stiefel; J. Harmoinen; P. Koski; S. Kaariainen; N. Wickstrand; K. Lindevall; N. J. Pultz; R. A. Bonomo; M. S. Helfand; C. J. Donskey, *Antimicrob. Agents Chemother.* **2005**, *49*, 5190.
192. A. M. Simm; E. J. Loveridge; J. Crosby; M. B. Avison; T. R. Walsh; P. M. Bennett, *Biochem. J.* **2005**, *387*, 585.
193. L. Olsen; S. Jost; H. W. Adolph; I. Pettersson; L. Hemmingsen; F. S. Jorgensen, *Bioorg. Med. Chem.* **2006**, *14*, 2627.
194. Y. Yamaguchi; W. Jin; K. Matsunaga; S. Ikemizu; Y. Yamagata; J. Wachino; N. Shibata; Y. Arakawa; H. Kurosaki, *J. Med. Chem.* **2007**, *50*, 6647.
195. B. M. Lienard; L. E. Horsfall; M. Galleni; J. M. Frere; C. J. Schofield, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 964.
196. L. Nauton; R. Kahn; G. Garau; J. F. Hernandez; O. Dideberg, *J. Mol. Biol.* **2008**, *375*, 257.
197. B. M. Lienard; R. Huting; P. Lassaux; M. Galleni; J. M. Frere; C. J. Schofield, *J. Med. Chem.* **2008**, *51*, 684.
198. J. R. Knowles, *Acc. Chem. Res.* **1985**, *18*, 97.
199. V. L. Thomas; D. Golemi-Kotra; C. Kim; S. B. Vakulenko; S. Mobashery; B. K. Shoichet, *Biochemistry* **2005**, *44*, 9330.
200. P. S. Padayatti; M. S. Helfand; M. A. Totir; M. P. Carey; P. R. Carey; R. A. Bonomo; F. van den Akker, *J. Biol. Chem.* **2005**, *280*, 34900.
201. D. Sulton; D. Pagan-Rodriguez; X. Zhou; Y. Liu; A. M. Hujer; C. R. Bethel; M. S. Helfand; J. M. Thomson; V. E. Anderson; J. D. Buynak; L. M. Ng; R. A. Bonomo, *J. Biol. Chem.* **2005**, *280*, 35528.
202. J. M. Thomson; A. M. Distler; F. Prati; R. A. Bonomo, *J. Biol. Chem.* **2006**, *281*, 26734.
203. M. S. Helfand; M. A. Taracila; M. A. Totir; R. A. Bonomo; J. D. Buynak; F. V. Akker; P. R. Carey, *Biochemistry* **2007**, *46*, 8689.

204. J. M. Thomson; A. M. Distler; R. A. Bonomo, *Biochemistry* **2007**, *46*, 11361.

205. M. A. Totir; M. S. Helfand; M. P. Carey; A. Sheri; J. D. Buynak; R. A. Bonomo; P. R. Carey, *Biochemistry* **2007**, *46*, 8980.

206. M. Perilli; G. Celenza; F. De Santis; C. Pellegrini; C. Forcella; G. M. Rossolini; S. Stefani; G. Amicosante, *Antimicrob. Agents Chemother.* **2008**, *52*, 915.

207. M. A. Totir; J. Cha; A. Ishiwata; B. Wang; A. Sheri; V. E. Anderson; J. Buynak; S. Mobashery; P. R. Carey, *Biochemistry* **2008**, *47*, 4094.

208. R. Canton; M. I. Morosini; O. Martin; S. de la Maza; E. G. de la Pedrosa, *Clin. Microbiol. Infect.* **2008**, *14* (Suppl. 1), 53.

209. I. Plantan; L. Selic; T. Mesar; P. S. Anderluh; M. Oblak; A. Prezelj; L. Hesse; M. Andrejasic; M. Vilar; D. Turk; A. Kocijan; T. Prevec; G. Vilfan; D. Kocjan; A. Copar; U. Urleb; T. Solmajer, *J. Med. Chem.* **2007**, *50*, 4113.

210. P. N. Wyrembak; K. Babaoglu; R. B. Pelto; B. K. Shoichet; R. F. Pratt, *J. Am. Chem. Soc.* **2007**, *129*, 9548.

211. K. Babaoglu; A. Simeonov; J. J. Irwin; M. E. Nelson; B. Feng; C. J. Thomas; L. Cancian; M. P. Costi; D. A. Maltby; A. Jadhav; J. Inglese; C. P. Austin; B. K. Shoichet, *J. Med. Chem.* **2008**, *51*, 2502.

212. A. M. Venkatesan; A. Agarwal; T. Abe; H. Ushirogochi; M. Ado; T. Tsuyoshi; O. Dos Santos; Z. Li; G. Francisco; Y. I. Lin; P. J. Petersen; Y. Yang; W. J. Weiss; D. M. Shlaes; T. S. Mansour, *Bioorg. Med. Chem.* **2008**, *16*, 1890.

213. J. E. Hugonnet; J. S. Blanchard, *Biochemistry* **2007**, *46*, 11998.

214. L. W. Tremblay; J. E. Hugonnet; J. S. Blanchard, *Biochemstry* **2008**, *47*, 5312.

215. B. T. Carter; H. Lin; S. D. Goldberg; E. A. Althoff; J. Raushel; V. W. Cornish, *ChemBioChem* **2005**, *6*, 2055.

216. S. T. Lefurgy; R. M. de Jong; V. W. Cornish, *Protein Sci.* **2007**, *16*, 2636.

217. J. M. Frère, *Biochem. Pharmacol.* **1989**, *38*, 1415.

218. J. M. Frère; B. Joris; M. Crine; H. H. Martin, *Biochem. Pharmacol.* **1989**, *38*, 1427.

219. Q. Shi; S. O. Meroueh; J. F. Fisher; S. Mobashery, *J. Am. Chem. Soc.* **2008**, *130*, 9293.

220. E. Sauvage; C. Duez; R. Herman; F. Kerff; S. Petrella; J. W. Anderson; S. A. Adediran; R. F. Pratt; J. M. Frere; P. Charlier, *J. Mol. Biol.* **2007**, *371*, 528.

221. P. Macheboeuf; D. Lemaire; A. Dos Santos Martins; O. Dideberg; M. Jamin; A. Dessen, *J. Mol. Biol.* **2008**, *376*, 405.

222. L. Maveyraud; I. Massova; C. Birck; K. Miyashita; J.-P. Samama; S. Mobashery, *J. Am. Chem. Soc.* **1996**, *118*, 7435.

223. G. Minasov; X. Wang; B. K. Shoichet, *J. Am. Chem. Soc.* **2002**, *124*, 5333.

224. J. C. Hermann; C. Hensen; L. Ridder; A. J. Mulholland; H. D. Holtje, *J. Am. Chem. Soc.* **2005**, *127*, 4454.

225. J. C. Hermann; L. Ridder; H. D. Holtje; A. J. Mulholland, *Org. Biomol. Chem.* **2006**, *4*, 206.

226. Y. Chen; R. Bonnet; B. K. Shoichet, *J. Am. Chem. Soc.* **2007**, *129*, 5378.

227. S. O. Meroueh; J. F. Fisher; H. B. Schlegel; S. Mobashery, *J. Am. Chem. Soc.* **2005**, *127*, 15397.

228. R. G. Finch; G. M. Eliopoulos, *J. Antimicrob. Chemother.* **2005**, *55* (Suppl. 2), ii5.

229. R. C. J. Moellering, *Clin. Infect. Dis.* **2006**, *42* (Suppl. 1), S3.

230. D. P. Levine, *Clin. Infect. Dis.* **2006**, *42* (Suppl. 1), S5.

231. Y. Gao, *Nat. Prod. Rep.* **2002**, *19*, 100.

232. J. L. Pace; G. Yang, *Biochem. Pharmacol.* **2006**, *71*, 968.

233. D. Kahne; C. Leimkuhler; W. Lu; C. T. Walsh, *Chem. Rev.* **2005**, *105*, 425.

234. R. J. Willems; J. Top; M. van Santen; D. A. Robinson; T. M. Coque; F. Baquero; H. Grundmann; M. J. Bonten, *Emerg. Infect. Dis.* **2005**, *11*, 821.

235. E. M. Mascini; M. J. Bonten, *Clin. Microbiol. Infect.* **2005**, *11* (Suppl. 4), 43.

236. E. Tacconelli; M. A. Cataldo, *Int. J. Antimicrob. Agents* **2008**, *31*, 99.

237. K. Hiramatsu, *Lancet Infect. Dis.* **2001**, *1*, 147.

238. F. Menichetti, *Clin. Microbiol. Infect.* **2005**, *11* (Suppl. 3), 22.

239. F. M. Walsh; S. G. Amyes, *Curr. Opin. Microbiol.* **2004**, *7*, 439.

240. R. F. Pfeltz; B. J. Wilkinson, *Curr. Drug Targets Infect. Disord.* **2004**, *4*, 273.

241. F. Van Bambeke; M. P. Mingeot-Leclercq; M. J. Struelens; P. M. Tulkens, *Trends Pharmacol. Sci.* **2008**, *29*, 124.

242. J. Pootoolal; J. Neu; G. D. Wright, *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 381.

243. T. R. Walsh; R. A. Howe, *Annu. Rev. Microbiol.* **2002**, *56*, 657.

244. D. H. Williams; B. Bardsley, *Angew. Chem. Int. Ed.* **1999**, *38*, 1172.

245. M. A. Cooper; D. H. Williams, *Chem. Biol.* **1999**, *6*, 891.

246. C. Walsh, *Science* **1999**, *284*, 442.

247. V. L. Healy; I. A. Lessard; D. I. Roper; J. R. Knox; C. T. Walsh, *Chem. Biol.* **2000**, *7*, R109.

248. M. Rekharsky; D. Hesek; M. Lee; S. O. Meroueh; Y. Inoue; S. Mobashery, *J. Am. Chem. Soc.* **2006**, *128*, 7736.

249. C. Leimkuhler; L. Chen; D. Barrett; G. Panzone; B. Sun; B. Falcone; M. Oberthur; S. Donadio; S. Walker; D. Kahne, *J. Am. Chem. Soc.* **2005**, *127*, 3250.

250. X. Fang; K. Tiyanont; Y. Zhang; J. Wanner; D. Boger; S. Walker, *Mol. Biosyst.* **2006**, *2*, 69.

251. K. Tiyanont; T. Doan; M. B. Lazarus; X. Fang; D. Z. Rudner; S. Walker, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11033.

252. P. M. Pereira; S. R. Filipe; A. Tomasz; M. G. Pinho, *Antimicrob. Agents Chemother.* **2007**, *51*, 3627.

253. V. R. Matias; T. J. Beveridge, *Mol. Microbiol.* **2007**, *64*, 195.

254. F. McAleese; S. W. Wu; K. Sieradzki; P. Dunman; E. Murphy; S. Projan; A. Tomasz, *J. Bacteriol.* **2006**, *188*, 1120.

255. L. Cui; A. Iwamoto; J. Q. Lian; H. M. Neoh; T. Maruyama; Y. Horikawa; K. Hiramatsu, *Antimicrob. Agents Chemother.* **2006**, *50*, 428.

256. L. Cui; H. Murakami; K. Kuwahara-Arai; H. Hanaki; K. Hiramatsu, *Antimicrob. Agents Chemother.* **2000**, *44*, 2276.

257. H. Komatsuzawa; K. Ohta; S. Yamada; K. Ehlert; H. Labischinski; J. Kajimura; T. Fujiwara; M. Sugai, *Antimicrob. Agents Chemother.* **2002**, *46*, 75.

258. K. Sieradzki; A. Tomasz, *J. Bacteriol.* **2003**, *185*, 7103.

259. A. Reipert; K. Ehlert; T. Kast; G. Bierbaum, *Antimicrob. Agents Chemother.* **2003**, *47*, 568.

260. L. Cui; X. Ma; K. Sato; K. Okuma; F. C. Tenover; E. M. Mamizuka; C. G. Gemmell; M. N. Kim; M. C. Ploy; N. El-Solh; V. Ferraz; K. Hiramatsu, *J. Clin. Microbiol.* **2003**, *41*, 5.

261. L. Cui; E. Tominaga; H. M. Neoh; K. Hiramatsu, *Antimicrob. Agents Chemother.* **2006**, *50*, 1079.

262. J. Davies, *Science* **1994**, *264*, 375.
263. J. R. Nodwell, *J. Bacteriol.* **2007**, *189*, 3683.
264. F. Beltrametti; A. Consolandi; L. Carrano; F. Bagatin; R. Rossi; L. Leoni; E. Zennaro; E. Selva; F. Marinelli, *Antimicrob. Agents Chemother.* **2007**, *51*, 1135.
265. T. D. Bugg; G. D. Wright; S. Dutka-Malen; M. Arthur; P. Courvalin; C. T. Walsh, *Biochemistry* **1991**, *30*, 10408.
266. C. C. McComas; B. M. Crowley; D. L. Boger, *J. Am. Chem. Soc.* **2003**, *125*, 9314.
267. B. M. Crowley; D. L. Boger, *J. Am. Chem. Soc.* **2006**, *128*, 2885.
268. P. E. Reynolds; P. Courvalin, *Antimicrob. Agents Chemother.* **2005**, *49*, 21.
269. J. Cremniter; J. L. Mainardi; N. Josseaume; J. C. Quincampoix; L. Dubost; J. E. Hugonnet; A. Marie; L. Gutmann; L. B. Rice; M. Arthur, *J. Biol. Chem.* **2006**, *281*, 32254.
270. M. Arthur; F. Depardieu; P. Reynolds; P. Courvalin, *Mol. Microbiol.* **1996**, *21*, 33.
271. G. D. Wright; T. R. Holman; C. T. Walsh, *Biochemistry* **1993**, *32*, 5057.
272. S. Evers; P. Courvalin, *J. Bacteriol.* **1996**, *178*, 1302.
273. A. P. Kuzin; T. Sun; J. Jorczak-Baillass; V. L. Healy; C. T. Walsh; J. R. Knox, *Structure* **2000**, *8*, 463.
274. H. J. Hong; M. I. Hutchings; J. M. Neu; G. D. Wright; M. S. Paget; M. J. Buttner, *Mol. Microbiol.* **2004**, *52*, 1107.
275. S. Serina; F. Radice; S. Maffioli; S. Donadio; M. Sosio, *FEMS Microbiol. Lett.* **2004**, *240*, 69.
276. M. Deghorain; P. Goffin; L. Fontaine; J. L. Mainardi; R. Daniel; J. Errington; B. Hallet; P. Hols, *J. Bacteriol.* **2007**, *189*, 4332.
277. L. Guardabassi; Y. Agerso, *FEMS Microbiol. Lett.* **2006**, *259*, 221.
278. Z. Wu; G. D. Wright; C. T. Walsh, *Biochemistry* **1995**, *34*, 2455.
279. I. A. Lessard; S. D. Pratt; D. G. McCafferty; D. E. Bussiere; C. Hutchins; B. L. Wanner; L. Katz; C. T. Walsh, *Chem. Biol.* **1998**, *5*, 489.
280. P. E. Reynolds, *Cell Mol. Life Sci.* **1998**, *54*, 325.
281. D. E. Bussiere; S. D. Pratt; L. Katz; J. M. Severin; T. Holzman; C. H. Park, *Mol. Cell* **1998**, *2*, 75.
282. I. A. Lessard; C. T. Walsh, *Chem. Biol.* **1999**, *6*, 177.
283. M. Anissimova; L. Yaouancq; F. Noor; M. A. Badet-Denisot; B. Badet, *J. Pept. Res.* **2003**, *62*, 88.
284. Y. P. Chang; M. J. Tseng; Y. H. Chu, *Anal. Biochem.* **2006**, *359*, 63.
285. M. L. Matthews; G. Periyannan; C. Hajdin; T. K. Sidgel; B. Bennett; M. W. Crowder, *J. Am. Chem. Soc.* **2006**, *128*, 13050.
286. M. Arthur; F. Depardieu; L. Cabanie; P. Reynolds; P. Courvalin, *Mol. Microbiol.* **1998**, *30*, 819.
287. G. D. Wright; C. Molinas; M. Arthur; P. Courvalin; C. T. Walsh, *Antimicrob. Agents Chemother.* **1992**, *36*, 1514.
288. A. H. Podmore; P. E. Reynolds, *Eur. J. Biochem.* **2002**, *269*, 2740.
289. F. Depardieu; M. G. Bonora; P. E. Reynolds; P. Courvalin, *Mol. Microbiol.* **2003**, *50*, 931.
290. Z. Wu; C. T. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11603.
291. R. Araoz; E. Anhalt; L. Rene; M. A. Badet-Denisot; P. Courvalin; B. Badet, *Biochemistry* **2000**, *39*, 15971.
292. K. W. Yang; J. J. Brandt; L. L. Chatwood; M. W. Crowder, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1085.
293. M. W. Crowder, *Infect. Disord. Drug Targets* **2006**, *6*, 147.
294. M. Arthur; F. Depardieu; P. Courvalin, *Microbiology* **1999**, *145*, 1849.
295. M. Arthur; C. Molinas; S. Dutka-Malen; P. Courvalin, *Gene* **1991**, *103*, 133.
296. W. M. Milewski; S. Boyle-Vavra; B. Moreira; C. C. Ebert; R. S. Daum, *Antimicrob. Agents Chemother.* **1996**, *40*, 166.
297. C. G. Marshall; M. Zolli; G. D. Wright, *Biochemistry* **1999**, *38*, 8485.
298. T. D. Bugg; S. Dutka-Malen; M. Arthur; P. Courvalin; C. T. Walsh, *Biochemistry* **1991**, *30*, 2017.
299. C. Fan; P. C. Moews; C. T. Walsh; J. R. Knox, *Science* **1994**, *266*, 439.
300. S. Evers; B. Casadewall; M. Charles; S. Dutka-Malen; M. Galimand; P. Courvalin, *J. Mol. Evol.* **1996**, *42*, 706.
301. D. I. Roper; T. Huyton; A. Vagin; G. Dodson, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8921.
302. M. W. Vetting; L. P. S. de Carvalho; M. Yu; S. S. Hegde; S. Magnet; S. L. Roderick; J. S. Blanchard, *Arch. Biochem. Biophys.* **2005**, *433*, 212.
303. H. J. Hong; M. I. Hutchings; L. M. Hill; M. J. Buttner, *J. Biol. Chem.* **2005**, *280*, 13055.
304. K. L. Longenecker; G. F. Stamper; P. J. Hajduk; E. H. Fry; C. G. Jakob; J. E. Harlan; R. Edalji; D. M. Bartley; K. A. Walter; L. R. Solomon; T. F. Holzman; Y. G. Gu; C. G. Lerner; B. A. Beutel; V. S. Stoll, *Protein Sci.* **2005**, *14*, 3039.
305. R. G. Sobral; A. M. Ludovice; H. de Lencastre; A. Tomasz, *J. Bacteriol.* **2006**, *188*, 2543.
306. R. G. Sobral; A. E. Jones; S. G. Des Etages; T. J. Dougherty; R. M. Peitzsch; T. Gaasterland; A. M. Ludovice; H. de Lencastre; A. Tomasz, *J. Bacteriol.* **2007**, *189*, 2376.
307. J. A. Schouten; S. Bagga; A. J. Lloyd; G. de Pascale; C. G. Dowson; D. I. Roper; T. D. Bugg, *Mol. Biosyst.* **2006**, *2*, 484.
308. G. F. Stamper; K. L. Longenecker; E. H. Fry; C. G. Jakob; A. S. Florjancic; Y. G. Gu; D. D. Anderson; C. S. Cooper; T. Zhang; R. F. Clark; Y. Cia; C. L. Black-Schaefer; J. Owen McCall; C. G. Lerner; P. J. Hajduk; B. A. Beutel; V. S. Stoll, *Chem. Biol. Drug Des.* **2006**, *67*, 58.
309. E. Z. Baum; S. M. Crespo-Carbone; A. Klinger; B. D. Foleno; I. Turchi; M. Macielag; K. Bush, *Antimicrob. Agents Chemother.* **2007**, *51*, 4420.
310. S. A. Khedkar; A. K. Malde; E. C. Coutinho, *J. Chem. Inf. Model.* **2007**, *47*, 1839.
311. M. Kotnik; P. S. Anderluh; A. Prezelj, *Curr. Pharm. Des.* **2007**, *13*, 2283.
312. M. O. Taha; N. Atallah; A. G. Al-Bakri; C. Paradis-Bleau; H. Zalloum; K. S. Younis; R. C. Levesque, *Bioorg. Med. Chem.* **2008**, *16*, 1218.
313. M. M. Lleo; B. Bonato; C. Signoretto; P. Canepari, *Antimicrob. Agents Chemother.* **2003**, *47*, 1154.
314. A. A. Ramadhan; E. Hegedus, *J. Clin. Pathol.* **2005**, *58*, 744.
315. J. L. Caplin; G. W. Hanlon; H. D. Taylor, *Environ. Microbiol.* **2008**, *10*, 885.
316. H. L. Leavis; M. J. Bonten; R. J. Willems, *Curr. Opin. Microbiol.* **2006**, *9*, 454.
317. R. J. Willems; M. J. Bonten, *Curr. Opin. Infect. Dis.* **2007**, *20*, 384.
318. K. Sieradzki; A. Tomasz, *Antimicrob. Agents Chemother.* **2006**, *50*, 527.
319. K. Sieradzki; A. Tomasz, *J. Bacteriol.* **1999**, *181*, 7566.
320. A. Severin; K. Tabei; F. Tenover; M. Chung; N. Clarke; A. Tomasz, *J. Biol. Chem.* **2004**, *279*, 3398.

321. A. Severin; S. W. Wu; K. Tabei; A. Tomasz, *Antimicrob. Agents Chemother.* **2004**, *48*, 4566.
322. N. McCallum; H. Karauzum; R. Getzmann; M. Bischoff; P. Majcherczyk; B. Berger-Bachi; R. Landmann, *Antimicrob. Agents Chemother.* **2006**, *50*, 2352.
323. B. L. De Jonge; D. Gage; N. Xu, *Antimicrob. Agents Chemother.* **2002**, *46*, 3151.
324. R. P. Adhikari; G. C. Scales; K. Kobayashi; J. M. Smith; B. Berger-Bachi; G. M. Cook, *J. Antimicrob. Chemother.* **2004**, *54*, 360.
325. M. J. Noto; P. M. Fox; G. L. Archer, *Antimicrob. Agents Chemother.* **2008**, *52*, 1221.
326. M. W. Climo; R. L. Patron; G. L. Archer, *Antimicrob. Agents Chemother.* **1999**, *43*, 1747.
327. P. M. Fox; R. J. Lampen; K. S. Stumpf; G. L. Archer; M. W. Climo, *Antimicrob. Agents Chemother.* **2006**, *50*, 2951.
328. S. T. Weiss; N. R. McIntyre; M. L. McLaughlin; D. J. Merkler, *Drug Disc. Today* **2006**, *11*, 819.
329. J. F. Barrett, *Curr. Opin. Invest. Drugs* **2005**, *6*, 781.
330. F. von Nussbaum; M. Brands; B. Hinzen; S. Weigand; D. Habich, *Angew. Chem. Int. Ed.* **2006**, *45*, 5072.
331. F. Van Bambeke, *Curr. Opin. Invest. Drugs* **2006**, *7*, 740.
332. S. T. Micek, *Clin. Infect. Dis.* **2007**, *45* (Suppl. 3), S184.
333. R. C. Mercier; L. Hrebickova, *Expert Rev. Anti Infect. Ther.* **2005**, *3*, 325.
334. G. Poulakou; H. Giamarellou, *Expert Opin. Invest. Drugs* **2008**, *17*, 225.
335. S. N. Leonard; M. J. Rybak, *Pharmacotherapy* **2008**, *28*, 458.
336. N. Scheinfeld, *Drugs Today (Barc)* **2007**, *43*, 305.
337. G. G. Zhanel; S. Trapp; A. S. Gin; M. DeCorby; P. R. Lagace-Wiens; E. Rubinstein; D. J. Hoban; J. A. Karlowsky, *Expert Rev. Anti Infect. Ther.* **2008**, *6*, 67.
338. M. Billeter; M. J. Zervos; A. Y. Chen; J. R. Dalovisio; C. Kurukularatne, *Clin. Infect. Dis.* **2008**, *46*, 577.
339. S. Walker; L. Chen; Y. Hu; Y. Rew; D. Shin; D. L. Boger, *Chem. Rev.* **2005**, *105*, 449.
340. P. Fulco; R. P. Wenzel, *Expert Rev. Anti Infect. Ther.* **2006**, *4*, 939.
341. J. Nam; D. Shin; Y. Rew; D. L. Boger, *J. Am. Chem. Soc.* **2007**, *129*, 8747.
342. R. Ciabatti; S. I. Maffioli; G. Panzone; A. Canavesi; E. Michelucci; P. S. Tiseni; E. Marzorati; A. Checchia; M. Giannone; D. Jabes; G. Romano; C. Brunati; G. Candiani; F. Castiglione, *J. Med. Chem.* **2007**, *50*, 3077.
343. M. Ge; Z. Chen; H. R. Onishi; J. Kohler; L. L. Silver; R. Kerns; S. Fukuzawa; C. Thompson; D. Kahne, *Science* **1999**, *284*, 507.
344. R. Kerns; S. D. Dong; S. Fukuzawa; J. Carbeck; J. Kohler; L. L. Silver; D. Kahne, *J. Am. Chem. Soc.* **2000**, *122*, 12608.
345. R. Sinha Roy; P. Yang; S. Kodali; Y. Xiong; R. M. Kim; P. R. Griffin; H. R. Onishi; J. Kohler; L. L. Silver; K. Chapman, *Chem. Biol.* **2001**, *8*, 1095.
346. D. L. Higgins; R. Chang; D. V. Debabov; J. Leung; T. Wu; K. M. Krause; E. Sandvik; J. M. Hubbard; K. Kaniga; D. E. J. Schmidt; Q. Gao; R. T. Cass; D. E. Karr; B. M. Benton; P. P. Humphrey, *Antimicrob. Agents Chemother.* **2005**, *49*, 1127.
347. F. C. Neuhaus; J. Baddiley, *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 686.
348. U. L. RajBhandary; D. Soll, *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5285.
349. A. Peschel; C. Vuong; M. Otto; F. Gotz, *Antimicrob. Agents Chemother.* **2000**, *44*, 2845.
350. L. Cegelski; D. Steuber; A. K. Mehta; D. W. Kulp; P. H. Axelsen; J. Schaefer, *J. Mol. Biol.* **2006**, *357*, 1253.
351. S. J. Kim; L. Cegelski; M. Preobrazhenskaya; J. Schaefer, *Biochemistry* **2006**, *45*, 5235.
352. S. J. Kim; L. Cegelski; D. Stueber; M. Singh; E. Dietrich; K. S. Tanaka; T. R. J. Parr; A. R. Far; J. Schaefer, *J. Mol. Biol.* **2008**, *377*, 281.
353. S. J. Kim; S. Matsuoka; G. J. Patti; J. Schaefer, *Biochemistry* **2008**, *47*, 3822.
354. N. Kresge; R. D. Simoni; R. L. Hill, *J. Biol. Chem.* **2004**, *279*, e7.
355. S. Magnet; J. S. Blanchard, *Chem. Rev.* **2005**, *105*, 477.
356. S. Shakil; R. Khan; R. Zarrilli; A. U. Khan, *J. Biomed. Sci.* **2008**, *15*, 5.
357. Q. Vicens; E. Westhof, *Biopolymers* **2003**, *70*, 42.
358. D. S. Pilch; M. Kaul; C. M. Barbieri, *Top. Curr. Chem.* **2005**, *253*, 179.
359. J. A. Sutcliffe, *Curr. Opin. Microbiol.* **2005**, *8*, 534.
360. E. Westhof, *Biochimie* **2006**, *88*, 931.
361. F. Jossinet; T. E. Ludwig; E. Westhof, *Curr. Opin. Microbiol.* **2007**, *10*, 279.
362. J. Haddad; L. P. Kotra; B. Llano-Sotelo; C. Kim; E. F. J. Azucena; M. Liu; S. B. Vakulenko; C. S. Chow; S. Mobashery, *J. Am. Chem. Soc.* **2002**, *124*, 3229.
363. J. B. Tok; L. Bi, *Curr. Top. Med. Chem.* **2003**, *3*, 1001.
364. S. B. Vakulenko; S. Mobashery, *Clin. Microbiol. Rev.* **2003**, *16*, 430.
365. S. Jana; J. K. Deb, *Curr. Drug Targets* **2005**, *6*, 353.
366. S. Hanessian, *ChemMedChem* **2006**, *1*, 1301.
367. J. G. Silva; I. Carvalho, *Curr. Med. Chem.* **2007**, *14*, 1101.
368. J. Zhou; G. Wang; L. H. Zhang; X. S. Ye, *Med. Res. Rev.* **2007**, *27*, 279.
369. J. R. Thomas; P. J. Hergenrother, *Chem. Rev.* **2008**, *108*, 1171.
370. M. Hainrichson; I. Nudelman; T. Baasov, *Org. Biomol. Chem.* **2008**, *6*, 227.
371. S. Napoli; G. M. Carbone; C. V. Catapano; N. Shaw; D. P. Arya, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3467.
372. M. Sainlos; M. Hauchecorne; N. Oudrhiri; S. Zertal-Zidani; A. Aissaoui; J. P. Vigneron; J. M. Lehn; P. Lehn, *ChemBioChem* **2005**, *6*, 1023.
373. L. Elson-Schwab; O. B. Garner; M. Schuksz; J. D. Esko; Y. Tor, *J. Biol. Chem.* **2007**, *282*, 13585.
374. I. Nudelman; A. Rebibo-Sabbah; D. Shallom-Shezifi; M. Hainrichson; I. Stahl; T. Ben-Yosef; T. Baasov, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 6310.
375. A. Rebibo-Sabbah; I. Nudelman; Z. M. Ahmed; T. Baasov; T. Ben-Yosef, *Hum. Genet.* **2007**, *122*, 373.
376. T. Hermann, *Cell Mol. Life Sci.* **2007**, *64*, 1841.
377. M. Kaiser; M. Sainlos; J. M. Lehn; S. Bombard; M. P. Teulade-Fichou, *ChemBioChem* **2006**, *7*, 321.
378. M. Kaiser; A. De Cian; M. Sainlos; C. Renner; J. L. Mergny; M. P. Teulade-Fichou, *Org. Biomol. Chem.* **2006**, *4*, 1049.
379. B. Willis; D. P. Arya, *Biochemistry* **2006**, *45*, 10217.
380. V. K. Tam; D. Kwong; Y. Tor, *J. Am. Chem. Soc.* **2007**, *129*, 3257.

381. R. Anupam; L. Denapoli; A. Muchenditsi; J. V. Hines, *Bioorg. Med. Chem.* **2008**, *16*, 4466.
382. C. A. Smith; E. N. Baker, *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 143.
383. M. Vetting; S. L. Roderick; S. Hegde; S. Magnet; J. S. Blanchard, *Biochem. Soc. Trans.* **2003**, *31*, 520.
384. L. Yu; T. K. Oost; J. M. Schkeryantz; J. Yang; D. Janowick; S. W. Fesik, *J. Am. Chem. Soc.* **2003**, *125*, 4444.
385. M. D. Disney; P. H. Seeberger, *Chem. Eur. J.* **2004**, *10*, 3308.
386. F. S. Liang; W. A. Greenberg; J. A. Hammond; J. Hoffmann; S. R. Head; C. H. Wong, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12311.
387. J. L. Childs-Disney; M. Wu; A. Pushechnikov; O. Aminova; M. D. Disney, *ACS Chem. Biol.* **2007**, *2*, 745.
388. M. D. Disney; O. J. Barrett, *Biochemistry* **2007**, *46*, 11223.
389. S. Hanessian; J. Szychowski; S. S. Adhikari; G. Vasquez; P. Kandasamy; E. E. Swayze; M. T. Migawa; R. Ranken; B. Francois; J. Wirmer-Bartoschek; J. Kondo; E. Westhof, *J. Med. Chem.* **2007**, *50*, 2352.
390. J. Kondo; B. Francois; A. Urzhumtsev; E. Westhof, *Angew. Chem. Int. Ed.* **2006**, *45*, 3310.
391. J. Kondo; M. Hainrichson; I. Nudelman; D. Shallom-Shezifi; C. M. Barbieri; D. S. Pilch; E. Westhof; T. Baasov, *ChemBioChem* **2007**, *8*, 1700.
392. J. Kondo; E. Westhof, *Nucleic Acids Res.* **2008**, *36*, 2654.
393. J. Revuelta; T. Vacas; M. Torrado; F. Corzana; C. Gonzalez; J. Jimenez-Barbero; M. Menendez; A. Bastida; J. L. Asensio, *J. Am. Chem. Soc.* **2008**, *130*, 5086.
394. S. M. McGaha; W. S. Champney, *Antimicrob. Agents Chemother.* **2007**, *51*, 591.
395. C. Foster; W. S. Champney, *Arch. Microbiol.* **2008**, *189*, 441.
396. G. G. Stone; D. Girard; S. Finegan; J. Duignan; J. Schafer; M. Maloney; R. P. Zaniewski; S. J. Brickner; S. K. Wade; P. T. Le; M. D. Huband, *Antimicrob. Agents Chemother.* **2008**, *52*, 2663.
397. C. M. Barbieri; D. S. Pilch, *Biophys. J.* **2006**, *90*, 1338.
398. M. Kaul; D. S. Pilch, *Biochemistry* **2002**, *41*, 7695.
399. M. Kaul; C. M. Barbieri; J. E. Kerrigan; D. S. Pilch, *J. Mol. Biol.* **2003**, *326*, 1373.
400. C. Ozen; J. M. Malek; E. H. Serpersu, *J. Am. Chem. Soc.* **2006**, *128*, 15248.
401. F. Freire; I. Cuesta; F. Corzana; J. Revuelta; C. Gonzalez; M. Hricovini; A. Bastida; J. Jimenez-Barbero; J. L. Asensio, *Chem. Commun.* **2007**, 174.
402. K. F. Blount; F. Zhao; T. Hermann; Y. Tor, *J. Am. Chem. Soc.* **2005**, *127*, 9818.
403. F. Corzana; I. Cuesta; F. Freire; J. Revuelta; M. Torrado; A. Bastida; J. Jimenez-Barbero; J. L. Asensio, *J. Am. Chem. Soc.* **2007**, *129*, 2849.
404. J. M. Ogle; V. Ramakrishnan, *Annu. Rev. Biochem.* **2005**, *74*, 129.
405. M. Kaul; C. M. Barbieri; D. S. Pilch, *J. Am. Chem. Soc.* **2006**, *128*, 1261.
406. C. M. Barbieri; M. Kaul; M. Bozza-Hingos; F. Zhao; Y. Tor; T. Hermann; D. S. Pilch, *Antimicrob. Agents Chemother.* **2007**, *51*, 1760.
407. S. O. Meroueh; S. Mobashery, *Chem. Biol. Drug. Dev.* **2007**, *69*, 291.
408. J. Kondo; A. Urzhumtsev; E. Westhof, *Nucleic Acids Res.* **2006**, *34*, 676.
409. Y. A. Puius; T. H. Stievater; T. Srikrishnan, *Carbohydr. Res.* **2006**, *341*, 2871.
410. R. Szilaghi; S. Shahzad-ul-Hussan; T. Weimar, *ChemBioChem* **2005**, *6*, 1270.
411. D. W. Staple; V. Venditti; N. Niccolai; L. Elson-Schwab; Y. Tor; S. E. Butcher, *ChemBioChem* **2008**, *9*, 93.
412. Q. Han; Q. Zhao; S. Fish; K. B. Simonsen; D. Vourloumis; J. M. Froelich; D. Wall; T. Hermann, *Angew. Chem. Int. Ed.* **2005**, *44*, 2694.
413. T. Hermann; V. Tereshko; E. Skripkin; D. J. Patel, *Blood Cells Mol. Dis.* **2007**, *38*, 193.
414. N. Moitessier; E. Westhof; S. Hanessian, *J. Med. Chem.* **2006**, *49*, 1023.
415. A. Bastida; A. Hidalgo; J. L. Chiara; M. Torrado; F. Corzana; J. M. Perez-Canadillas; P. Groves; E. Garcia-Junceda; C. Gonzalez; J. Jimenez-Barbero; J. L. Asensio, *J. Am. Chem. Soc.* **2006**, *128*, 100.
416. S. Hanessian; J. Szychowski; N. B. Campos-Reales Pineda; A. Furtos; J. W. Keillor, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3221.
417. D. Kling; D. Hesek; Q. Shi; S. Mobashery, *J. Org. Chem.* **2007**, *72*, 5450.
418. G. F. Busscher; S. A. M. W. van den Broek; F. P. J. T. Rutjes; F. L. van Delft, *Tetrahedron* **2007**, *63*, 3183.
419. J. Kondo; K. Pachamuthu; B. Francois; J. Szychowski; S. Hanessian; E. Westhof, *ChemMedChem* **2007**, *2*, 1631.
420. J. Kondo; B. Francois; R. J. Russell; J. B. Murray; E. Westhof, *Biochimie* **2006**, *88*, 1027.
421. S. Eda; H. Maseda; T. Nakae, *J. Biol. Chem.* **2003**, *278*, 2085.
422. I. Marchand; L. Damier-Piolle; P. Courvalin; T. Lambert, *Antimicrob. Agents Chemother.* **2004**, *48*, 3298.
423. R. A. Bonomo; D. Szabo, *Clin. Infect. Dis.* **2006**, *43* (Suppl. 2), S49.
424. F. El'garch; K. Jeannot; D. Hocquet; C. Llanes-Barakat; P. Plesiat, *Antimicrob. Agents Chemother.* **2007**, *51*, 1016.
425. Y. Hiraiwa; T. Usui; Y. Akiyama; K. Maebashi; N. Minowa; D. Ikeda, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3540.
426. E. H. Nakamatsu; E. Fujihira; R. C. Ferreira; A. Balan; S. O. Costa; L. C. Ferreira, *FEMS Microbiol. Lett.* **2007**, *269*, 229.
427. J. L. Kadurugamuwa; J. S. Lam; T. J. Beveridge, *Antimicrob. Agents Chemother.* **1993**, *37*, 715.
428. J. Ramos-Aires; P. Plesiat; L. Kocjancic-Curty; T. Kohler, *Antimicrob. Agents Chemother.* **2004**, *48*, 843.
429. S. Yokota; N. Fujii, *Comp. Immunol. Microbiol. Infect. Dis.* **2007**, *30*, 97.
430. K. G. Payie; A. J. Clarke, *J. Bacteriol.* **1997**, *179*, 4106.
431. K. Franklin; A. J. Clarke, *Antimicrob. Agents Chemother.* **2001**, *45*, 2238.
432. F. Maurice; I. Broutin; I. Podglajen; P. Benas; E. Collatz; F. Dardel, *EMBO Rep.* **2008**, *9*, 344.
433. M. W. Vetting; S. Magnet; E. Nieves; S. L. Roderick; J. S. Blanchard, *Chem. Biol.* **2004**, *11*, 565.
434. S. S. Hegde; F. Javid-Majd; J. S. Blanchard, *J. Biol. Chem.* **2001**, *276*, 45876.
435. S. Magnet; T. Lambert; P. Courvalin; J. S. Blanchard, *Biochemistry* **2001**, *40*, 3700.
436. M. A. Owston; E. H. Serpersu, *Biochemistry* **2002**, *41*, 10764.
437. Y. Hamano; Y. Hoshino; S. Nakamori; H. Takagi, *J. Biochem.* **2004**, *136*, 517.
438. S. Magnet; T. A. Smith; R. Zheng; P. Nordmann; J. S. Blanchard, *Antimicrob. Agents Chemother.* **2003**, *47*, 1577.
439. S. S. Hegde; T. K. Dam; C. F. Brewer; J. S. Blanchard, *Biochemistry* **2002**, *41*, 7519.

440. M. L. B. Magalhaes; J. S. Blanchard, *Biochemistry* **2005**, *44*, 16275.
441. M. Yu; M. L. Magalhaes; P. F. Cook; J. S. Blanchard, *Biochemistry* **2006**, *45*, 14788.
442. F. Gao; X. Yan; T. Shakya; O. M. Baettig; S. Ait-Mohand-Brunet; A. M. Berghuis; G. D. Wright; K. Auclair, *J. Med. Chem.* **2006**, *49*, 5273.
443. M. L. B. Magalhaes; M. W. Vetting; F. Gao; L. Freiburger; K. Auclair; J. S. Blanchard, *Biochemistry* **2008**, *47*, 579.
444. X. Yan; F. Gao; S. Yotphan; P. Bakirtzian; K. Auclair, *Bioorg. Med. Chem.* **2007**, *15*, 2944.
445. Y. Hiraiwa; N. Minowa; T. Usui; Y. Akiyama; K. Maebashi; D. Ikeda, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 6369.
446. C. Kim; S. Mobashery, *Bioorg. Chem.* **2005**, *33*, 149.
447. N. Ardic; B. Sareyyupoglu; M. Ozyurt; T. Haznedaroglu; U. Ilga, *Microbiol. Res.* **2006**, *161*, 49.
448. R. Kelmani Chandrakanth; S. Raju; S. A. Patil, *Curr. Microbiol.* **2008**, *56*, 558.
449. D. D. Boehr; S. I. Jenkins; G. D. Wright, *J. Biol. Chem.* **2003**, *278*, 12873.
450. M. Toth; J. Zajicek; C. Kim; J. W. Chow; C. Smith; S. Mobashery; S. Vakulenko, *Biochemistry* **2007**, *46*, 5570.
451. D. Nurizzo; S. C. Shewry; M. H. Perlin; S. A. Brown; J. N. Dholakia; R. L. Fuchs; T. Deva; E. N. Baker; C. A. Smith, *J. Mol. Biol.* **2003**, *327*, 491.
452. D. E. Paschon; Z. S. Patel; M. Ostermeier, *J. Mol. Biol.* **2005**, *353*, 26.
453. C. Kim; J. Y. Cha; H. Yan; S. B. Vakulenko; S. Mobashery, *J. Biol. Chem.* **2006**, *281*, 6964.
454. P. R. Thompson; J. Schwartzenhauer; D. W. Hughes; A. M. Berghuis; G. D. Wright, *J. Biol. Chem.* **1999**, *274*, 30697.
455. D. D. Boehr; A. R. Farley; G. D. Wright; J. R. Cox, *Chem. Biol.* **2002**, *9*, 1209.
456. P. R. Thompson; D. D. Boehr; A. M. Berghuis; G. D. Wright, *Biochemistry* **2002**, *41*, 7001.
457. C. Ozen; A. L. Norris; M. L. Land; E. Tjioe; E. H. Serpersu, *Biochemistry* **2008**, *47*, 40.
458. J. Roestamadji; I. Grapsas; S. Mobashery, *J. Am. Chem. Soc.* **1995**, *117*, 11060.
459. M. Kaul; C. M. Barbieri; A. R. Srinivasan; D. S. Pilch, *J. Mol. Biol.* **2007**, *369*, 142.
460. C. Kim; J. Haddad; S. B. Vakulenko; S. O. Meroueh; Y. Wu; H. Yan; S. Mobashery, *Biochemistry* **2004**, *43*, 2373.
461. A. Badarau; Q. Shi; J. W. Chow; J. Zajicek; S. Mobashery; S. Vakulenko, *J. Biol. Chem.* **2008**, *283*, 7638.
462. L. J. Byrnes; A. Badarau; S. B. Vakulenko; C. A. Smith, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2008**, *64*, 126.
463. R. Walanj; P. Young; H. M. Baker; E. N. Baker; P. Metcalf; J. W. Chow; S. Lerner; S. Vakulenko; C. A. Smith, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2005**, *61*, 410.
464. C. T. Lemke; J. Hwang; B. Xiong; N. P. Cianciotto; A. M. Berghuis, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2005**, *61*, 606.
465. D. Iino; Y. Takakura; M. Kuroiwa; R. Kawakami; Y. Sasaki; T. Hoshino; K. Ohsawa; A. Nakamura; S. Yajima, *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2007**, *63*, 685.
466. E. Wright; E. H. Serpersu, *Protein Expr. Purif.* **2004**, *35*, 373.
467. E. Wright; E. H. Serpersu, *Biochemistry* **2005**, *44*, 11581.
468. E. Wright; E. H. Serpersu, *Biochemistry* **2006**, *45*, 10243.
469. F. Corzana; I. Cuesta; A. Bastida; A. Hidalgo; M. Latorre; C. Gonzalez; E. Garcia-Junceda; J. Jimenez-Barbero; J. L. Asensio, *Chem. Eur. J.* **2005**, *11*, 5102.
470. J. L. Asensio; A. Hidalgo; A. Bastida; M. Torrado; F. Corzana; J. L. Chiara; E. Garcia-Junceda; J. Canada; J. Jimenez-Barbero, *J. Am. Chem. Soc.* **2005**, *127*, 8278.
471. D. D. Boehr; D. M. Daigle; G. D. Wright, *Biochemistry* **2004**, *43*, 9846.
472. H. Matsuo; M. Kobayashi; T. Kumagai; M. Kuwabara; M. Sugiyama, *FEBS Lett.* **2003**, *546*, 401.
473. N. Minowa; Y. Akiyama; Y. Hiraiwa; K. Maebashi; T. Usui; D. Ikeda, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 6351.
474. C. Kim; D. Hesek; J. Zajicek; S. B. Vakulenko; S. Mobashery, *Biochemistry* **2006**, *45*, 8368.
475. C. Kim; A. Villegas-Estrada; D. Hesek; S. Mobashery, *Biochemistry* **2007**, *46*, 5270.
476. G. F. Liou; S. Yoshizawa; P. Courvalin; M. Galimand, *J. Mol. Biol.* **2006**, *359*, 358.
477. K. Yokoyama; Y. Doi; K. Yamane; H. Kurokawa; N. Shibata; K. Shibayama; T. Yagi; H. Kato; Y. Arakawa, *Lancet* **2003**, *362*, 1888.
478. J. Wachino; K. Yamane; K. Shibayama; H. Kurokawa; N. Shibata; S. Suzuki; Y. Doi; K. Kimura; Y. Ike; Y. Arakawa, *Antimicrob. Agents Chemother.* **2006**, *50*, 178.
479. P. Bogaerts; M. Galimand; C. Bauraing; A. Deplano; R. Vanhoof; R. De Mendonca; H. Rodriguez-Villalobos; M. Struelens; Y. Glupczynski, *J. Antimicrob. Chemother.* **2007**, *59*, 459.
480. J. Wachino; K. Shibayama; H. Kurokawa; K. Kimura; K. Yamane; S. Suzuki; N. Shibata; Y. Ike; Y. Arakawa, *Antimicrob. Agents Chemother.* **2007**, *51*, 4401.
481. K. Yamane; J. Wachino; Y. Doi; H. Kurokawa; Y. Arakawa, *Emerg. Infect. Dis.* **2005**, *11*, 951.
482. Y. S. Yu; H. Zhou; Q. Yang; Y. G. Chen; L. J. Li, *J. Antimicrob. Chemother.* **2007**, *60*, 454.
483. Y. Doi; J. M. Adams; K. Yamane; D. L. Paterson, *Antimicrob. Agents Chemother.* **2007**, *51*, 4209.
484. T. R. Fritsche; M. Castanheira; G. H. Miller; R. N. Jones; E. S. Armstrong, *Antimicrob. Agents Chemother.* **2008**, *52*, 1843.
485. Y. Doi; J. I. Wachino; Y. Arakawa, *Antimicrob. Agents Chemother.* **2008**, *52*, 2287.
486. J. Y. Zhao; Z. J. Xia; X. Sun; L. Zhong; D. M. Jiang; H. Liu; J. Wang; Z. J. Qin; Y. Z. Li, *Biochem. Biophys. Res. Commun.* **2008**, *370*, 140.
487. G. M. Vlahovicek; S. Cubrilo; K. L. Tkaczuk; J. M. Bujnicki, *Biochim. Biophys. Acta* **2008**, *1784*, 582.
488. T. Tenson; M. Lovmar; M. Ehrenberg, *J. Mol. Biol.* **2003**, *330*, 1005.
489. R. Jain; L. H. Danziger, *Curr. Pharm. Des.* **2004**, *10*, 3045.
490. B. S. Schuwirth; M. A. Borovinskaya; C. W. Hau; W. Zhang; A. Vila-Sanjurjo; J. M. Holton; J. H. Cate, *Science* **2005**, *310*, 827.
491. L. Katz; G. W. Ashley, *Chem. Rev.* **2005**, *105*, 499.
492. J. Poehlsgaard; S. Douthwaite, *Nat. Rev. Microbiol.* **2005**, *3*, 870.
493. T. Hermann, *Curr. Opin. Struct. Biol.* **2005**, *15*, 355.
494. T. Tenson; A. Mankin, *Mol. Microbiol.* **2006**, *59*, 1664.
495. K. S. Long; J. Poehlsgaard; C. Kehrenberg; S. Schwarz; B. Vester, *Antimicrob. Agents Chemother.* **2006**, *50*, 2500.
496. M. Johansson; M. Lovmar; M. Ehrenberg, *Curr. Opin. Microbiol.* **2008**, *11*, 141.

497. F. Schlunzen; R. Zarivach; J. Harms; A. Bashan; A. Tocilj; R. Albrecht; A. Yonath; F. Franceschi, *Nature* **2001**, *413*, 814.
498. R. Berisio; J. Harms; F. Schluenzen; R. Zarivach; H. A. Hansen; P. Fucini; A. Yonath, *J. Bacteriol.* **2003**, *185*, 4276.
499. F. Schlunzen; J. M. Harms; F. Franceschi; H. A. Hansen; H. Bartels; R. Zarivach; A. Yonath, *Structure* **2003**, *11*, 329.
500. A. Yonath, *Annu. Rev. Biochem.* **2005**, *74*, 649.
501. A. Yonath, *Mol. Cells* **2005**, *20*, 1.
502. J. L. Hansen; J. A. Ippolito; N. Ban; P. Nissen; P. B. Moore; T. A. Steitz, *Mol. Cell* **2002**, *10*, 117.
503. J. L. Hansen; P. B. Moore; T. A. Steitz, *J.Mol. Biol.* **2003**, *330*, 1061.
504. D. Tu; G. Blaha; P. B. Moore; T. A. Steitz, *Cell,* **2005**, *121*, 257.
505. B. Vester; S. Douthwaite, *Antimicrob. Agents Chemother.* **2001**, *45*, 1.
506. M. Liu; S. Douthwaite, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14658.
507. M. Lovmar; K. Nilsson; V. Vimberg; T. Tenson; M. Nervall; M. Ehrenberg, *J. Biol. Chem.* **2006**, *281*, 6742.
508. A. D. Petropoulos; E. C. Kouvela; G. P. Dinos; D. L. Kalpaxis, *J. Biol. Chem.* **2008**, *283*, 4756.
509. M. Lovmar; T. Tenson; M. Ehrenberg, *J. Biol. Chem.* **2004**, *279*, 53506.
510. W. S. Champney; C. L. Tober, *Curr. Microbiol.* **1998**, *37*, 418.
511. G. Garza-Ramos; L. Xiong; P. Zhong; A. Mankin, *J. Bacteriol.* **2001**, *183*, 6898.
512. J. Thompson; C. A. Pratt; A. E. Dahlberg, *Antimicrob. Agents Chemother.* **2004**, *48*, 4889.
513. C. Cagliero; C. Mouline; S. Payot; A. Cloeckaert, *J. Antimicrob. Chemother.* **2005**, *56*, 948.
514. L. Mamelli; V. Prouzet-Mauleon; J. M. Pages; F. Megraud; J. M. Bolla, *J. Antimicrob. Chemother.* **2005**, *56*, 491.
515. A. K. Wierzbowski; D. Boyd; M. Mulvey; D. J. Hoban; G. G. Zhanel, *Antimicrob. Agents Chemother.* **2005**, *49*, 4635.
516. S. Stepanovic; A. Martel; I. Dakic; A. Decostere; D. Vukovic; L. Ranin; L. A. Devriese; F. Haesebrouck, *Microb. Drug Resist.* **2006**, *12*, 115.
517. T. Bogdanovich; B. Bozdogan; P. C. Appelbaum, *Antimicrob. Agents Chemother.* **2006**, *50*, 893.
518. M. Kurincic; N. Botteldoorn; L. Herman; S. Smole Mozina, *Int. J. Food Microbiol.* **2007**, *120*, 186.
519. A. Gibreel; N. M. Wetsch; D. E. Taylor, *Antimicrob. Agents Chemother.* **2007**, *51*, 3212.
520. M. C. Roberts, *Mol. Biotechnol.* **2004**, *28*, 47.
521. A. K. Wierzbowski; K. Nichol; N. Laing; T. Hisanaga; A. Nikulin; J. A. Karlowsky; D. J. Hoban; G. G. Zhanel, *J. Antimicrob. Chemother.* **2007**, *60*, 733.
522. R. Szczepanowski; I. Krahn; N. Bohn; A. Puhler; A. Schluter, *Antimicrob. Agents Chemother.* **2007**, *51*, 673.
523. O. Chesneau; K. Tsvetkova; P. Courvalin, *FEMS Microbiol. Lett.* **2007**, *269*, 317.
524. D. N. Bolam; S. Roberts; M. R. Proctor; J. P. Turkenburg; E. J. Dodson; C. Martinez-Fleites; M. Yang; B. G. Davis; G. J. Davies; H. J. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5336.
525. B. Weisblum, *Drug Resist. Updat.* **1998**, *1*, 29.
526. I. D. Villsen; B. Vester; S. Douthwaite, *J. Mol. Biol.* **1999**, *286*, 365.
527. K. Buriankova; F. Doucet-Populaire; O. Dorson; A. Gondran; J. C. Ghnassia; J. Weiser; J. L. Pernodet, *Antimicrob. Agents Chemother.* **2004**, *48*, 143.
528. M. Bailey; T. Chettiath; A. S. Mankin, *Antimicrob. Agents Chemother.* **2008**, *52*, 866.
529. D. N. Wilson; J. M. Harms; K. H. Nierhaus; F. Schlunzen; P. Fucini, *Biol. Chem.* **2005**, *386*, 1239.
530. P. Pfister; N. Corti; S. Hobbie; C. Bruell; R. Zarivach; A. Yonath; E. C. Bottger, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 5180.
531. L. Xiong; Y. Korkhin; A. S. Mankin, *Antimicrob. Agents Chemother.* **2005**, *49*, 281.
532. R. Berisio; N. Corti; P. Pfister; A. Yonath; E. C. Bottger, *Antimicrob. Agents Chemother.* **2006**, *50*, 3816.
533. S. Zaman; M. Fitzpatrick; L. Lindahl; J. Zengel, *Mol. Microbiol.* **2007**, *66*, 1039.
534. V. Vimberg; L. Xiong; M. Bailey; T. Tenson; A. Mankin, *Mol. Microbiol.* **2004**, *54*, 376.
535. B. Weisblum, *Antimicrob. Agents Chemother.* **1995**, *39*, 577.
536. B. Weisblum, *Antimicrob. Agents Chemother.* **1995**, *39*, 797.
537. A. R. Kwon; Y. H. Min; E. J. Yoon; J. A. Kim; M. J. Shim; E. C. Choi, *Arch. Pharm. Res.* **2006**, *29*, 1154.
538. N. Vazquez-Laslop; C. Thum; A. S. Mankin, *Mol. Cell* **2008**, *30*, 190.
539. D. A. Low; J. Casadesus, *Curr. Opin. Microbiol.* **2008**, *11*, 106.
540. M. C. Roberts, *FEMS Microbiol. Lett.* **2008**, *282*, 147.
541. D. E. Bussiere; S. W. Muchmore; C. G. Dealwis; G. Schluckebier; V. L. Nienaber; R. P. Edalji; K. A. Walter; U. S. Ladror; T. F. Holzman; C. Abad-Zapatero, *Biochemistry* **1998**, *37*, 7103.
542. G. Schluckebier; P. Zhong; K. D. Stewart; T. J. Kavanaugh; C. Abad-Zapatero, *J. Mol. Biol.* **1999**, *289*, 277.
543. H. C. O'Farrell; J. N. Scarsdale; J. P. Rife, *J. Mol. Biol.* **2004**, *339*, 337.
544. G. Maravic; M. Feder; S. Pongor; M. Flogel; J. M. Bujnicki, *J. Mol. Biol.* **2003**, *332*, 99.
545. W. S. Champney; H. S. Chittum; C. L. Tober, *Curr. Microbiol.* **2003**, *46*, 453.
546. G. Maravic; J. M. Bujnicki; M. Feder; S. Pongor; M. Flogel, *Nucleic Acids Res.* **2003**, *31*, 4941.
547. N. Wolter; A. M. Smith; D. J. Farrell; J. B. Northwood; S. Douthwaite; K. P. Klugman, *Antimicrob. Agents Chemother.* **2008**, *52*, 435.
548. I. Lebars; C. Husson; S. Yoshizawa; S. Douthwaite; D. Fourmy, *J. Mol. Biol.* **2007**, *372*, 525.
549. I. Pokkunuri; W. S. Champney, *RNA Biol.* **2007**, *4*, 147.
550. S. Douthwaite; L. Jakobsen; S. Yoshizawa; D. Fourmy, *J. Mol. Biol.* **2008**, *378*, 969.
551. C. T. Madsen; L. Jakobsen; K. Buriankova; F. Doucet-Populaire; J. L. Pernodet; S. Douthwaite, *J. Biol. Chem.* **2005**, *280*, 38942.
552. M. Feder; E. Purta; L. Koscinski; S. Cubrilo; G. Maravic Vlahovicek; J. M. Bujnicki, *ChemMedChem* **2008**, *3*, 316.
553. M. T. Burger; C. Hiebert; M. Seid; D. T. Chu; L. Barker; M. Langhorne; R. Shawar; J. Kidney; M. C. Desai; J. J. Plattner, *Bioorg. Med. Chem.* **2006**, *14*, 5592.
554. X. Lin; A. C. Rico; D. T. Chu; G. L. Carroll; L. Barker; R. Shawar; M. C. Desai; J. J. Plattner, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4692.
555. L. E. Bermudez; N. Motamedi; C. Chee; G. Baimukanova; P. Kolonoski; C. Inderlied; P. Aralar; G. Wang; L. T. Phan; L. S. Young, *Antimicrob. Agents Chemother.* **2007**, *51*, 1666.
556. J. Dreier; E. Amantea; L. Kellenberger; M. G. Page, *Antimicrob. Agents Chemother.* **2007**, *51*, 4361.

557. B. Zhu; B. A. Marinelli; D. Abbanat; B. D. Foleno; K. Bush; M. J. Macielag, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 3900.
558. T. Kaneko; W. McMillen; M. K. Lynch, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5013.
559. T. Kaneko; K. Romero; B. Li; R. Buzon, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5049.
560. I. Miroshnyk; S. Mirza; P. M. Zorky; J. Heinamaki; J. Yli-Kauhaluoma; J. Yliruusi, *Bioorg. Med. Chem.* **2008**, *16*, 232.
561. T. Furuuchi; T. Miura; K. I. Kurihara; T. Yoshida; T. Watanabe; K. Ajito, *Bioorg. Med. Chem.* **2008**, *16*, 4401.
562. V. T. Andriole, *Clin. Infect. Dis.* **2005**, *41* (Suppl. 2), S113.
563. G. A. Jacoby, *Clin. Infect. Dis.* **2005**, *41* (Suppl. 2), S120.
564. L. A. Mitscher, *Chem. Rev.* **2005**, *105*, 559.
565. B. Périchon; P. Courvalin; M. Galimand, *Antimicrob. Agents Chemother.* **2007**, *51*, 2464.
566. Q. C. Truong-Bolduc; D. C. Hooper, *J. Bacteriol.* **2007**, *189*, 2996.
567. K. Yamane; J. Wachino; S. Suzuki; K. Kimura; N. Shibata; H. Kato; K. Shibayama; T. Konda; Y. Arakawa, *Antimicrob. Agents Chemother.* **2007**, *51*, 3354.
568. S. Begic; E. A. Worobec, *Microbiology* **2008**, *154*, 454.
569. P. A. Lambert, *Adv. Drug Deliver. Rev.* **2005**, *57*, 1471.
570. A. K. Turner; S. Nair; J. Wain, *J. Antimicrob. Chemother.* **2006**, *58*, 733.
571. L. F. Hu; J. B. Li; Y. Ye; X. Li, *J. Microbiol.* **2007**, *45*, 168.
572. S. N. Richter; G. Giaretta; V. Comuzzi; E. Leo; L. A. Mitchenall; L. M. Fisher; A. Maxwell; M. Palumbo, *Nucleic Acids Res.* **2007**, *35*, 6075.
573. M. Ip; S. S. Chau; F. Chi; J. Tang; P. K. Chan, *Antimicrob. Agents Chemother.* **2007**, *51*, 2690.
574. D. E. Rozen; L. McGee; B. R. Levin; K. P. Klugman, *Antimicrob. Agents Chemother.* **2007**, *51*, 412.
575. J. H. Tran; G. A. Jacoby, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5638.
576. G. A. Jacoby; N. Chow; K. B. Waites, *Antimicrob. Agents Chemother.* **2003**, *47*, 559.
577. E. Giraud; S. Baucheron; A. Cloeckaert, *Microbes Infect.* **2006**, *8*, 1937.
578. M. J. Ellington; N. Woodford, *J. Antimicrob. Chemother.* **2006**, *57*, 1026.
579. A. Robicsek; G. A. Jacoby; D. C. Hooper, *Lancet Infect. Dis.* **2006**, *6*, 629.
580. L. Poirel; V. Cattoir; P. Nordmann, *Clin. Microbiol. Infect.* **2008**, *14*, 295.
581. P. Courvalin, *Arch. Microbiol.* **2008**, *189*, 289.
582. S. S. Hegde; M. W. Vetting; S. L. Roderick; L. A. Mitchenall; A. Maxwell; H. E. Takiff; J. S. Blanchard, *Science* **2005**, *308*, 1480.
583. M. W. Vetting; S. S. Hegde; J. E. Fajardo; A. Fiser; S. L. Roderick; H. E. Takiff; J. S. Blanchard, *Biochemistry* **2006**, *45*, 1.
584. M. W. Vetting; S. S. Hegde; K. Z. Hazleton; J. S. Blanchard, *Protein Sci.* **2007**, *16*, 755.
585. J. H. Tran; G. A. Jacoby; D. C. Hooper, *Antimicrob. Agents Chemother.* **2005**, *49*, 118.
586. J. H. Tran; G. A. Jacoby; D. C. Hooper, *Antimicrob. Agents Chemother.* **2005**, *49*, 305.
587. K. P. Klugman; B. R. Levin, *Nat. Med.* **2006**, *12*, 19.
588. A. Robicsek; J. Strahilevitz; G. A. Jacoby; M. Macielag; D. Abbanat; C. Hye Park; K. Bush; D. C. Hooper, *Nat. Med.* **2006**, *12*, 83.
589. M. W. Vetting; C. H. Park; S. S. Hegde; G. A. Jacoby; J. S. Blanchard, *Biochemistry* **2008**, *47*, 9825.
590. S. K. Henderson-Begg; D. M. Livermore; L. M. Hall, *J. Antimicrob. Chemother.* **2006**, *57*, 849.
591. W. L. Kelley, *Mol. Microbiol.* **2006**, *62*, 1228.
592. M. Prudhomme; L. Attaiech; G. Sanchez; B. Martin; J. P. Claverys, *Science* **2006**, *313*, 89.
593. E. Lopez; M. Elez; I. Matic; J. Blazquez, *Mol. Microbiol.* **2007**, *64*, 83.
594. J. Overhage; M. Bains; M. D. Brazas; R. E. Hancock, *J. Bacteriol.* **2008**, *190*, 2671.
595. D. J. Dwyer; M. A. Kohanski; B. Hayete; J. J. Collins, *Mol. Syst. Biol.* **2007**, *3*, 91.
596. M. A. Kohanski; D. J. Dwyer; B. Hayete; C. A. Lawrence; J. J. Collins, *Cell* **2007**, *130*, 797.
597. R. E. Hancock, *Mol. Syst. Biol.* **2007**, *3*, 142.
598. G. D. Wright, *Cell* **2007**, *130*, 781.
599. D. J. Hassett; J. A. Imlay, *ACS Chem. Biol.* **2007**, *2*, 708.

Biographical Sketches

Jed F. Fisher was born and bred in Yonkers, New York. His first research experience was as an undergraduate in the laboratory of Bill Fowler at S.U.N.Y. Stony Brook, examining the

Diels–Alder chemistry of dihydropyridines and the Grob fragmentations of the resulting cycloadducts. He arrived in the fall of 1972 at M.I.T. at the same time as a new faculty member in the Department of Chemistry, Chris Walsh. His 4 years in the Walsh group focused on the use of riboflavin analogues to study flavin coenzyme mechanisms. He remained in Massachusetts as a postdoctoral fellow with Jeremy Knowles at Harvard, where he deciphered the acyl-enzyme mechanism of the β-lactamases, particularly with regard to the slow hydrolysis of cefoxitin and inactivation by clavulanate. Following a faculty position at the University of Minnesota (redox chemistry of the anthracyclines and mito-mycins, inhibitor of protein–protein association), and nearly two decades as a medicinal chemist (Upjohn), he joined the University of Notre Dame in 2004. He enjoys the Lake Michigan shoreline on weekends, and reads, writes, and teaches chemistry during the week.

Shahriar Mobashery received his undergraduate (1981) and doctoral (1985) degrees from the University of Southern California and University of Chicago, respectively. Subsequent to postdoctoral research (1986–88) at the Rockefeller University, he joined the faculty at Wayne State University in 1989. Mobashery and his research group relocated to the University of Notre Dame in 2003, where he holds the position of Navari Family Professor of Life Sciences in the Department of Chemistry and Biochemistry. He was awarded the 2008 Astellas USA Foundation Award for his research on the mechanisms of antibiotic resistance, antibiotic design, and the structure of the bacterial cell wall.

8.14 Copper Metalloenzymes

Albrecht Messerschmidt, Max-Planck-Institute of Biochemistry, Martinsried, Germany

8.14.1 Introduction

8.14.1.1 General Remarks on Copper Proteins

A group of proteins contain copper in sites in a permanently coordinated state. These sites may simply function as depository for copper, in molecules participating in copper trafficking, as electron transfer center or as substantial part of the active site of a copper enzyme. Research on copper proteins has been summarized in a three volume book by Lontie,[1] in volume 2 of the *Handbook of Copper Proteins*,[2] in the *Handbook on Metalloproteins*,[3] in a book edited by Valentine and Gralla,[4] and in several review articles (see, e.g., Abolmaali *et al.*[5] or Messerschmidt[6]).

Copper is redox active and in most cases is involved directly in the catalytic cycle. The copper-dependent catalysis proceeds via redox change of the copper ions involved independently of the concrete mechanism and the number of participating copper ions. During the redox change, electrons are transferred between the redox partners, and the redox potentials of the redox couples involved determine the direction of the electron flow. In biological systems copper has been found in the Cu(I) and Cu(II) oxidation states only. Copper proteins occur in all three biological kingdoms – archaea, bacteria, and eukarya. The occurrence and functions of copper proteins suggest that copper gained biological importance only after the oxygenation of the atmosphere.

- Copper proteins have not been detected in anaerobically living archaea.
- They function as oxygen transport or electron transfer proteins in the range of very high redox potentials.
- In copper trafficking they act as depositary, transporters, and chaperones (see, e.g., Arnesano and Banci[7]).
- Copper thioneins have two functions, namely, copper storage[8,9] and detoxification.[10]
- In all the reactions catalyzed by copper enzymes except for copper-dependent nitrogen oxide reductases, dioxygen either functions as a substrate or electron receptor.
- While the extracellular milieu is oxidizing, cytosol has reducing properties. Except for cellular superoxide dismutase and proteins involved in copper trafficking, all known copper proteins and enzymes are found in the extracellular space or in organelles and vesicles.

The participation of copper electron transfer proteins in photosynthesis appears strange, as photosynthesis originated before the oxygenation of the atmosphere and the copper proteins only thereafter. However, in early photosynthesis the function of the copper electron transfer proteins was performed by cytochromes that subsequently have been replaced by the respective copper proteins. The reason for the late use of copper during the evolution lies in the bioavailability of copper. Copper was not utilizable by organisms during the first 2 billion years after the origin of life. The reducing atmosphere and the relatively high H_2S concentration $(10^{-3}\,mol\,l^{-1})$ of the seas maintained copper in its singly oxidized form (Cu^+) and caused its precipitation as copper sulfide (Cu_2S).[11] As a result, uptake and incorporation of copper was initially denied to the organisms. In contrast, iron was present as Fe^{2+} under these conditions and was used in proteins in this form. After the oxygenation of the atmosphere and the seas, the ion equilibria of the transition metals shifted to higher oxidation numbers. Cu^+ became oxidized to soluble Cu^{2+} and gained biological relevance. In this context, it is interesting to note that iron was now in an oxidized form, which exists in water as poorly soluble $Fe(OH)_3$. For iron, new mechanisms for enrichment and storage had to be found to allow the iron proteins already developed to continue to function (da Silva and Williams[12]). As this chapter is in principle dedicated to copper enzymes, a subgroup of copper proteins, two other subgroups, proteins from copper trafficking and electron transfer proteins, will be treated very briefly in this introductory section because of their general interest in the context of copper enzymes.

8.14.1.2 Copper Trafficking

Copper trafficking has been recently reviewed.[7,13,14] The need for special copper trafficking systems is due to the redox properties of copper, which are also responsible for its cellular toxicity. Indeed, free copper ions catalyze Fenton reactions, which produce highly reactive oxygen species (ROS) that may damage cell membranes, proteins, and nucleic acids.[15]

To supply copper to proteins and enzymes, avoiding damage caused by free copper ions, cells have developed mechanisms for copper transport and homeostasis, which maintain the cellular concentration of copper within narrow limits and prevent the occurrence of free copper ions in the cytosol,[8] and also make the needed copper ions available in concentrations suitable for copper metalation. The tight control exerted by these cell machineries regulates copper levels via uptake and efflux, mediated by high-affinity membrane transporters, as well as guarantees copper delivery to the relevant target proteins and compartments.[16,17] The latter task is done by soluble copper chaperones, which provide an efficient copper distribution to specific cellular pathways and of copper incorporation into designated copper enzymes.[18,19]

The class of copper transporters encompasses proteins with very different architectures and functional mechanisms. Nonetheless, some features are conserved among different subclasses. Membrane copper transporters contain intramembrane copper-binding site(s) and soluble copper-binding domains, while soluble
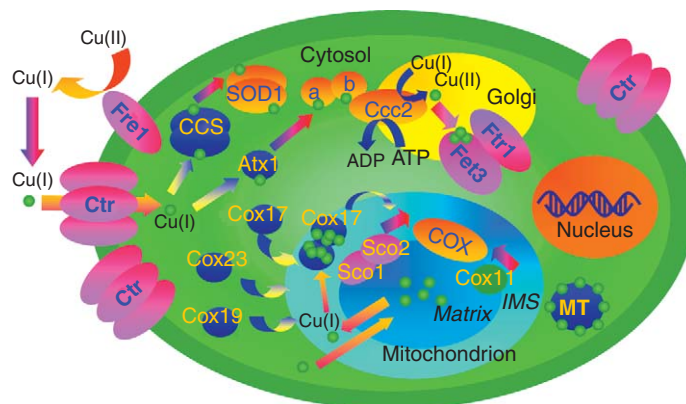
Figure 1 Copper trafficking pathways in yeast. Atx, antioxidant; Ccc1, cross-complements Ca^{2+}-sensitive phenotype of csg1; CCS, copper chaperone for SOD; COX, cytochrome c oxidase; Ctr, copper transporter; Fet, ferrous transport; Fre, FerriReductase-encoding; Ftr, Fe transporter; IMS, intermembrane space; MT, metallothienoin; Sco, suppressor of cytochrome c oxidase deficiency; SOD, superoxide dismutase. Reproduced from F. Arnesano; L. Banci, In *Handbook of Metalloproteins*; A. Messerschmidt, Ed.; John Wiley & Sons: Chichester, 2007; Vol. 4, pp 1–21, with permission from John Wiley & Sons.

copper chaperones are characterized by exposed metal sites, and thermodynamic stability and kinetic lability of copper binding, which ensure tight binding as well as fast copper transfer.[20] A classification of copper transporters and chaperones can be made on the basis of the functional pathway in which they are involved. So far, three well-characterized trafficking pathways (**Figure 1**) have been studied: (1) copper transport into the Golgi/thylakoid compartment and incorporation into multicopper oxidases/plastocyanin and other copper enzymes as well as excess copper excretion from the cell, (2) copper incorporation into Cu,Zn-superoxide dismutase (Cu,Zn-SOD) in the cytosol and mitochondria, and (3) copper delivery to mitochondria/periplasm and incorporation into specific subunits of cytochrome c oxidase (COX). The components of the different trafficking pathways are explained in the caption of **Figure 1**. The Ccc2 component is a copper-transporting ATPase delivering copper to the Golgi/thylakoid compartment in eukaryotes and cyanobacteria, respectively, which has homologues in numerous bacteria, reflecting a highly conserved mechanism of copper transport, adapted among different phyla and cellular organizations.[21] For more details on copper trafficking, the reader is referred to the cited review articles.[7,13,14]

Another interesting protein is the amyloid precursor protein (APP), which is a transmembrane glycoprotein implicated in the pathogenesis of Alzheimer's disease (AD).[22] APP is a multidomain protein with metal-binding sites critical to its function. There are two copper-binding domains, one lies in the N-terminus, adjacent to the zinc-binding domain, and the other is in the amyloid-beta (Aβ) domain. Aβ is derived via a series of protease cleavages of APP by the secretases and is the main constituent of the amyloid plaques that are a key hallmark of AD. The physiological role of APP is as yet unknown. It can reduce Cu^{2+} to Cu^+ and the physiological and three-dimensional structure suggests a role as a copper chaperone. The binding of Cu to Aβ is toxic in neuronal cultures and this may contribute to the oxidative stress that is commonly observed in AD.

8.14.1.3 Copper in Medicine

Genetic abnormalities involving copper in humans are responsible for a number of diseases with Wilson's disease and Menke's disease as the most important ones. Wilsons's disease is an autosomal genetic disorder and is caused by disabling mutations in both copies of the ATP7B gene.[23] This gene codes for a copper-transporting P-type ATPase, which functions in a pathway in the liver for biliary excretion of excess copper. With absent or reduced function, copper accumulates and causes progressive damage in the liver and often in the brain.[24] The therapy of Wilson's disease continues to advance with the addition of zinc to the food.[25] Zinc's mechanism of action involves induction of intestinal cell metallothionein, which binds food copper and copper in gastro-intestinal secretions and prevents its absorption.

Menke's disease is an X-linked inherited disorder and is caused by a mutation in the ATP7A gene.[26] The syndrome that results is one of severe prenatal and postnatal copper deficiency in affected males, leading to brain damage and mental retardation. The ATP7A gene also codes for a membrane-bound copper-transporting ATPase. The expression profiles and functions are different, with ATP7B expressed primarily in the liver with functions as mentioned above, whereas ATP7A is much more ubiquitously expressed and has important functions in several organs. ATP7A is crucial in causing copper efflux from cells. Failure of function of ATP7A in the intestine leads to a failure of copper efflux from intestinal cells, accumulation of excess copper in the intestine, a failure of copper absorption into the blood, and generalized copper deficiency. Failure of function of ATP7A in the blood–brain barrier leads to a failure of copper efflux from cells of this barrier, accumulation of copper in these cells, and a failure of copper uptake in the brain, even if circulating copper levels are normalized by parenteral copper therapy.

Aceruloplasminemia is an autosomal recessive disease, which is caused by mutations in the ceruloplasmin (Cp) gene and results in a total absence of Cp in the blood. It is an iron accumulation disorder causing clinical problems in the brain and liver.[27] Cp is a ferroxidase, necessary to convert Fe^{2+} to Fe^{3+} so that the iron can be bound to transferrin and mobilized from cells.

The role of copper in Alzheimer's disease has already been discussed in the previous section.

For a more detailed discussion of copper in medicine see the review of Brewer.[28]

8.14.1.4   Different Copper Site Geometries in Copper Proteins

Copper is bound to the copper proteins in different geometries according to the requirements of their function. In aqueous solution and in the presence of suitable ligands, copper forms complexes in different oxidation numbers that differ both in the geometry and in the kind and number of ligands. According to the HSAB (hard soft acid base) concept of Pearson,[29] the soft Cu^+ in proteins prefers the sulfur atoms or ions of the amino acids Cys, Cys⁻, and Met. The harder Cu^{2+} can be coordinated by oxygen and nitrogen atoms and ions of the harder amino acids such as Tyr, Tyr⁻, Thr, His, OH⁻, and H_2O. When no steric hindrance is present the preferred coordination of Cu^+ is tetrahedral, but triginal planar and linear coordination are also possible. The ligands of Cu^{2+} arrange in a square planar configuration. Other possibilities are distorted tetrahedral, square pyramidal, trigonal bipyramidal, and distorted octahedral.[30] In the copper-binding sites of proteins, both ligands and their conformations deviate considerably from the preferred coordination of the corresponding oxidation state. Copper complexes are most stable if the favored ligands are in the preferred coordination geometry around the central atom. Changes in this arrangement lead to a destabilization of the corresponding oxidation state. Copper centers in cuproproteins except for a row of components of copper trafficking must be redox active to be suited to their function; that is, that the copper center should be able to change easily between the cupric (Cu^{2+}) and cuprous (Cu^+) oxidation states. This has been realized by using ligands that favor the cuprous state and other ligands that prefer the cupric state. As a second solution, arrangements of the ligands have been found in such geometries that are intermediate between the demands for Cu^+ and Cu^{2+}. In this case, the relatively rigid protein backbone provides a preformed copper-binding site.[31] The nature of ligands, the coordination geometry, and the surrounding protein matrix together determine the redox potential of the copper site. Thus a distortion of the square planar geometry preferred by Cu^{2+} to tetrahedral facilitates the reduction of Cu^{2+} to Cu^+, that is, the redox potential is increased. Therefore, soft ligands in the coordination sphere also raise the Cu^+/Cu^{2+} redox potential.[32] The Cu^+/Cu^{2+} redox potential in aqueous solution is $E_0 = +153$ mV. In contrast, the range of observed redox potentials in copper proteins is from $+183$ mV for halocyanin[33] to $+785$ mV for the type-1 copper center in fungal laccase.[34] Seven different types of copper sites have so far been characterized in copper proteins. They will be described below.
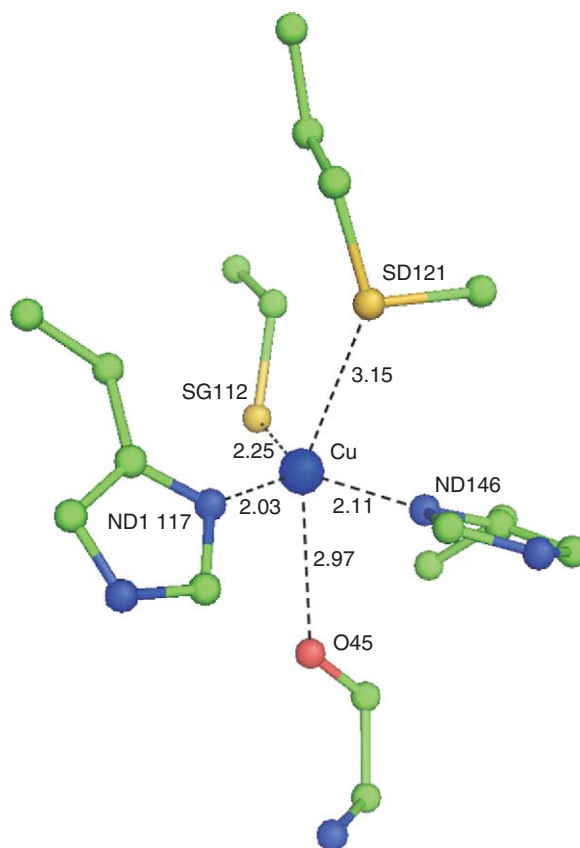
8.14.1.4.1   Type-1 copper sites

Small blue copper proteins contain a mononuclear copper site with three characteristic properties: (1) an intense blue color at approximately 600 nm, with absorption coefficients of 2000–6000 $mol^{-1} l cm^{-1}$, due to a S(Cys) → Cu(II) charge transfer and (2) an unusually narrow hyperfine coupling ($A_{||}$ values of 0.0035–0.0063 cm^{-1}) in the electron paramagnetic resonance (EPR) spectrum of the Cu(II) protein; high redox potentials (183–680 mV) as compared to the aqua Cu(II/I) couple. This copper site has been classified as

type-1 copper site. The small blue copper proteins function as electron transfer proteins in such fundamental processes as photosynthesis and respiration. Research on them has been summarized in several reviews in the past (Sykes,[35] Adman,[36] Canters and Gilardi,[37] Messerschmidt[38]). Type-1 copper sites are not only present in the small blue copper proteins but also in blue multicopper oxidases and copper-containing nitrite reductase (see Messerschmidt[39]). Small blue copper proteins (cupredoxins) fold into a single domain mainly consisting of a β-sandwich or β-barrel. This β-sandwich may be built up by 6–13 β-strands. The arrangement of the β-strands is basically antiparallel showing the Greek-key motif. A typical fold represented by azurin is depicted in **Figure 2**. In this case the β-sandwich consists of eight β-strands. Subunits II in COX[40,41] exhibits in its membrane-exposed domain a cupredoxin fold. But this domain contains a binuclear copper site denoted as Cu_A site discussed in (Section 8.14.1.4.5). The blue multicopper oxidases ascorbate oxidase, laccase, and ceruloplasmin are built up of three (ascorbate oxidase, laccase) and six (ceruloplasmin) domains with the cupredoxin fold (see, e.g., Messerschmidt[39]). The mononuclear blue copper site is located in the C-terminal domain of ascorbate oxidase and laccase. Ceruloplasmin has three mononuclear coppers bound to domains 2, 4, and 6.[42] A copper-containing nitrite reductase from *Achromobacter cycloclastes* or different *Alcaligenes* species consists of two domains with cupredoxin fold, which arrange to the enzymatic active trimer (six domains).[43] The mononuclear site is found in the N-terminal domain. The cupredoxin and their multidomain relatives form a large evolutionary family. This family also comprises proteins that have lost their copper-binding capability during evolution (see, e.g., Messerschmidt[44]). The type-1 copper center is located close to the surface of the protein (**Figure 2**). In native type-1 proteins the copper has four protein side chains as ligands and in some cases (e.g., azurin) a weak main chain carbonyl oxygen as fifth ligand (see **Figure 3**). The four canonical type-1 copper ligands are His, Cys, His, and Met arranged in this sequence on the polypeptide chain. The Cys, His, Met ligands come from a loop between the two C-terminal strands of the β-sandwich. The type-1 copper sites can be subdivided into T1 trigonal and T1 distorted tetrahedral based on the optical, EPR, and RR spectroscopy[45–47] as well as on X-ray crystallography.[36] The most abundant geometry of the type-1 sites is T1 distorted tetrahedral as found in plastocyanin, ascorbate oxidase, amicyanin, pseudoazurin, and cucumber basic protein. This unusual copper coordination is a compromise between the preferred tetrahedral Cu(I) and trigonal Cu(II) coordination. It reduces the reorganization energy between both redox states, which helps to speed up electron transfer.[48] The crystal structures of wild-type azurin at two pH values in both redox states have been determined.[49,50] There are little changes in the copper site geometry. On reduction, the bond distances are

Figure 2   Ribbon diagram of azurin from *Pseudomonas aeruginosa*, a typical cupredoxin; type-1 copper site and disulfide bridge are included (PDB-code: 4AZU); prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

Figure 3 Type-1 copper site in wild-type azurin from *Pseudomonas aeruginosa* (PDB-code: 4AZU); prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

slightly increased by about 0.05–0.1 Å $(1\,\text{Å} = 10^{-10}\,\text{m})$ as also observed in reduced poplar plastocyanin at pH 7.8[51] and pseudoazurin from *Alcaligenes faecalis* at pH 7.8.[52] This means that both T1 trigonal and T1 distorted tetrahedral copper sites are well designed for rapid electron transfer between different redox partners.

An interesting cupredoxin is nitrosocyanin (NC) from the ammonia-oxidizing bacterium *Nitrosomonas europaea*, which has a red color. The X-ray structure of NC[53] shows that the coordination of the copper ion is distinct from that of blue copper centers. In particular, the red copper center has a higher coordination number and lacks the long Cu–S(Met) and short Cu–S(Cys) bond distances characteristic of blue copper. Moreover, the red copper center is square pyramidal whereas the blue copper is typically distorted tetrahedral.

Stellacyanins, a subclass of phytocyanins occurring exclusively in plants, have a glutamine as axial ligand instead of methionine. The coordination of the copper can be described as trigonal pyramidal with a short bond distance of 2.2 Å from the copper to the OE1 atom of the axial glutamine ligand.[54] This is the shortest axial bond distance observed in cupredoxins so far.

The C-terminal loop in cupredoxins holding three of the copper ligands vary in length for the different cupredoxins. Cupredoxin chimeras have been generated where the C-terminal copper-binding loop of another cupredoxin has been grafted into the protein scaffold of another cupredoxin (for a recent review see Dennison[55]). These studies demonstrate the importance of loop–scaffold interactions for metal sites in cupredoxins and generally in proteins.

8.14.1.4.2 Type-2 copper sites

Type-2, or normal, Cu^{2+} has undetectable absorption and the EPR line shape of the low-molecular-mass copper complexes $(A_{||} > 0.0140\,\text{cm}^{-1})$. Type-2 copper centers are present in copper-containing oxidases
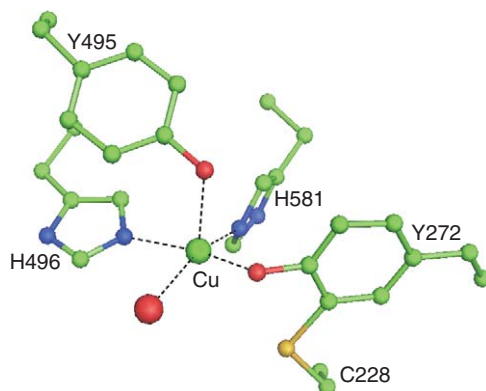
Figure 4 Type-2 copper site in galactose oxidase from *Dactylium dendroides* (PDB-code: 1GOG); prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

(not in blue oxidases and COX), oxygenases, nitrite reductase, and Cu,Zn-SOD. The type-2 copper centers are involved in the direct activation of dioxygen or additionally responsible for the formation of the internal cofactor as in amine oxidase and galactose oxidase. As in the low-molecular-mass copper complexes the type-2 Cu^{2+} sites prefer quadratic planar or tetragonal pyramidal coordinations. The ligand types vary. However, a common feature is the fact that in the active form of the enzyme one of the coordination sites of the copper is empty for the binding of the dioxygen. Ligands of type-2 copper are histidine, tyrosine (often modified), methionine, and cysteine with decreasing abundance in the sequence of citation. The active site structure of galactose oxidase as revealed by X-ray crystallography[56] is illustrated in **Figure 4**, showing the coordination of the copper ion by two histidine residues, a simple tyrosinate and a covalently modified tyrosine cross-linked to a cysteinyl residue to form a new dimeric amino acid (cysteinyl-tyrosine). Ligands in amine oxidase and lysyl oxidase are each three histidines and a modified tyrosine.[57] In diamine oxidase two of the histidine residues seem to be replaced by cysteines.[58] According to Blackburn *et al.*[59] the oxidized enzyme of dopamine-β-monooxygenase contains two tetragonal Cu^{2+} centers coordinated to no more than three histidine ligands per Cu. They propose that the Cu^{2+} sites are inequivalent, with a $Cu_H(His)_3(H_2O)\ldots Cu_M(His)_2X(H_2O)$ type configuration. The identity of X as Met has been concluded from the related structure of peptidylglycine α-hydroxylating monooxygenase (PHM).[60] Cu,Zn-SOD shows a type-2 copper center with four histidine ligands whose coordinating nitrogen atoms are arranged in a quadratic planar fashion around the copper. One of the histidines is also linked to the zinc atom, the second metal within the active center.[61,62]

8.14.1.4.3 Type-3 copper sites

Type-3 copper centers are binuclear sites whose copper ions are antiferro-magnetically coupled in the oxidized state[63,64] and exhibit very similar spectroscopic properties. They occur in the oxygen transport protein hemocyanin, in the monooxygenase tyrosinase, and in catechol oxidase. The copper site geometries of hemocyanins have been determined by X-ray crystallography.[65,66] The copper ions are each trigonal-planar coordinated by three histidine ligands in the deoxy form and the copper–copper distance is 4.6 Å.[66] In oxyhemocyanin the dioxygen is bound to the dinuclear copper site as peroxide in a μ–η^2:η^2 fashion[67] (**Figure 5**) and the copper–copper distance is 3.6 Å. The geometry of the type-3 copper site in tyrosinase is very similar to that of hemocyanin due to the similar spectroscopic features and has now been derived from the X-ray structure of a bacterial tyrosinase.[68]

8.14.1.4.4 Trinuclear copper centers

Besides the blue type-1 copper sites, the blue oxidases contain a trinuclear copper center, which is located between the N- and C-terminal domains. It can be described spectroscopically as a coupled type-2/type-3 copper center.[69] The atomic structure of the trinuclear copper site for oxidized ascorbate oxidase as derived from X-ray crystallography[70] is displayed in **Figure 6**. The trinuclear cluster has eight histidine ligands symmetrically supplied from
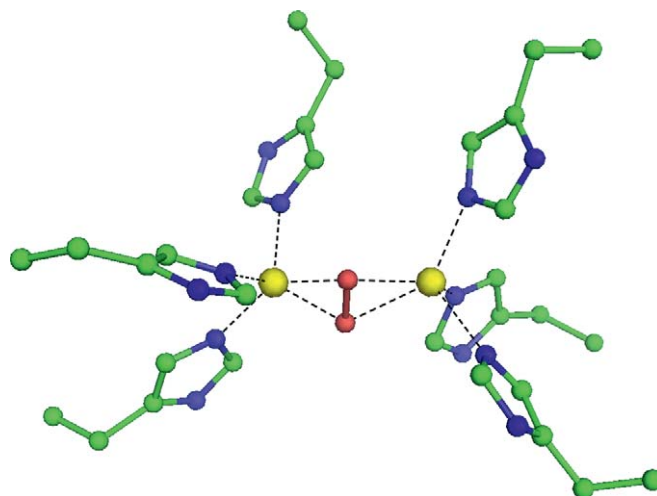
Figure 5 Type-3 copper site in oxyhemocyanin from *Limulus polyphemus* (PDB-code: 1OXY); prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).
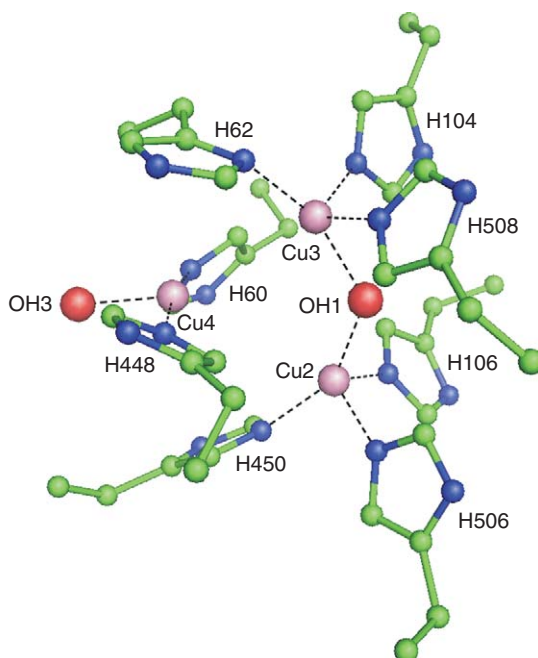
Figure 6 The trinuclear copper site as determined for oxidized ascorbate oxidase (PDB-code: 1AOZ);[70] prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

the N- and C-terminal domains. It may be subdivided into a pair of copper atoms with six histidine ligands whose coordinating N-atoms are arranged trigonal-prismatic in contrast to hemocyanin where the N-atoms form a trigonal antiprism. The pair is the putative type-3 copper. The remaining copper has two histidine ligands and is the putative spectroscopic type-2 copper. Two oxygens are bound to the trinuclear species; as OH^- or O^{2-} and bridging the type-3 copper pair, and as OH^- or H_2O bound to the putative type-2 copper trans to the copper pair. An oxygen ligand in the center of the copper ions could not be detected. The average copper–copper distances in the trinuclear copper site of ascorbate oxidase is 3.74 Å. The structure of the trinuclear copper site in human ceruloplasmin as determined by X-ray crystallography[42] is similar to that in ascorbate oxidase. The same is valid

for laccases, where X-ray structures of a number of fungal laccases have been recently determined.[71–75] These trinuclear copper centers are the active sites for the dioxygen reduction.

8.14.1.4.5   Cu_A copper centers

The purple Cu_A copper site is one of four metal sites in COX and functions as primary electron acceptor from cytochrome *c* (see, e.g., Ramirez *et al.*[48] and Gennis and Ferguson-Miller[76]). N_2O reductase, another cupredoxin-domain-containing enzyme, also contains a Cu_A site as one of its metal sites and it has been characterized both in COX and N_2O reductase as a mixed valence, Cu(1.5)–Cu(1.5) redox state, due to its seven-line EPR spectrum and other similar spectroscopic properties.[77,78] A Cu_A site has been engineered into purple Cyo A (membrane-exposed domain from quinol oxidase).[79] The structure of the Cu_A site in COX from *Rhodobacter sphaeroides* as derived from the crystal structure[80] is shown in **Figure 7**. Two thiolate groups of different cysteines are bridging the two coppers. Two histidines act as terminal ligands to each copper. The arrangement of these four ligands is symmetrical with respect to the copper pair. Asymmetry is introduced by the coordination of a methionine and main-chain carbonyl oxygen, which each complete the distorted tetrahedral coordination of the individual coppers. All four Cu–S(Cys) bonds are approximately 2.2 Å long. The copper–copper distance is 2.6 Å. The Cu_A site in engineered Cyo A resembles this of COX.[79] The copper–copper distances in the X-ray structures of COX from *Paracoccus denitrificans*[40] and bovine heart[41] are 2.6 and 2.7 Å, respectively. The structure of the Cu_A center with its [2Cu–2S] center resembles that of a [2Fe–2S] iron–sulfur center. It is remarkable that the distorted tetrahedral coordination of each copper is conserved and that the Cu_2S_2Im_2 cluster allows for complete electron delocalization, thereby ensuring that the small reorganization energy is spread over both copper ions.[48]

8.14.1.4.6   Cu_B copper centers

COXs contain a second copper center called Cu_B. Cu_B is always part of a heme–Cu_B pair whose structure is now known from the crystal structures of bacterial (e.g., from *P. denitrificans* and *R. sphaeroides*) and bovine heart
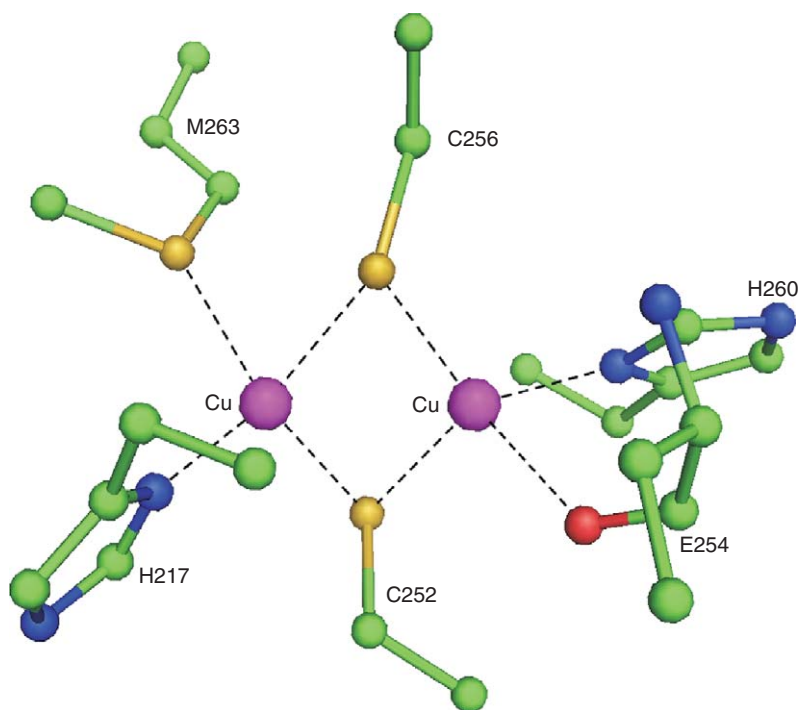
Figure 7   Cu_A copper site as determined for COX from *Rhodobacter sphaeroides* (PDB-code: 1M56); prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).
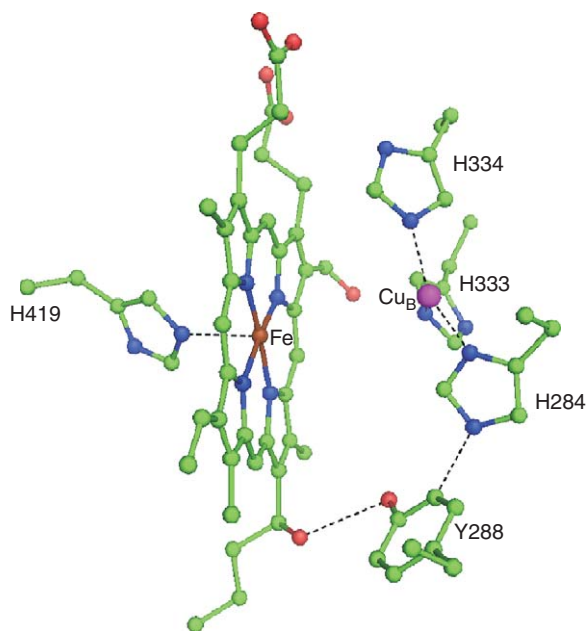
Figure 8 Cu$_B$–heme a_3 pair in COX from *Rhodobacter sphaeroides* (PDB-code: 1M56); prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

COXs (**Figure 8**). The metal ions of both redox-active centers are antiferro-magnetically coupled. In oxidized bovine heart COX as well as in COX from *R. sphaeroides*, Cu$_B$ is coordinated by three histidines in a trigonal-planar arrangement.[81] The Cu$_B$–Fe$_{heme\ a3}$ distance is 4.5 Å. A ligand directly bridging the Fe$_{a3}$ and Cu$_B$ was not found. The Cu$_B$–heme pairs are the binding and reduction sites for the dioxygen and besides the trinuclear copper clusters of the blue oxidases the only structures that catalyze the four-electron reduction of dioxygen to water.

8.14.1.4.7   Cu$_Z$ copper centers

On the basis of the analysis of EPR, optical, and MCD data, the catalytic Cu$_Z$ center of N$_2$OR was proposed to be a second binuclear center with cysteine coordination.[82,83] When *nosZ* DNA sequences revealed that there were an insufficient number of conserved Cys residues to bind two thiolate-bridged centers, coordination by a set of conserved histidine residues was suggested.[84] This suggestion was subsequently confirmed by the crystal structures of N$_2$OR. What was completely unexpected was the tetranuclear nature of the Cu$_Z$ site,[85,86] which had remained undiscovered despite expensive spectroscopic study. The recognition of the tetranuclear structure of the Cu$_Z$ center provided a framework for the interpretation of the spectroscopic data.

The crystal structures of N$_2$OR from three different bacteria (*Pseudomonas nautica*,[86] *P. denitrificans*,[87] and *A. cycloclastes*[88]) determined so far show that the Cu$_Z$ center is located nearly in the middle of the central channel of the β-propeller domain and reveal the novel nature of the Cu$_Z$ site, in which four Cu ions were coordinated by seven histidine residues (**Figure 9**). In addition, an inorganic S bridges the Cu atoms forming the novel μ$_4$-sulfide-bridged tetranuclear Cu cluster. **Figure 9** shows the Cu$_Z$ center from *A. cycloclastes*[88] with the inhibitor I$^-$ bound to the center. The inhibitor bridges the Cu1 and Cu4 ions and provides clear evidence for the proposal that the Cu1–Cu4 edge forms the catalytic edge of the Cu$_Z$ center. The native as isolated structure of *A. cycloclastes* N$_2$OR has two oxygen atoms bound to the Cu1–Cu4 edge. Oxygen 1 is ligated to Cu1 and oxygen 2 to Cu4 at 2.2 and 2.5 Å, respectively. It has been suggested that this structure represents the Cu$_Z$ species of the catalytic cluster and that the longer O distance of 2.5 Å arises from a H$_2$O molecule while the shorter distance is more compatible with it being an OH$^-$. The metal cluster adopts three different redox states relevant to enzyme turnover, namely [4Cu$^+$ S]$^{2+}$, [3Cu$^+$ 1Cu^{2+}S]$^{3+}$, and [2Cu$^+$ 2Cu^{2+}]$^{4+}$.
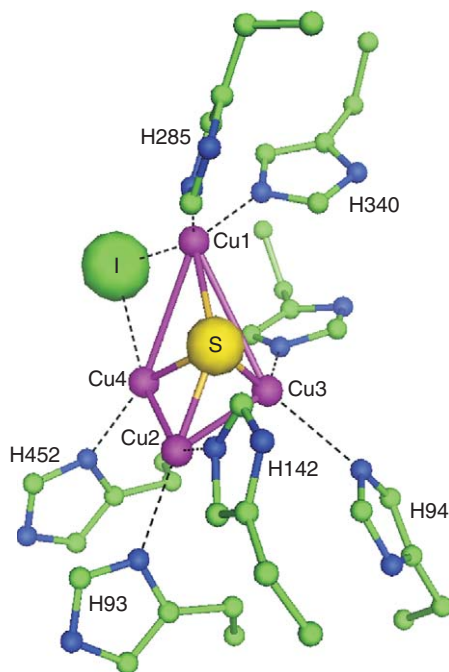
Figure 9 Cu$_Z$ cluster of the inhibitor-bound N$_2$OR structure from *Achromobacter cycloclastes* (PDB-code: 2IWK); prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

8.14.2 Copper Enzyme Catalysis

8.14.2.1 Nonblue Oxidases

The nonblue oxidases belong to a class of oxidases that extract two electrons from the reducing substrate and transfer them to dioxygen, which is reduced to hydrogen peroxide. No oxygen is transferred during this reaction.
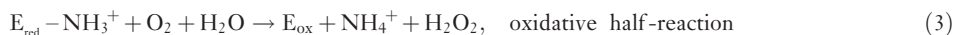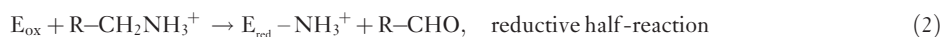
8.14.2.1.1 Amine oxidases (EC 1.4.3.6)

Amine oxidases are important regulatory enzymes catalyzing the oxidation of a wide range of biogenic amines including many neurotransmitters, histamine, and xenobiotic amines. There are two classes of amine oxidases; the flavin-containing monoamine oxidases[89] and the copper-containing amine oxidases (CuAOs).[90] For recent reviews on CuAOs see, for example, Brazeau *et al.*[91] and DuBois and Klinman.[92] The latter class occurs in all orders of organisms but little is known about their precise biological function in higher organisms. In bacteria, the CuAOs provide a route for the utilization of various amine substrates as alternative sources of nitrogen and carbon to support growth. In higher organisms, the role of CuAOs have been implicated as key components in complex processes such as leukocyte trafficking involving the CuAO, vascular adhesion protein-1 (VAP-1).[93] In animals CuAOs have been suggested to be related to development and detoxification, and in plants to development, wound and resistance responses, and secondary metabolism. Despite these functional differences, CuAOs from every known source are dimers, have subunit masses ranging from 70 to 95 kDa, and share fundamentally identical chemistry. CuAOs have been shown to contain a covalently bound cofactor, 2,4,5-trihydroxyphenylalanine quinone (TPQ), that derived from the modification of an endogenous tyrosine residue.[94] Furthermore, they also contain a single copper ion in the active site that is involved in the biogenesis of TPQ as well as the catalytic cycle involving oxidation of primary amines. In fact, CuAOs can be viewed as catalyzing three different reactions: (1) the biogenesis of TPQ, (2) the oxidation of amine substrates to generate reduced TPQ (termed the reductive half-reaction), and (3) the reduction of molecular oxygen by reduced TPQ

(termed the oxidative half-reaction). As can be seen from the stoichiometry, both the biogenesis reaction and the oxidative half-reaction require molecular oxygen.

Biogenesis:

$$E_{apo} + 2O_2 + Cu(II) \rightarrow E_{ox} + H_2O_2 \tag{1}$$

Catalysis:

$$E_{ox} + R-CH_2NH_3^+ \rightarrow E_{red}-NH_3^+ + R-CHO, \quad \text{reductive half-reaction} \tag{2}$$

$$E_{red}-NH_3^+ + O_2 + H_2O \rightarrow E_{ox} + NH_4^+ + H_2O_2, \quad \text{oxidative half-reaction} \tag{3}$$

E_{apo} is the apo-enzyme, containing the precursor tyrosine to the TPQ cofactor and no bound copper. E_{ox} is the mature native holoenzyme containing TPQ and Cu(II). E_{red} is the substrate reduced from of the enzyme, principally containing the aminoquinol from of the cofactor and Cu(II), which in some CuAOs is in equilibrium with a Cu(I)-semiquinone from of the cofactor.

Biochemical studies in solution by many different laboratories[95,96] have led to mechanistic proposals for the reactions catalyzed by the CuAOs. Recently, each mechanism has been probed by studies that have been done on CuAOs in the crystalline state using X-ray crystallographic techniques.[97–100]

The crystal structure of CuAO has been solved from *Escherichia coli* (ECAO),[57] pea seedling (PSAO),[101] *Arthrobacter globiformis* (AGAO),[102] *Hansenula polymorpha* (HPAO),[103] *Pichia pastoris* (PPLO),[104] bovine serum amine oxidase (BSAO),[105] and human vascular adhesion protein (VAP-1).[106,107]

The tertiary structures of the CuAOs that have been solved are very similar. **Figure 10** shows the structure of ECAO, which is representative of all of the known CuAO structures. The primary difference is the presence or absence of the N-terminal domain (**Figure 10**, D1), which is present in ECAO, but not in AGAO, HPAO, or PSAO. The biological function of this domain is currently unknown.

The C-terminal domain (**Figure 10**, D4) folds into a large β-sandwich composed of twisted 8- and 10-stranded β-sheets. This domain forms a large part of the intimate dimer interface, with two long β-hairpin arms reaching across from one monomer to the other. In addition, the active site of each monomer is also located in this domain. The ligands of the copper ion and the conserved sequence that includes the modified tyrosine residue are all provided by adjacent β-sheets in this domain.

In the resting native CuAO, each active site consists of Cu(II) coordinated in a distorted square pyramidal arrangement by three equatorial N ligands provided by histidines at a distance of approximately 2.0 Å, and axial water ligand (Wa) at a distance of approximately 2.4 Å and a labile equatorial water ligand (We) if present (**Figure 11**). The TPQ lies close to, but is not directly coordinated to the copper ion. The O2 position of the
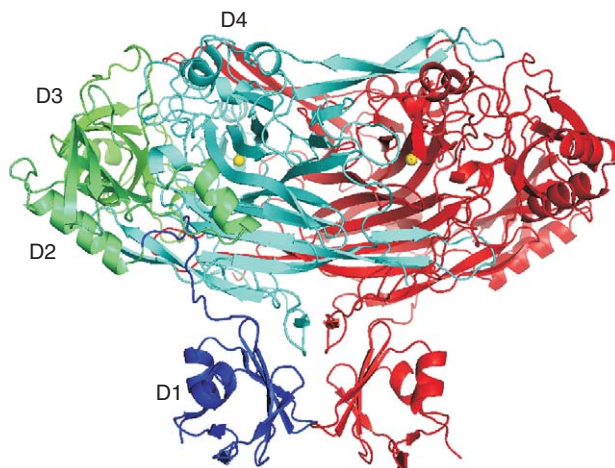
Figure 10 Ribbon diagram of the *Escherichia coli* copper-containing amine oxidase (ECAO) homodimer (PDB-code: 1DYU). One monomer is colored red. The other monomer is colored by domains (D1 blue; D2 lime green; D3 green; D4 cyan). The copper ions are shown as yellow spheres; prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).
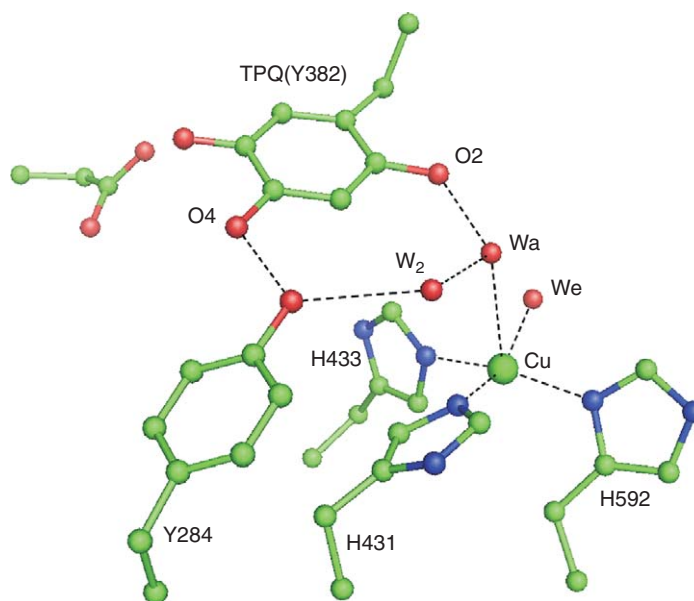
Figure 11 Active site of holo-*Arthrobacter globiformis* copper containing amine axidase (AGAO) (PDB-code: 1IVX). The holo-AGAO formed from soaking an apo-AGAO crystal in Cu(II) solution for 1 week; prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

cofactor is hydrogen bonded to Wa, while the O4 atom is hydrogen bonded to the hydroxyl of a conserved tyrosine residue. The O5 position of TPQ points into the amine substrate-binding pocket that contains a conserved aspartic acid, which functions as a catalytic base during turnover.

A proposal for the mechanism of TPQ biosynthesis (**Figure 12**) has been made on the basis of experiments involving HPAO in solution[109,110] and X-ray crystallographic studies of AGAO in which several intermediates along the biogenesis pathway of TPQ were trapped and the structures solved.[97,102]

The first step in TPQ biogenesis is presumably copper binding to apo-AGAO (**Figure 12(a)**). The phenyl group of Tyr382, the precursor to TPQ, is thought to be protonated. After binding of dioxygen at a site away from the copper center (**Figure 12(b)**), it is assumed that the precursor tyrosine becomes deprotonated and forms a Cu(II)–tyrosinate complex (**Figure 12(c)**). The fact that no other metal ion will significantly support TPQ biogenesis indicated that there is a crucial role for copper in initiating the reaction with molecular oxygen. It may be that the key role for copper is in forming a transient Cu(I)–tyrosine radical intermediate, that is, the species that reacts with dioxygen to yield the peroxo-bridged adduct (**Figure 12(d)**). The spontaneous breakdown of this species would yield 3,4-dihydroxy phenylalanine quinone (DPQ) and Cu(II)-hydroxide. An early intermediate in the biogenesis of TPQ could be obtained by soaking apo-AGAO crystals in aerobic Cu(II) solution and flash-freezing after 10 min only.[97] The crystal structure showed the Tyr382 had been modified and was either 3,4-dihydroxyphenylalanine or the oxidized quinone form DPQ as depicted in **Figure 12(e)**. In order for the TPQ biogenesis reaction to continue from the DPQ intermediate, the DPQ ring must rotate to bring C6 close to the copper center (**Figure 12(f)**). C6 of DPQ is now situated for attack by the copper-bound hydroxide (**Figure 12(f)**). After this attack and the replacement of the copper-bound hydroxyl by a water from the solvent we proceed to intermediate G in **Figure 12**. In the final step of the biogenesis reaction, TPQ is oxidized to yield holo-AGAO (**Figures 12(g)** → **12(h)**). This step requires molecular oxygen, presumably to oxidize reduced TPQ yielding hydrogen peroxide.

The stoichiometry of the reductive half-reaction is shown in Equation (2). The initial step in this reaction, as depicted in **Figure 13**, involves the nucleophilic attack of the primary amine substrate at C5 of oxidized TPQ (**Figures 13(a)** → **13(b)**). This results in the formation of a substrate Schiff base (**Figure 13(b)**). In the next step, an active site base (a conserved aspartic acid) abstracts a proton to form a carbanion species (**Figure 13(c)**), which quickly rearranges to form a product Schiff base (**Figure 13(d)**). The product aldehyde is formed by
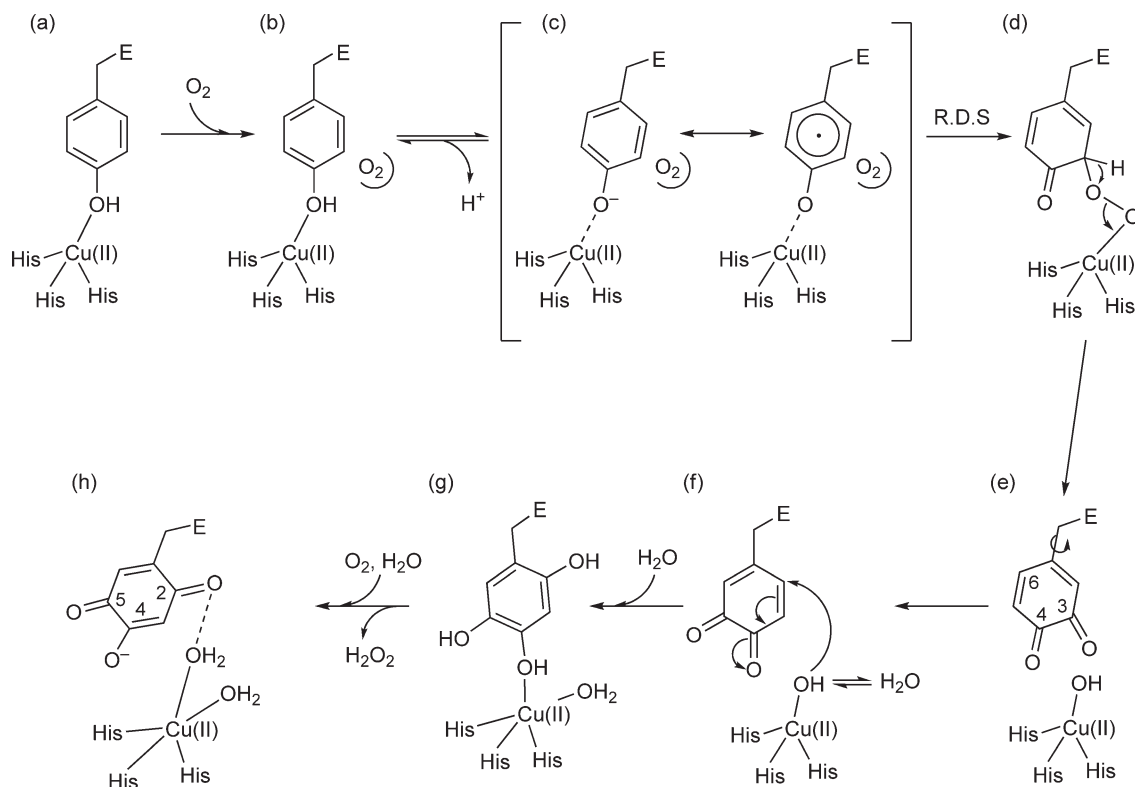
Figure 12 Scheme of the current proposed mechanism for 2,4,5-trihydroxyphenylalanine quinone (TPQ) biogenesis in copper-containing amine oxidases.[97,102,108] E represents the polypeptide of the enzyme, of which the precursor Tyr is part. R.D.S. is rate-determining step. Reproduced from B. J. Brazeau; B. J. Johnson; C. M. Wilmot, *Arch. Biochem. Biophys.* **2004**, *428*, 22–31, with permission from Elsevier Inc.

hydrolysis and the oxidized TPQ has been reduced to the aminoquinol state, with the amine nitrogen replacing the oxygen bound to C5 of the cofactor (**Figure 13(e)**). Taken together this series of steps has been termed the reductive half-reaction in reference to the final oxidation state of TPQ, which has been reduced by two electrons. The mechanisms of the reductive half-reaction depicted in **Figure 13** is well understood and widely accepted for all known CuAOs (**Figures 13(a)** → **13(e)**).[111] The chemical evidence for the reductive half-reaction mechanism has recently been reviewed.[95] For the reductive half-reaction, there is still much to be done; in particular, the current lack of a true Schiff base structure. A series of X-ray structures of intermediates that shed light on the reductive half-reaction of ECAO has been determined.[99,100] One of these intermediates is a TPQ/2-hydrazinopyridine (2-HP) adduct, where the 2-HP is covalently bound to the C5 position of the TPQ ring. In the structures of all known holo-CuAOs, the TPQ cofactor is situated in a wedge such that the C5-carbonyl is oriented toward the substrate entry channel.[98] The fact that 2-HP is covalently bound to this position is a good indication that the 2-HP reacts at the same carbon as the substrate. There are several structural features that strongly argue for this being a substrate Schiff base mimic and not a model of the product Schiff base.

The end of the reductive half-reaction leads directly to the oxidative half-reaction. In fact, it is reasonable to view the aminoquinol state (**Figure 13(e)**) as the first intermediate of the oxidative half-reaction. This is the stage in which molecular oxygen plays a key role in the mechanism. As shown in **Figures 13(e)** → **13(g)**, molecular oxygen binds to the enzyme and accepts two electrons and two protons from the aminoquinol yielding hydrogen peroxide and an iminoquinone intermediate. The hydrolysis of the iminoquinone releases ammonia and regenerates oxidized TPQ, which can then participate in another catalytic cycle.
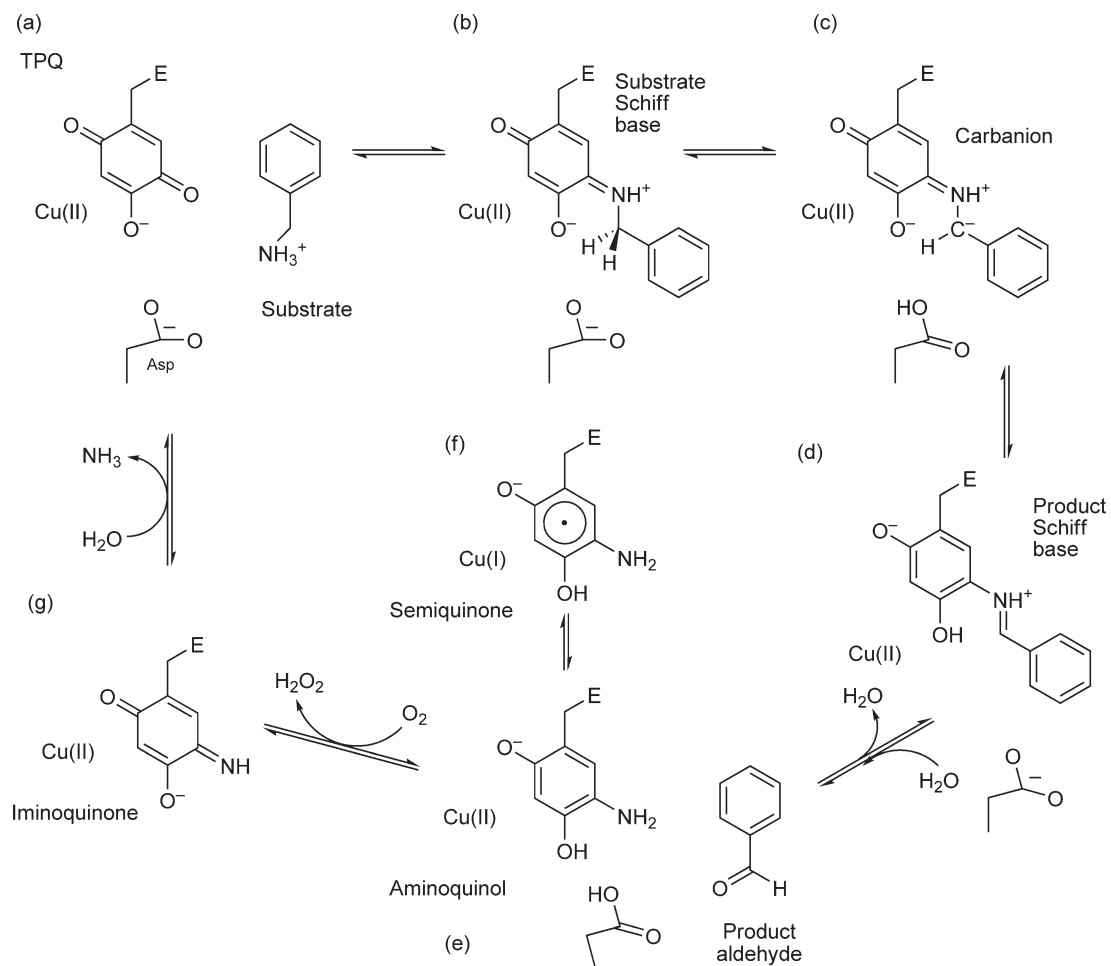
Figure 13 Scheme of the proposed catalytic cycle of copper-containing amine oxidases (CuAOs),[95,99] TPQ is the cofactor 2,4,5-trihydroxy-phenylalanine quinone, which is derived from a constitutive Tyr that is part of the polypeptide of the enzyme (E). The substrate shown is the aromatic primary amine, benzylamine, a typical substrate of some amine oxidases, such as CuAO from *E. coli* (ECAO). Reproduced from B. J. Brazeau; B. J. Johnson; C. M. Wilmot, *Arch. Biochem. Biophys.* **2004**, *428*, 22–31, with permission from Elsevier Inc.

The mechanism of the oxidative half-reaction is not as well understood as the reductive half-reaction and is the subject of a great deal of current investigation. One of the most intensely debated issues in CuAO research is what species does molecular oxygen react with? In solution, the Cu(II)–aminoquinol state of the enzyme (**Figure 13(e)**) has been observed to be in equilibrium with a Cu(I)–semiquinone species (**Figure 13(f)**).[112] However, the amount of Cu(I)–semiquinone observed differs depending on the source of the enzyme and can be undetectable in some anaerobically substrate-reduced CuAOs. Thus, the question has been raised as to whether this species actually lies on the catalytic pathway. Yet, this debate is especially intriguing because Cu(I) would chemically be the most obvious choice to donate the first electron to molecular oxygen, as has been previously proposed.[112] Several recent studies have put forth the hypothesis that the copper does not change oxidation state during the oxidative half-reaction.[95,113,114] Especially compelling was a kinetic study with HPAO in which it was shown that the turnover rate (k_{cat}) with molecular oxygen was very similar when the copper was replaced with cobalt.[115] As cobalt is very unlikely to participate in a redox reaction it has been proposed that molecular oxygen does not react with Cu(I), but instead binds to a site in the active site that is 'off-metal' and accepts the first electron from the aminoquinol.[114]

As pointed out, considerable insights into the biogenesis of TPQ and the catalytic mechanism of CuAOs have been obtained by a combination of spectroscopic, kinetic, biochemical, and X-ray structural investigations. Nevertheless, many details are still unclear and make these enzymes an interesting research field further on.

Mono- and diamine oxidases belong to the same group of amine oxidases. Diamine oxidases particularly catalyze the oxidative deamination of putrescine and histamine to aminoaldehydes, hydrogen peroxide, and ammonia.[116] Diamine oxidase activity is found in a wide range of bacteria, plants, and animals. An alignment of the amino acid sequences of human kidney amiloride-binding protein, human placenta diamine oxidase, pig kidney diamine oxidase, rat colon amiloride-binding protein, *H. polymorpha* amine oxidase, and lentil seedling amine oxidase shows that these proteins belong to the same group of enzymes and identifies the amiloride-binding protein and diamine oxidase as identical. All copper ligand residues and the TPQ-consensus sequence are conserved.[58] Human kidney diamine oxidase could recently be overexpressed as a secreted enzyme under the control of a metallothionein promoter in *Drosophila* S2 cell culture.[117] The enzyme material has been characterized by visible absorption, CD, EPR, and resonance Raman spectroscopy. The metal content was determined either by inductively coupled plasma (ICP) emission spectroscopy or flame atomic absorption. The recombinant enzyme contains the cofactors TPQ and copper at stoichiometries of up to 1.1 and 1.5 mol mol^{-1} homodimer, respectively. In addition, tightly bound and stoichiometric calcium ions were identified and proposed to occupy a second metal-binding site. The apparent molecular weight of the recombinant protein, determined by analytical ultracentrifugation, suggests 20–26% glycosylation by weight. Kinetic studies indicate that the preferred substrates of human diamine oxidase are, in order, histamine, l-methylhistamine, and putrescine, with K_M values of 2.8, 3.4, and 20 μmol l^{-1}, respectively. An X-ray crystal structure of the human kidney diamine oxidase is still lacking probably due to the high carbohydrate content, which hampers its crystallization.

8.14.2.1.2   *Lysyl oxidase (EC 1.4.3.13)*

Mammalian lysyl oxidase is a copper amine oxidase that initiates covalent cross-linkage formation in elastin and collagen by oxidizing peptidyl lysine in these proteins to aminoadipic semialdehyde.[118] The catalytic reaction can be considered to be the sum of two component reactions similar to the scheme for the other mono- and diamine oxidases.[119] Removal of copper from the preparations of lysyl oxidase results in the loss of enzyme activity as does the exposure of the enzyme to copper chelating agents.[120,121] Therefore, copper-deficient nutrition leads to a reduction of cross-linking of elastin and collagen and thus to faulty connective tissue formation. Among the copper amine oxidases the physiologically important lysyl oxidase has been a puzzle. Although formally a member of the copper amine oxidase family, lysyl oxidase shows a number of essential differences. These include the smaller size (monomer of 32 kDa) in relation to other copper amine oxidases (dimers of 75–85 kDa subunits) and the absence of the conserved consensus sequence in all known TPQ-containing enzymes.[122] Later, a previously unknown redox cofactor has been identified in the active site of lysyl oxidase from bovine aorta.[123] The cofactor is a quinone whose structure is derived from the cross-linking of the ε-amino group of a peptidyl lysine with the modified side chain of a tyrosyl residue and it has been designated lysine tyrosylquinone (LTQ). EPR spectra analysis is consistent with a single type-2 copper ion with at least three nitrogen, that is, histidine ligands.[119]

A lysyl oxidase from the yeast *P. pastoris* (PPLO) has been overexpressed in *P. pastoris* GS115 cells and its crystal structure determined.[104,124] The PPLO is a CuAO with a TPQ cofactor and structurally closely related to the amine oxidases discussed in Section 8.14.2.1.1. But they differ from most other members of the CuAO enzyme family in possessing the ability to oxidize the side chain of lysine residues in a polypeptide. The important findings of these structure determinations are that the TPQ cofactor is disordered while the rest of the active site is well ordered, that there is a gated pathway to the Cu site from the central 'lake' and that the enzyme has apparently oxidized one of its own lysine side chains with the resulting formation of an intramolecular lysine–lysine cross-link. This clearly demonstrates the peptidyl lysyl activity of PPLO. In the absence of structural information for a true mammalian LTQ containing lysyl oxidase, PPLO is currently the best model for the lysyl oxidases.

### *8.14.2.1.3   Galactose oxidase (EC 1.1.3.9)*

Galactose oxidase[125–127] is a fungal secretory enzyme widely used in bioanalytical and histological applications. It is one of the best-characterized free radical enzyme. The overall reaction catalyzed by galactose oxidase is the oxidation of a primary alcohol to the corresponding aldehyde, coupled to the reduction of dioxygen to hydrogen peroxide,[128] the biologically important product:

$$R–CH_2OH + O_2 \rightarrow R–CHO + H_2O_2 \tag{4}$$

Since both alcoholic oxidation and O_2 reduction are two-electron processes, the catalytic reaction is conceptually equivalent to a transfer of the elements of dihydrogen between the two substrates. Biological hydrogen transfer generally involves specialized organic redox factors (e.g., flavins, nicotinamide, quinones), with well-characterized reaction mechanisms. Galactose oxidase does not contain any of these conventional redox factors and instead utilizes a very different type of active site, a free radical-coupled copper complex, to perform this chemistry.[129] The new type of active site structure implies that the reaction follows a novel biochemical redox mechanisms based on free radicals and the two-electron reactivity of the metalloradical complex.

Substrate specificity is very broad, ranging from small alcohols to polysaccharides with D-galactose at the nonreducing terminus.[130] The best substrate reported so far is dihydroxyacetone, which is more than three times better than D-galactose. Despite this, galactose oxidase is strictly stereospecific and does not oxidize either D-glucose or L-galactose.

The crystal structure of galactose oxidase from the fungus *Dactylium dendroides* has been determined.[56,131] Accordingly, galactose oxidase (639 amino acid residues) consists of three domains predominantly formed from β-structures (**Figure 14**). The first domain (residues 1–155) has a β-sandwich structure. The catalytic domain (residues 156–532) comprises a seven-fold β-propeller based on the kelch structural motif. The copper lies on the solvent-accessible surface of this domain close to the pseudo seven-fold axis. The third domain (residues 533–639) is comprised of seven β-strands. The copper site on the second domain lies in a region extremely rich
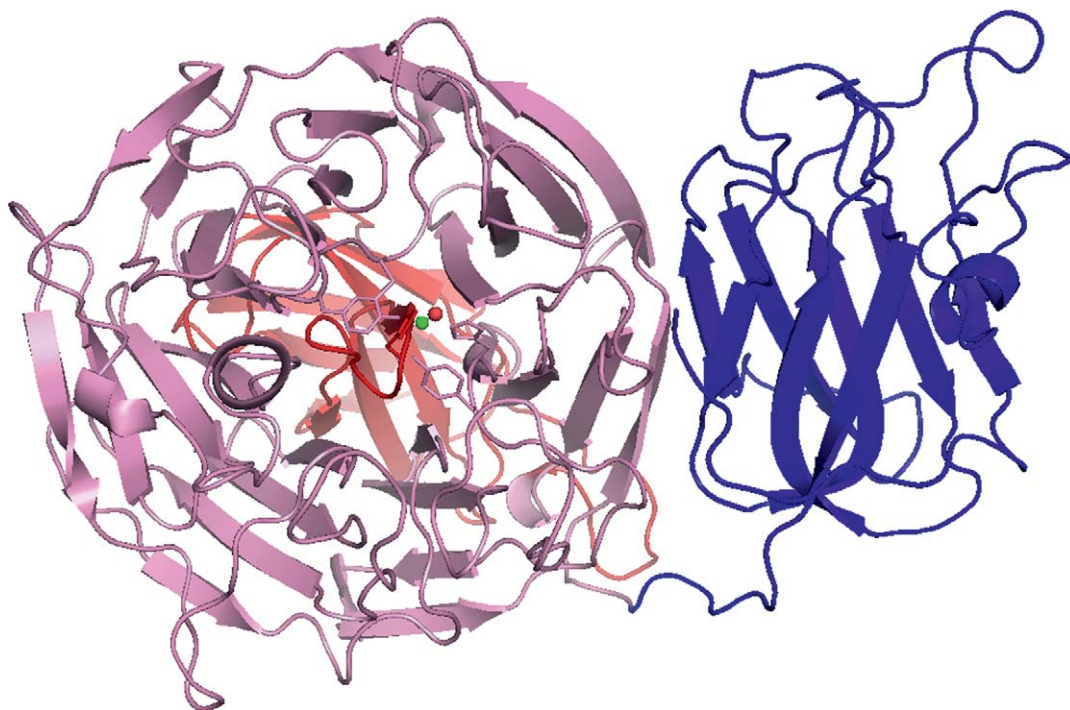
Figure 14   Ribbon diagram of galactose oxidase from *Dactylium dendroides* (PDB-code: 1GOG). The colors of the individual domains are: domain 1 (1–155), blue; domain 2 (156–532), pink; domain 3 (533–639), red; prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

in aromatic residues. Side chains of three tyrosines, three phenylalanines, three histidines, and one tryptophan all lie within 7 Å of the copper. Some seem to be involved in the formation and stabilization of the free radical whereas others form a hydrophobic wall for the putative substrate-binding pocket. The active site structure is illustrated in **Figure 4**, showing the coordination of the copper ion by two histidine residues (His496 and His581), a simple tyrosinate (Tyr495), and a covalently modified tyrosine (Tyr272) cross-linked to a cysteinyl residue (Cys228) to form a new dimeric amino acid (cysteinyl-tyrosine). This feature has been identified by spectroscopic and modeling studies as the radical-forming site in galactose oxidase.[132,133] As the copper is EPR-inactive it was initially assumed that the copper changes between the Cu(I) and Cu(III) oxidation states. According to recent investigations the copper is in the Cu(II) state and is antiferro-magnetically coupled with a tyrosyl radical (Tyr272) and therefore EPR-inactive.[56,134] The thioether bond that links the two residues affects both the structure and reactivity of the protein. Structurally, the cross-link contributes to the rigidity of the active site, similar to the effect a disulfide bond would have on the protein. However, unlike a disulfide bond, the thioether bond is formed irreversibly and is not susceptible to reductive cleavage. The cross-link forms spontaneously in the protein in the presence of reduced copper (Cu^+) and dioxygen.[135]

The reaction of dioxygen with the Cu^+ precursor complex is expected to result in one-electron reduction of O_2, implying the involvement of free radical intermediates in the biogenesis reaction, although there is no spectroscopic evidence for build-up of radicals in the rapid reaction experiments and further investigations will be necessary to identify the radicals involved. A detailed mechanism consistent with the experimental results has been proposed (**Figure 15**).[135] The reaction is initiated by Cu^+ binding to a preorganized active site (**Figures 15(a) and 15(b)**). Dioxygen reacts with this complex to produce an oxygenated species. An end-on superoxo complex (**Figure 15(c)**) would be expected to react in the outer sphere with the Cys228 thiol to form a thiyl free radical (**Figure 15(d)**). Addition to the thiyl free radical to the Tyr272 ring system would break the aromatic conjugation of the ring (**Figure 15(e)**), which would be restored on deprotonation and reduction of the metal ion (**Figure 15(f)**). This mature, cross-linked complex corresponds to the fully reduced active site formed during turnover and is expected to undergo very rapid reaction with a second molecule dioxygen to form the oxidized metalloradical complex. An alternative path, involving side-on bound oxy species and inner-sphere reaction of the initial complex, might proceed via a tyrosyl phenoxyl radical intermediate. This mechanism accounts for the metal oxidation-state dependence of the reaction, the O_2 stoichiometry, the pH and isotope sensitivity, and the requirement for a metalloradical complex in the product.

Galactose oxidase operates by a free radical mechanism. The enzyme is isolated as a mixed redox state in which only some 5% of molecules are active. Full activation can be achieved *in vitro* by a number of one-electron oxidants that facilitate the removal of one electron to create the Tyr272-based radical (**Figure 16**, R → O).

Molecular modeling studies for substrate binding are consistent with direct binding of the C6 alcohol group to copper at the equatorial site normally occupied by water (**Figure 16**, O → 1).[56,136] Displacement of a water close to the copper has been demonstrated by water proton relaxation data using dihydroxyactetone as substrate.[137]

The first step in catalysis is the activation of substrate by the transfer of the O6 hydroxyl proton to Tyr495, the axial copper ligand, leading to its dissociation (**Figure 16**, 1 → 2). This step was proposed from azide-binding studies, which demonstrated uptake of a solvent-derived proton.[138] The next step (**Figure 16**, 2 → 4) is the abstraction of the *pro*-S hydrogen from the C6 methylene group, a process that displays a significant kinetic isotope effect with C6 deuterated D-galactose.[139] A hemolytic radical hydrogen transfer from substrate to Tyr272 is strongly implied by the presence of the free radical on Tyr272 and this is supported by the lack of a solvent isotope effect on the rate-limiting step of the reductive half-reaction as recently demonstrated.[139] At low galactose concentrations the rate-limiting step is reoxidation of the reduced enzyme. Since the kinetic isotope effect is also observed when oxygen is rate limiting it is most likely that the hydrogen atom from the substrate is retained in the reduced enzyme active site and contributes in the later reoxidation. The transfer of a single electron from substrate to Cu^{2+} could precede or follow hydrogen atom transfer to Tyr272; however, Whittaker *et al.*[139] favor the former as a consequence of the now 4-coordinate geometry of the Cu^{2+} that would allow rapid inner sphere electron transfer. The aldehyde product would then be expected to be released from reduced enzyme according to the ping-pong
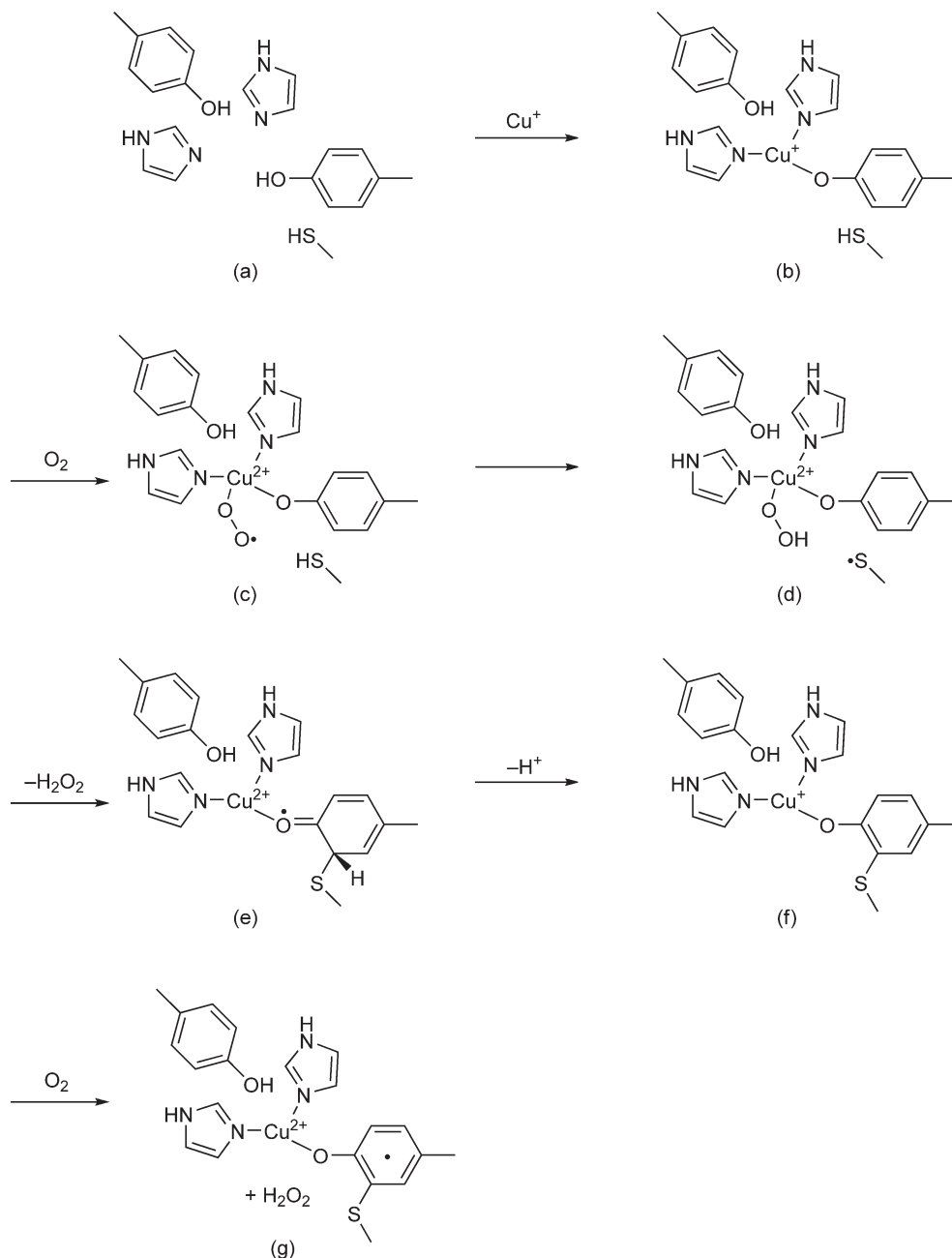
Figure 15 Proposed mechanism for biogenesis of the Tyr–Cys cofactor in galactose oxidase.[135] Reproduced from M. M. Whittaker; J. W. Whittaker, *J. Biol. Chem.* **2003**, *278*, 22090–22101, with permission from Elsevier Inc.

mechanism with dioxygen which will bind then (**Figure 16**, $4 \rightarrow 5$). Reoxidation of the enzyme would occur by electron and proton transfers to dioxygen, which is proposed to bind to the Cu^+ leading to the formation of the hydrogen peroxide product (**Figure 16**, $5 \rightarrow 6$). There is no direct evidence of oxygen binding to copper and it is not clear whether it would occupy the site vacated by aldehyde. In this regard, the benefits of cryocrystallography of trapped intermediates are obvious and may allow the direct visualization of oxygen bound to copper. Unfortunately, such experiments have been unsuccessful up to now.
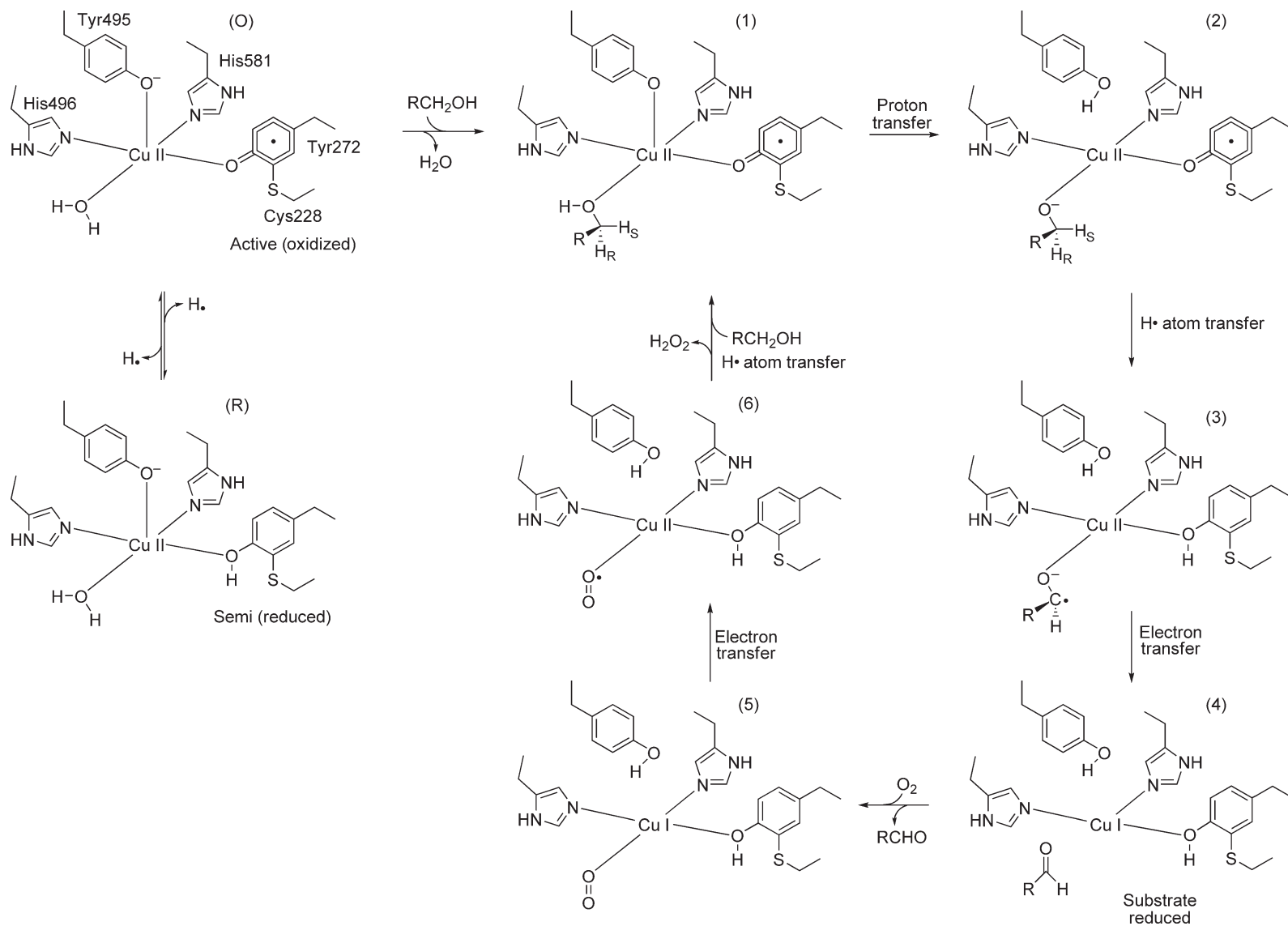
Figure 16 Proposed catalytic mechanism for galactose oxidase.[127] Reproduced from M. J. McPherson; M. R. Parsons; R. K. Spooner; C. M. Wilmot, In *Handbook of Metalloproteins*; A. Messerschmidt, R. Huber, T. Poulos, K. Wieghardt, Eds.; John Wiley & Sons: Chichester, 2001; Vol. 2, pp 1272–1283, with permission from [126] John Wiley & Sons.

8.14.2.1.4 Glyoxal oxidase

Glyoxal oxidase from the white-rot wood-metabolizing basidiomycete *Phanerochaete chrysosporium* has been characterized as a novel radical–copper oxidase.[140] This enzyme exhibits a wide substrate specificity for oxidation of simple aldehydes to the corresponding carboxylic acids according to Equation (5).

$$RCHO + H_2O + O_2 \rightarrow RCO_2H + H_2O_2 \tag{5}$$
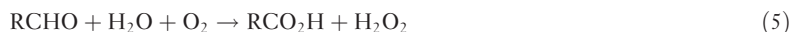
Spectroscopically, a cysteinyl–tyrosyl active complex has been determined similar to that in galactose oxidase. Glyoxal oxidase shares many properties with galactose oxidase and a modest protein sequence similarity including 28% identity over their primary structures, but catalyzes a distinct reaction. The sequence of glyoxal oxidase, the smaller of the two proteins, matches the C-terminal three-quarters of galactose oxidase, implying that glyoxal oxidase lacks an approximately 150-residue N-terminal domain that is present in the larger enzyme. However, alignment of the protein sequences indicates that the critical active site residues character-istic of the radical–copper oxidases are conserved between these structures, consistent with spectroscopic comparisons[140] that demonstrate a close structural similarity of the active sites in these enzymes.

Three of four targeted active site residues (Y377F, C70A, and Y135F) could heterologously be expressed in both a filamentous fungus (*Aspergillus nidulans*) and in a methylotrophic yeast (*P. pastoris*).[141] Biochemical and spectroscopic characterization of the mutant enzymes supports structural correlation between both enzymes identifying the catalytically important residues in glyoxal oxidase and demonstrating the functions of each of these residues.[141]

The enzymes represent members of a growing class of free radical metalloenzymes based on the radical–copper catalytic motif and appear to represent functional variants that have evolved to distinct catalytic roles.

8.14.2.2 Monooxygenases

Monooxygenases represent a class of enzymes catalyzing a reductive activation of dioxygen, linked to the insertion of an oxygen atom into a C–H bond according to Equation (6).

$$R{-}H + O_2 \xrightarrow{2e^-,\,2H^+} R{-}OH + H_2O \tag{6}$$

The basic properties of monooxygenase-type systems that must be taken into account in the development of a reaction mechanism include (1) the splitting of the dioxygen bond, (2) the interaction of spin unpaired (triplet) oxygen with spin paired (singlet) substrates, and (3) the large thermodynamic driving force deriving from the greater bond energy of the O–H of water relative to the C–H of substrate. In order to circumvent the spin forbidden properties of these reactions, nature has evolved a number of strategies to activate dioxygen involving the use of either organic cofactors or transition metal ions. Although the total number of enzyme systems found to use copper in this capacity is small, eukaryotic copper-containing enzymes catalyze physio-logically important monooxygenase reactions, for example, those catalyzed by tyrosinase,[142] dopamine-β-monooxygenase,[143] and peptidyl-α-amidating enzyme.[144]

8.14.2.2.1 Peptidylglycine α-hydroxylating monooxygenase

Many peptide hormones and neuropeptides require amidation of their C-terminus for biological activity. Peptidylglycine α-hydroxylating monooxygenase (PHM) is one part of the peptidylglycine α-amidating monooxygenase (PAM) and catalyzes the first step of the two-step amidation reaction, the hydroxylation of the Cα of a C-terminal glycine.[144] PHM, together with dopamine β-monooxygenase (DβM), belongs to a small class of copper proteins found exclusively in higher eukaryotes. Both enzymes are localized in subcellular compartments: the chromaffin vesicles of the adrenal glands and synaptic vesicles of the sympathetic nervous system (DβM) and the secretory vesicles of the pituitary gland (PHM). Kinetic and mutagenesis studies,[145,146] detailed kinetic isotope effect measurements,[147] and extended X-ray absorption fine structure (EXAFS) measurements[146,148] as well as determination of the crystal structures of several forms of the enzyme[60,149,150] have provided notable insights into the architecture and function of PHM.

Figure 17 Ribbon diagram of peptidylglycine α-hydroxylating monooxygenase (PHM) (PDB-code: 1SDW). The copper ions are displayed as yellow spheres; prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

The catalytic core of PHM is composed of two nine-stranded β-sandwich domains. Each domain is about 150 residues long and binds one catalytic Cu (**Figure 17**). The domains have similar topologies, each containing a common eight-stranded antiparallel jelly-roll motif. The interior of both domains are highly hydrophobic and lack any charged residues. Domain 1 binds one active site Cu, Cu_A, with three histidines (His107, His108, and His172) as ligands, which occupy three of the four equatorial positions of a square pyramidal complex. Domain 2 binds the second catalytic Cu, Cu_B, with two histidines and a methionine (His242, His244, and Met314). Together with a bound water molecule, Cu_B shows a tetrahedral coordination. Unfortunately, these copper ions have been designated as Cu_A and Cu_B as well in the X-ray structures of PHM[60,149,150] and must not be confused with the binuclear Cu_A site and the Cu_B site from COX. In DβM they have been unambiguously termed as Cu_H and Cu_M, respectively. The two coppers are 11 Å apart and face the interdomain space in such a way that the cleft between them is fully accessible to solvent. The precatalytic complex of PHM with the bound peptide N-acetyl-diiodo-tyrosyl-D-threonine (IYT) and dioxygen as determined from the crystal structure[150] is shown in **Figure 18**. The D-threonine C-terminus is anchored by a bidentate salt bridge with the guanidinium group of Arg240 and a hydrogen bond with side chain of Tyr318. The side chain of Asn316 forms an additional hydrogen bond to the D-threonine main chain amide.

Dioxygen binds to Cu_B with an end-on η^1 geometry in the precatalytic complex (**Figure 19**). This geometry is compatible with dioxygen or superoxide bound to copper, but not with Cu-peroxo species. The catalytic mechanism of PHM will be discussed in the next section because the reaction schemes of PHM and DβM are very similar.

8.14.2.2.2 *Dopamine β-monooxygenase (EC 1.14.13.1)*

DβM, found within vesicles of the adrenal medulla and noradrenergic nerve cells, catalyzes the conversion of dopamine to norepinephrine. It is the third step in the catecholamine biosynthetic pathway and provides norepinephrine for the sympathetic nervous system (see, e.g., Stewart and Klinman,[143] Klinman[111]). The conversion of dopamine to norepinephrine involves an insertion of an atom of oxygen into the benzylic position of the ethylamine side chain of dopamine and other phenylethylamine analogues according to Equation (7).

$$RCH_2CH_2NH_3^+ + O_2 \xrightarrow{2e^-,\ 2H^+} RCHOHCH_2NH_3^+ + H_2O \qquad (7)$$
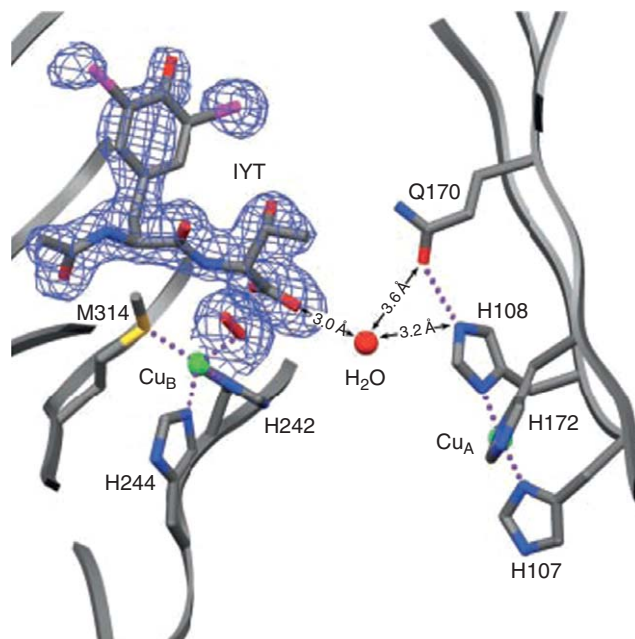
Figure 18 The precatalytic complex of PHM with bound IYT peptide and dioxygen. The 2Fo–Fc electron density is shown for the IYT peptide. Reproduced from S. T. Prigge; B. A. Eipper; R. E. Mains; L. M. Amzel, *Science* **2004**, *304*, 864–867, with permission from American Association for the Advancement of Science.
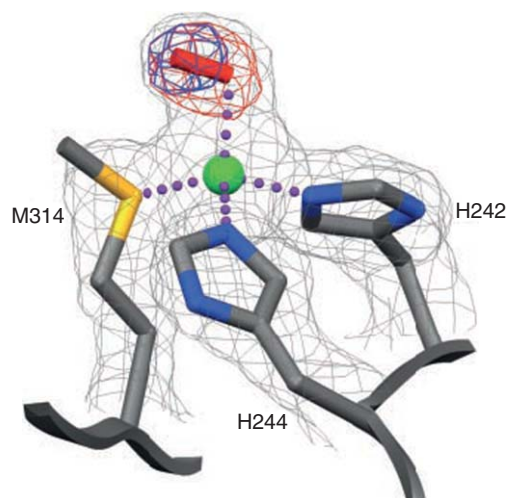
Figure 19 The structure of the binding site. Dioxygen (the red rod) is shown bound to Cu_B (the green sphere) in an end-on manner. The 2Fo–Fc omit map is displayed as well. Reproduced from S. T. Prigge; B. A. Eipper; R. E. Mains; L. M. Amzel, *Science* **2004**, *304*, 864–867, with permission from American Association for the Advancement of Science.

As this reaction implies, the reduction of oxygen to water is a four-electron process, with two electrons coming from the substrate and two electrons from an exogenous electron donor. The naturally occuring donor is ascorbic acid, which undergoes oxidation in two sequential steps to generate $2 \, mol \, l^{-1}$ of semihydroascorbate:

$$2(\text{ascorbate}) \rightarrow 2(\text{semidehydroascorbate}) \tag{8}$$
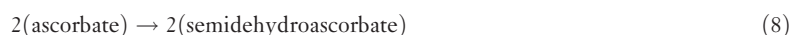
The enzyme exists in both membrane-bound and soluble forms. The holoenzyme is a tetrameric glycoprotein monooxygenase with a molecular weight of 290 kDa, consisting of two disulfide-linked dimers.[151,152] When

total cellular DβM from bovine or rat tissue is analyzed by gel electrophoresis, two subunit sizes of 72–73 kDa and 75–77 kDa are observed. The lower molecular mass subunit is the primary component of the soluble form of the enzyme. The higher molecular mass subunit is associated with the membrane-bound form of the enzyme.[153] Both membrane-bound and soluble forms of DβM are generated from one primary translation product and the recombinant protein is enzymatically active.[154]

DβM contains two coppers per monomer and these Cu atoms are not equivalent. Based on spectroscopic data, it has been proposed that the oxidized enzyme contains two tetragonal Cu(II) centers with no more than three histidine ligands per Cu. As already mentioned, the Cu^{2+} sites are inequivalent with Cu$_H$(His)$_3$(H$_2$O)...Cu$_M$(His)$_2$X(H$_2$O) coordination. The identity of X was unknown, Reduction by ascorbate causes loss of the bound water molecules since water is a poor ligand for Cu$^+$. This leaves two 3-coordinate centers, one of which, Cu$_M$, now coordinates a S from methionine. Alternatively, this S could already be present in the oxidized form as a weakly bound axial ligand. The mononuclearity of the Cu centers requires Cu$_H$ to be at some distance from Cu$_M$ and consequently must serve to shuttle electrons into the hydroxylating site.[155,156] Cu$_M$ has been identified as the CO, and by inference, O$_2$-binding site.[157]

Comparison of the primary sequence of PHM with the larger DβM indicates a central core of approximately 300 amino acids from DβM that is 27% identical to PHM. In addition, DβM contains approximately 200 amino acids toward its N-terminus and another approximately 100 amino acids toward the C-terminus that bear no relationship to PHM. Of particular note is the conservation of the ligands of both copper sites in the two related enzymes, which were identified in the crystal structure of PHM.[60] The structure of PHM reveals many striking and unexpected features. These include a lack of bridging ligands between the copper sites. It is known that two electrons consumed during substrate hydroxylation are stored in the Cu$_H$ and Cu$_M$ sites.[156] Although it is conceivable that the metal centers could approach each other during the catalytic cycle, there is no structural evidence for a hinge region capable of facilitating such a movement. Although DβM and PHM belong to a family of multicopper proteins, the coppers appear to perform different functions, that of substrate hydroxylation (Cu$_M$) and electron storage/transfer (Cu$_H$). Perhaps the most startling feature to emerge from the X-ray studies is the fully solvent-exposed nature of the copper sites, raising the questions of (1) how DβM and PHM carry out the region- and stereospecific hydroxylations and (2) how they carry out controlled ET from Cu$_H$ to Cu$_M$ through bulk solvent.

The copper enzyme family of DβM and PHM has recently been reviewed.[158] It could be stated that the solution studies indicate that DβM and PHM are mechanistically interchangeable. The large body of functional data obtained from studies in solution has been evaluated in the context of the now available structural information and a mechanism common for DβM and PHM has been proposed.[159] The mechanism is shown in **Figure 20**.

The formation of a copper–superoxo intermediate appears to provide a working mechanism that is capable of rationalizing the voluminous amount of data available for DβM and PHM. The expanded mechanism of **Figure 20** also provides an answer to the long-standing question of how these enzymes catalyze ET across bulk water at a rate that is compatible with catalytic turnover.

Starting with the fully reduced enzyme on the upper left-hand side of **Figure 20**, substrate and O$_2$ bind to produce the ternary complex. This is the trigger for initial O$_2$ activation involving ET from Cu$_M$(I) to O$_2$ to form the EPR-silent copper–superoxo intermediate. The latter is believed to be generated in an energetically uphill process, consistent with the impact of substrate deuteration on the O-18 isotope effects and the fact that oxygen uptake and product formation are so tightly coupled. Recent density functional theory calculations support such an energetic view of Cu$_M$(II)($^-$O$_2^{\cdot}$) in DβM/PHM.[160] Within the protein active sites, the reactivity of the Cu(II)–superoxo species is expected to be tightly linked to the degree of charge transfer from metal to O$_2$, together with the tightness of binding of the resultant superoxide anion to Cu(II). Subsequent transfer of a hydrogen atom from substrate via tunneling will also be linked to the ability of the protein to sample many different configurational substrates; only a subset of these configurations is expected to possess the requisite energetic and internuclear distance requirements that can give rise to efficient wave function overlap from the hydrogen in the donor substrate to acceptor oxygen.[161] These properties suggest that a great deal of 'subtle tuning of reactivity' is at work within the active sites of DβM and PHM.

One of the more gratifying aspects of the mechanism in **Figure 20** concerns the stage at which the second electron from Cu$_H$(I) enters into the reaction mechanism. The rate constant for the C–H bond cleavage step with the natural substrate dopamine in DβM is almost 103 s^{-1}.[162] If ET from Cu$_H$ to Cu$_M$ preceded the
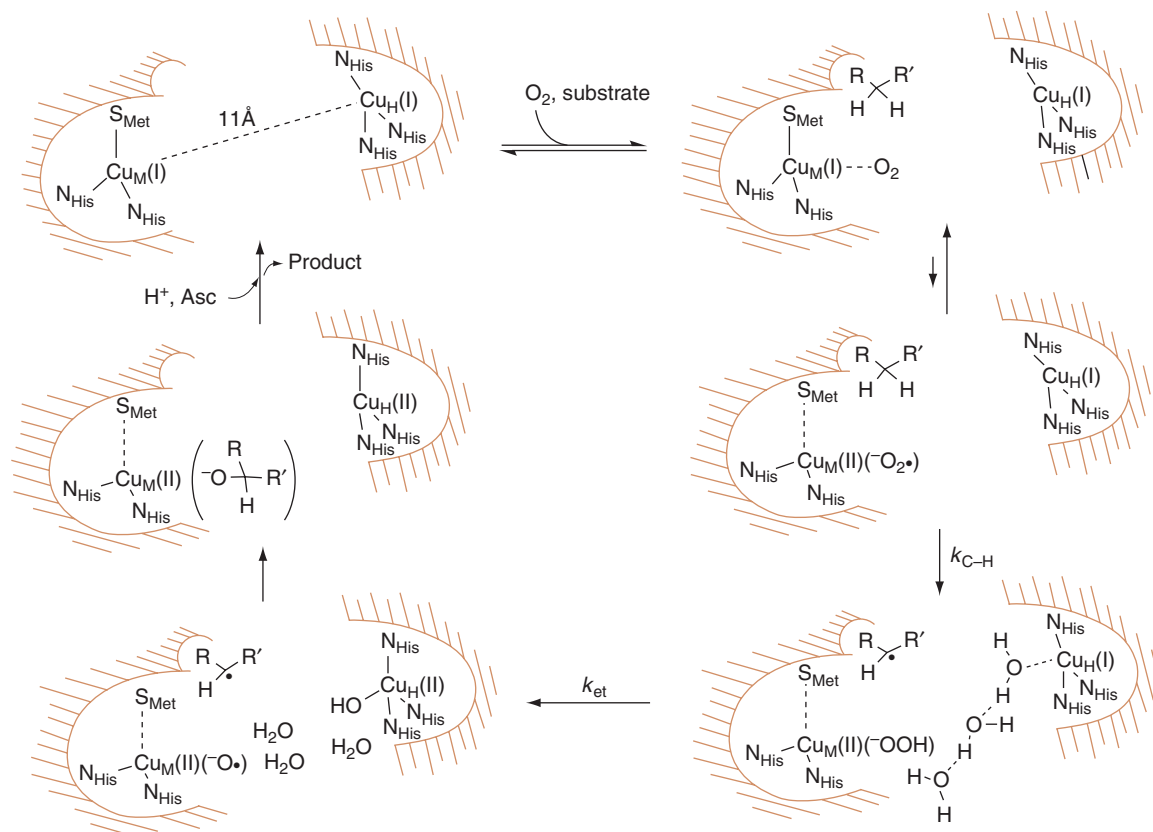
Figure 20 Copper–superoxo mechanism for DβM and PHM.[159] Reproduced from J. P. Klinman, *J. Biol. Chem.* **2006**, *281*, 3013–3016, with permission from The American Society for Biochemistry and Molecular Biology.

substrate activation step, it would have to occur significantly faster than $103\,s^{-1}$. Many authors have debated how the electron could move this quickly through bulk water, proposing pathways that involve portions of the protein and/or the substrate (e.g., Bell *et al*.[145] and Francisco *et al*.[163]). Experimental testing of these proposals has, thus far, failed to provide support for their existence. In the context of **Figure 20**, the second electron does not transfer to the Cu$_M$ site until after an irreversible hydrogen atom transfer, placing the long-range ET into k_{cat}, which is a much slower process, approximately $40\,s^{-1}$.[145]

This not only relaxes the kinetic constraints for the electron transfer step but alters its thermodynamic driving force. In the original mechanism for DβM proposed by Evans *et al*.,[159] the intermediate Cu$_M$(II)–OOH is proposed to undergo reductive cleavage by Cu$_H$ via k_{et} to produce water and a CuM(II)–oxo radical, which then rapidly recombines with the substrate-derived radical to give an inner sphere alcohol product. The computations of Solomon and coworkers[160] suggest an alternative in which the substrate-derived radical abstracts hydroxyl radical directly from the Cu$_M$(II)–OOH to form unliganded product; the resulting CuM(II)–oxo species then undergoes reduction by Cu$_H$ to Cu$_M$(II)–OH in a thermodynamic favorable process. One feature of the DβM mechanism that may favor the inner sphere product complex shown in **Figure 20** is the observation that k_{cat} is faster with phenethylamine substrates containing electron-withdrawing substituents. This has been used to argue for the intermediacy of an inner sphere alcohol product complex that undergoes (partially) rate-limiting dissociation to free product as part of the k_{cat} process.[164]

A new stage has been reached where these bizarre and beautiful enzymes have begun to reveal their unique chemistry. Many experimental challenges remain, which include the precise tuning of the active site for hydrogen transfer and the possible participation of regions of the protein distal from the active site in this process. The vexing question of the exact mechanism of long-range ET between the Cu$_H$ and Cu$_M$ sites also awaits elaboration.

8.14.2.2.3 *Particulate methane monooxygenase (EC 1.14.13.25)*

Methane monooxygenase (MMO) is the first enzyme in the metabolic pathway of methanotrophs, which are bacteria that use methane as their sole source of carbon and energy.[165] Methanotrophs have been used to combat emissions of methane,[166] a potent greenhouse gas, and are also useful for bioremediation because they can oxidize halogenated hydrocarbons.[165,167] Membrane-bound particulate MMO (pMMO), soluble MMO (sMMO),[168] and the related enzyme ammonia monooxygenase (AMO)[169] are the only known enzymes capable of methane hydroxylation by activating the inert ($104 \, \text{kcal} \, \text{mol}^{-1}$) C–H bond in methane. Therefore, they are relevant to the development of new industrial catalysts. All methanotrophs produce pMMO, which is housed in intracytoplasmic membranes. Under copper-limiting conditions, several strains also produce sMMO.[165] The well-studied sMMO system comprises a hydroxylase (MMOH), a reductase, and a regulatory protein.[168] The crystal structure of MMOH, which contains a carboxylate-bridged di-iron center, has been known for a decade.[170] In contrast, most questions surrounding the biochemistry, structure, and mechanism of the predominant methane oxidation enzyme, pMMO, have remained unanswered despite considerable research efforts in the past 20 years.[171]

pMMO is composed of three subunits, pmoB (α, ~47 kDa), pmoA (β, ~24 kDa),[172] and pmoC (γ, ~22 kDa),[173] each containing predicted membrane-spanning helices. The molecular mass and oligomerization state of pMMO are not established, but 100 kDa $\alpha\beta\gamma$[174] and 200 kDa $\alpha_2\beta_2\gamma_2$[175] polypeptide arrangements have been proposed. The metal content of pMMO is controversial, with reported values of 2–15 copper ions[174] and 0–2 iron ions[175] per 100 kDa purified pMMO.

The characterization of pMMO has led to three different models of the metal center(s) discussed in Lieberman and Rosenzweig.[176] As the crystal structure determination of pMMO from *Methylococcus capsulatus* (Bath)[176] has established the nature of the metal sites details on these different models are not provided here. The pMMO structure shows that three copies each of the pmoA, pmoB, and pmoC subunits form a cylindrical $\alpha_3\beta_3\gamma_3$ trimer approximately 105 Å long and approximately 90 Å in diameter. A soluble region composed mainly of six β-barrel structures, two from each protomer, extends approximately 45 Å away from the membrane and is supported by 42 transmembrane (TM) helices, 14 from each protomer. A hole is formed in the center of the trimer. The trimeric structure of pMMO was not anticipated and provides the first experimental evidence for a 1:1:1 subunit ratio.

Each protomer in the trimer comprises single copies of the pmoB, pmoA, and pmoC subunits (**Figure 21**). pmoB includes residues 33–414. The first 32 residues are proposed to be a leader sequence.[172] The soluble regions are derived primarily from pmoB and include two antiparallel β-barrel structures, one at the N-terminus and the other at the C-terminus. The N-terminal β-barrel is composed of seven strands, and is oriented approximately 90° from the eight-stranded C-terminal β-barrel. The two β-barrel structures are separated by a β-hairpin followed by two TM helices. A 22-residue loop links the TM helices to the C-terminal β-barrel. Residues from this loop participate in trimer interface interactions with the β-barrel structures from the adjacent protomer, whereas the β-hairpin is involved in intraprotomer stabilization. The pmoB subunit houses a dinuclear copper center as well, in the N-terminal β-barrel (see below in this section).

The pmoA and pmoC subunits reside primarily in the membrane. The pmoA subunit consists of seven TM helices and packs against the two TM helices from pmoB (**Figure 21**). These helices, which span the range of observed TM helix lengths, are inclined with respect to one another, and several are quite tilted with respect to the lipid bilayer normal. The two C-terminal helices face the opening at the trimer center and interact with their counterparts from the other two protomers. A short helix and a β-hairpin structure protrude from the membrane, and interact with the soluble region of pmoB. pmoC comprises five TM helices that are oriented approximately parallel to the membrane normal and to one another (**Figure 21**).

The crystal structure reveals three metal centers per protomer (**Figures 21 and 22**). The first site (**Figure 22(a)**) is located in pmoB approximately 25 Å above the membrane and near the surface of the N-terminal β-barrel. The metal ion was assigned to be copper from the strong peaks in anomalous Fourier maps calculated using data collected near the copper and zinc absorption edges. There is no evidence to suggest that this site contains more than one copper ion, and attempts to model additional copper ions were not consistent with anomalous Fourier maps. The copper ion is coordinated by the δ nitrogen atoms of His48 and His72 with a nearly linear geometry. Gln404 is also within 3 Å of the copper ion. The second copper site (**Figure 22(b)**) is dinuclear and also located in the N-terminal β-barrel of pmoB, approximately 21 Å away from the mononuclear site (**Figures 21 and 22**). The site is situated approximately 10 Å above the lipid bilayer
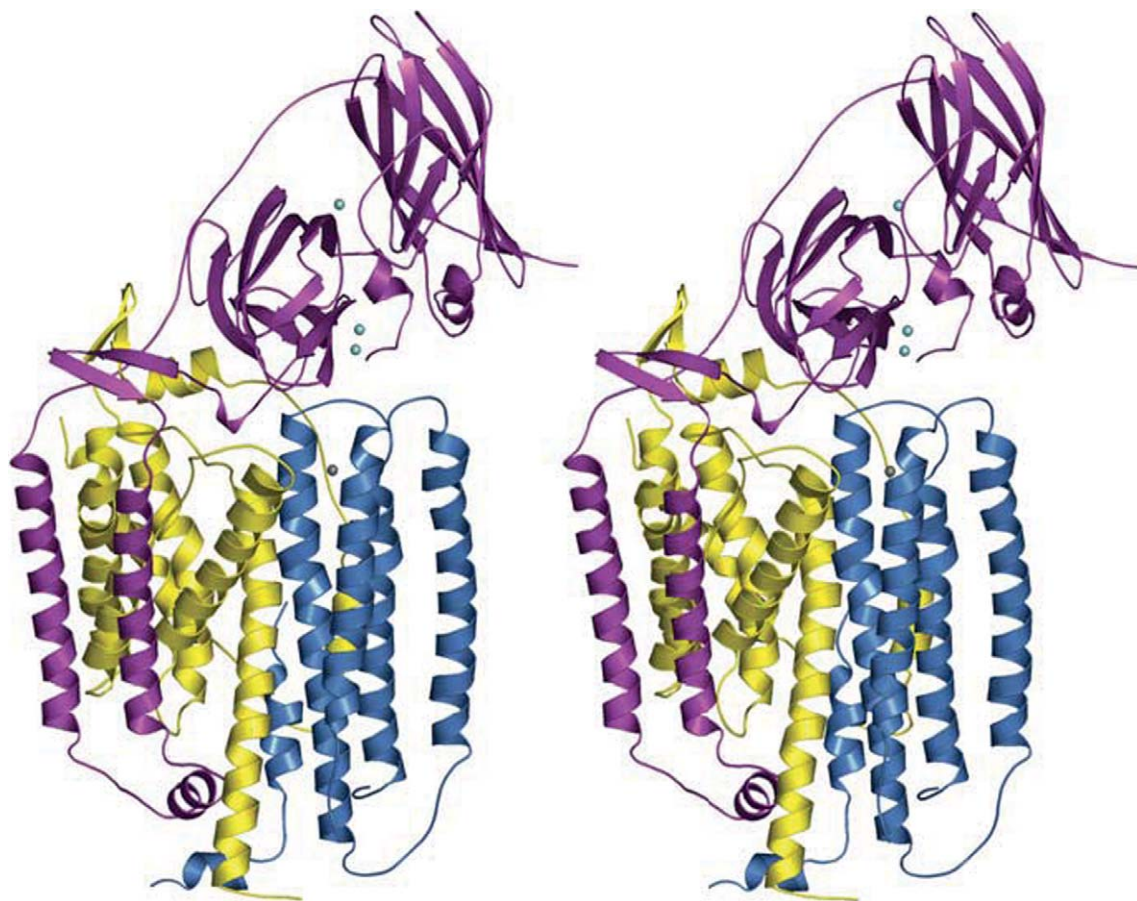
Figure 21 Stereo view of a single protomer of pMMO with pmoB shown in magenta, pmoA displayed in yellow, and pmoC shown in blue. Copper ions are shown as cyan spheres and zinc is depicted as a gray sphere. Reproduced from R. L. Lieberman; A. C. Rosenzweig, *Nature* **2005**, *434*, 177–182, with permission from Nature Publishing Group.

interface. Consistent with the 2.57 Å Cu–metal interaction determined by EXAFS,[175] the Cu–Cu distance in the dinuclear model refines to approximately 2.6 Å. One copper ion is coordinated by the N-terminal residue of pmoB, His33. Both the N-terminal amino nitrogen and the side chain δ nitrogen are within coordinating distance. The second copper ion is coordinated by the δ nitrogen of His137 and the ε nitrogen of His139 (**Figure 22(b)**). These residues are highly conserved in pMMO and AMO from a number of organisms. Residues His33 and His139 are held in position by hydrogen bonds to the side chain of Glu35 and the carbonyl oxygen of Gly152, respectively. Both of these residues are also conserved. Additional terminal or bridging ligands may be present, but are not observed in the 2.8 Å resolution electron density maps.

The third site, occupied by a Zn ion in the crystal, is situated approximately 13 Å below the surface of the membrane and is coordinated by Asp156, His160, and His173 from pmoC and Glu195 from pmoA (**Figure 22(c)**). The presence of Zn at this site is a crystallization artifact and the physiological metal ion is unknown. As the 2.8 Å resolution of the crystal structure determination does not allow to obtain accurate descriptions of the coordination geometry of the metal sites, pMMO has been reinvestigated by X-ray absorption spectroscopy to analyze the oxidation states and coordination environments of the metal centers in as-isolated (pMMO$_{iso}$), chemically reduced (pMMO$_{red}$), and chemically oxidized (pMMO$_{ox}$) samples.[177] X-ray absorption near-edge spectra (XANES) indicate that pMMO$_{iso}$ contains both Cu(I) and Cu(II) and that the pMMO centers undergo redox chemistry. EXAFS analysis reveals a Cu–Cu interaction in all redox forms of the enzyme. The Cu–Cu distance increases from 2.51 to 2.65 Å upon reduction, accompanied with an increase
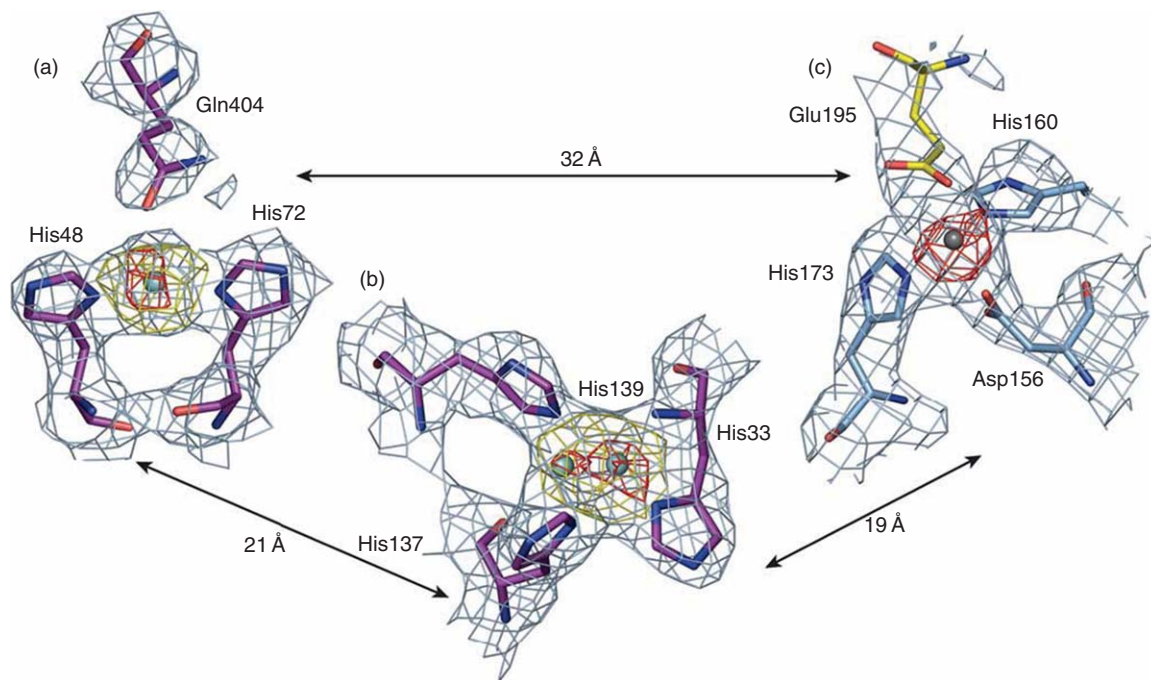
Figure 22 Metals centers of pMMO. The distances are measured between metal centers. Anomalous difference Fourier maps, yellow for Cu absorption edge and red for Zn absorption edge, are superimposed on the final 2Fo–Fc electron density map. (a) The mononuclear copper site. (b) The dinuclear copper site. (c) The zinc site. Reproduced from R. L. Lieberman; A. C. Rosenzweig, *Nature* **2005**, *434*, 177–182, with permission from Nature Publishing Group.

in the average Cu–O/N bond lengths. The findings are complementary to the crystallographic data and provide new insight into the oxidation states and possible electronic structures of the pMMO Cu ions.

In conclusion, the pMMO structure reveals an unexpected trimeric arrangement and the overall folds of the three subunits. Two of the three metal centers, modeled as mononuclear and dinuclear copper, are located within the soluble regions of the pmoB subunit. The third metal center, occupied by zinc in the crystal, lies within the membrane with ligands derived from both pmoC and pmoA. Direct ET between metal centers may be possible. Neither the site of methane oxidation nor the pathway(s) of substrate entry and product egress have yet been identified so far.

8.14.2.2.4   *Tyrosinase (EC 1.14.18.1) and catechol oxidase (EC 1.10.3.1)*

Tyrosinase, which belongs to a protein family having the catalytic center formed by dinuclear type-3 copper, catalyzes the orthohydroxylation of monophenol and the subsequent oxidation of the diphenolic product to the resulting quinone.[178] A series of reactions occurs under the concomitant reduction of molecular oxygen to water. The quinone product is a reactive precursor for the synthesis of melanin pigments. Tyrosinase, which is contained in vegetables, fruits, and mushrooms, is a key enzyme in the browning that occurs upon bruising or long-term storage. In mammals, the enzyme is responsible for skin pigmentation abnormalities, such as flecks and defects.[179] Thus, tyrosinase is quite significant in the fields of agriculture and industry. In the cosmetic industry, the development and screening of potent inhibitors of tyrosinase are especially attractive.

Tyrosinase is classified into the type-3 copper protein family, as are catechol oxidase and the respiratory pigment hemocyanin. During the catalytic reaction, the type-3 copper center of tyrosinase exists in three redox forms.[178] The deoxy form (Cu(I)–Cu(I)) is a reduced species, which binds oxygen to give the oxy form (Cu(II)–O_2^{2-}–Cu(II)). In the oxy form, molecular oxygen is bound as peroxide in a μ-η^2:η^2 side-on bridging mode, which destabilizes the O–O bond and activates it. The met form (Cu(II)–Cu(II)) is assumed as a resting enzymatic form, where Cu(II) ions are normally bridged to a small ligand, such as a water molecule or hydroxide ion.

Catechol oxidase oxidizes *ortho*-diphenols to the corresponding quinones but lacks monooxygenase or cresolase activity. Hemocyanin acts as an oxygen carrier in arthropods and mollusks.

The crystal structures of tyrosinase from *Streptomyces castaneoglobisporus* HUT 6202[68] and catechol oxidase from the sweet potato *Ipomoea batatas*[180] have been determined. They confirm that the coordination of the type-3 copper site in tyrosinase and catechol oxidase is very similar to that found in hemocyanin. This had been deduced before from the similarity of spectroscopic properties and a comparison of many tyrosinase and hemocyanin primary structures.[181–183] On the basis of the biological source of the proteins seven different domain organizations could be identified. Plant catechol oxidases of different organisms have a sequence identity of about 40–60%. The sequence identity between catechol oxidases and mulluscan hemocyanins is about 35% over almost the whole length of the sequences. In contrast, the sequence identity between plant catechol oxidases and other type-3 copper proteins from any nonplant source is limited to the two copper-binding regions.

The two copper-binding regions show the highest conservation throughout all type-3 copper proteins. Especially the region binding Cu_B is highly conserved, whereas the Cu_A-binding region shows more sequence variety and has been held responsible for the different functions of tyrosinase, catechol oxidase, and hemocyanin.

The overall structure of tyrosinase from *S. castaneoglobisporus* in complex with open reading frame ORF378 is displayed in **Figure 23**. Tyrosinase takes α-helical structures with the core of the enzyme, which is formed by a four-helix bundle. The catalytic dinuclear copper center is lodged in the helical bundle (**Figure 23**). Each of the two copper ions in an active site is coordinated by three His residues (**Figure 24**), which are derived from the four helices of the α-bundle except His54. One copper ion (designated Cu_A) is coordinated by His38, His54, and His63. His38 and His63 are located in the middle of $\alpha2$ and $\alpha3$, respectively. The second copper ion (Cu_B) is coordinated by His190, His194, and His216. The residues His190 and His194 are at the beginning and in the middle of $\alpha6$, respectively, and His216 is in the middle of $\alpha7$. This dicopper center is located at the bottom of the large concavity as a putative substrate-binding pocket, which is formed by the hydrophobic residues. In addition to the helical structure, tyrosinase has a few β-structures, as judged from the backbone torsion angles. In these, only the N- and C-terminal β-strands form a sheet structure.
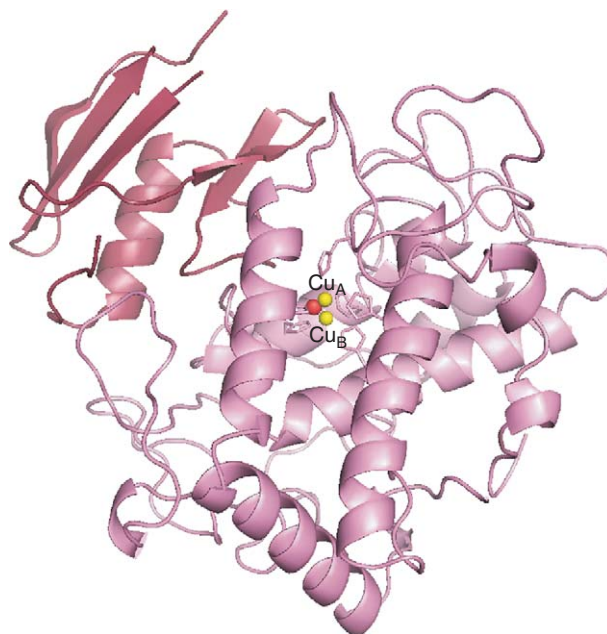
Figure 23   Ribbon diagram of tyrosinase from *Streptomyces castaneoglobisporus* in complex with ORF378 (PDB-code: 1WX3). Tyrosinase and ORF378 are shown in pink and raspberry, respectively. The copper ions Cu_A and Cu_B are depicted as yellow spheres; prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).
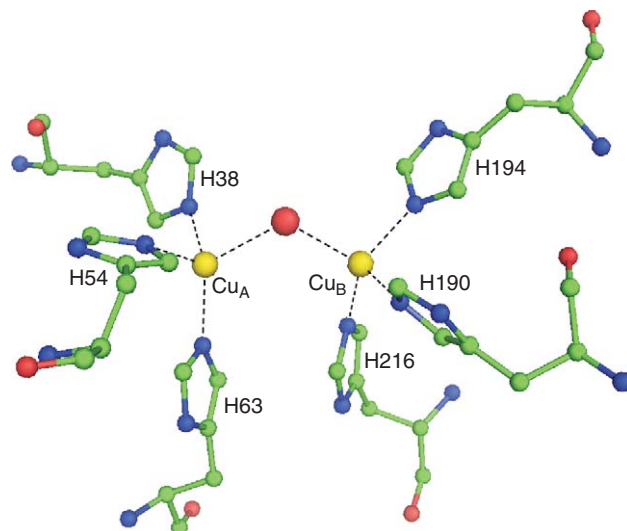
Figure 24   The met form of the active center of tyrosinase from *Streptomyces castaneoglobisporus* (PDB-code: 1WX3); prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

Although the amino acid sequence of tyrosinase has only 25.3 and 26.0% identities with those of the *I. batatas* catechol oxidase[180] and the odg domain of the *Octopus dofleini* hemocyanin,[184] respectively, its overall structure is quite similar to theirs. Among these three proteins, a high degree of conservation is observed in the core domain composed of the α-bundle. The tyrosinase and hemocyanins from *Panulirus interruputus*[185] and *L. polyphemus*[66] show no significant homology and no resemblance in their structures, but the catalytic core domains of these proteins are superimposable.

For tyrosinase from *S. castaneoglobisporus* five different states of the active site could be characterized in crystal structures, namely, copper-free form, met form I, met form II, deoxy, and oxy. In crystal structures of catechol oxidase from *I. batatas* the met and deoxy states as well as an inhibitor complex have been elucidated. In the met state (Cu(II), Cu(II)) the two cupric ions are at a distance of 2.9 Å, each of them being coordinated by three histidines. They are bridged by another atom, most likely a hydroxide ion, at a distance of about 1.8 Å from each cupric ion, so that each of them has a coordination number of 4 (see **Figure 24**, which shows the same situation for tyrosinase from *S. castaneoglobisporus*). In the *deoxy* or reduced state, both copper atoms are in the +1 oxidation state. The copper–copper distance is 4.4 Å. The coordination numbers are 4 for Cu_A (three histidine ligands and a coordinating water molecule) and 3 for Cu_B (three histidine ligands). The coordination sphere is distorted trigonal pyramidal for Cu_A and square planar for Cu_B (the coordination site occupied by the bridging OH^- in the met state is vacant). In the inhibitor complex with phenylthiourea (PTU), the copper–copper distance increases to 4.2 Å with the sulfur atom of PTU replacing the hydroxo bridge of the met state. The coordination spheres of the two coppers remain similar to those of the met state, but there are conformational changes at the active site residues. The most significant change is a rotation of the aromatic ring of Phe261 (catechol oxidase numbering).

Compared with the met state, the coordinating residues have only slightly different positions in the reduced state, indicating a rather rigid pocket. The changes in coordination are associated with movements of the copper atoms in the pocket. The inhibitor complex shows that Phe261 is located above the active site like a gate, which rotates after the inhibitor is bound. Thus, access of the substrate to the catalytic metal center seems to be controlled by this 'gate residue'.

The catalytic mechanism of tyrosinase was first studied in detail by Solomon *et al.*[178] Solomon proposed a mechanism for both the cresolase and catecholase activities of tyrosinase (**Figure 25**). This mechanism suggests the oxy state to be the starting point of cresolase activity (inner circle). This state is present in the resting form of tyrosinase in a proportion of about 15% (85% met state). A monophenol substrate binds to the oxy state and is monooxygenated to *o*-diphenol. This diphenol subsequently binds to the copper center of met tyrosinase in a
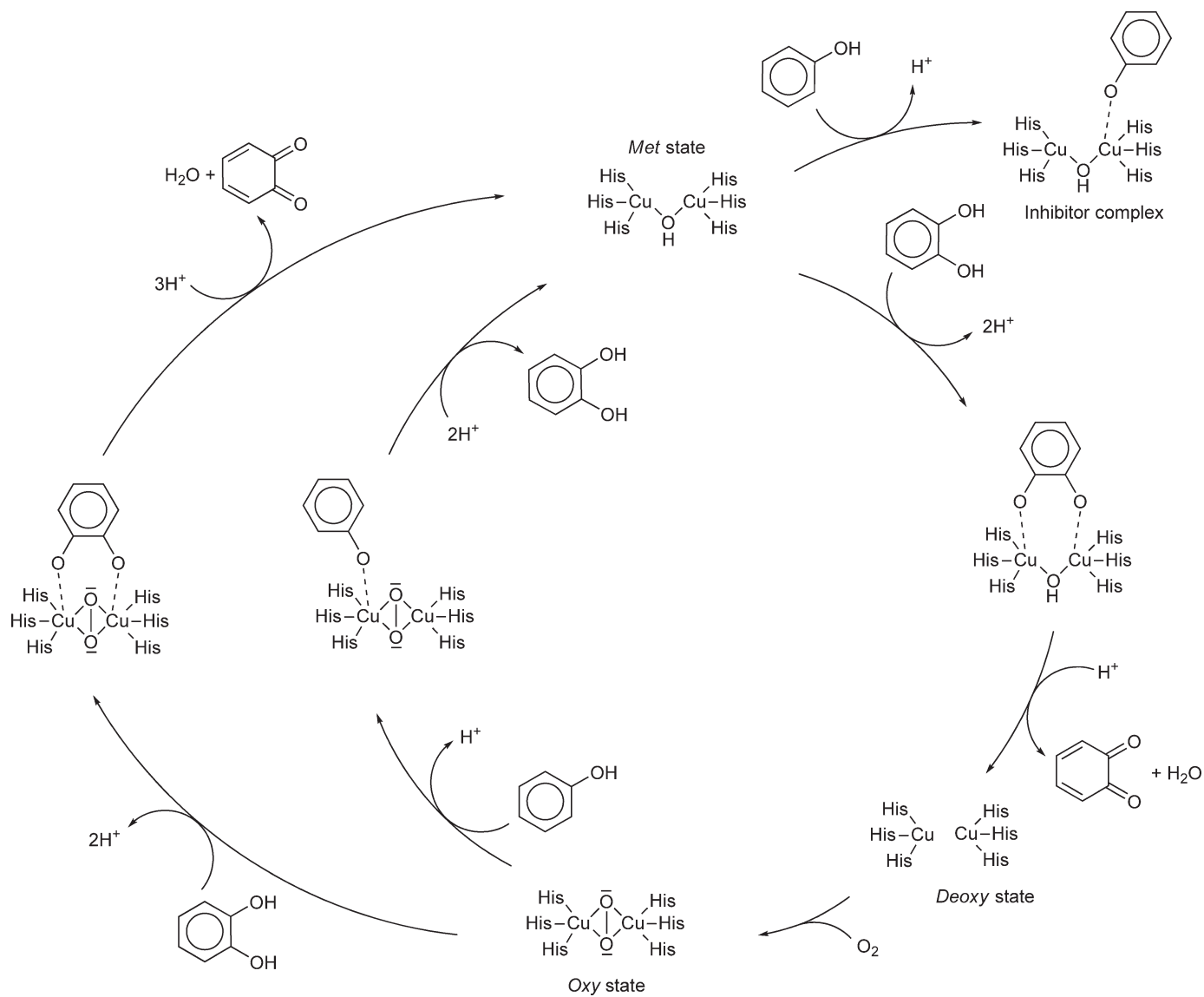
Figure 25 Mechanism of cresolase and catecholase activity of tyrosinase and catechol oxidase developed on the basis of an initial proposal by Solomon and coworkers[178] and including more recent results.[186,187] Reproduced from C. Gerdemann; C. Eicken; B. Krebs, *Acc. Chem. Res.* **2002**, *35*, 183–191, with permission from American Chemical Society.

bidentate binding mode proposed on the basis of a model compound.[188] Oxidation of the diphenol substrate leads to the reduced state of the dinuclear copper center. Reoxidation of the reduced state to the oxy state occurs by attack of dioxygen and closes the catalytic cycle.

The mechanism of catecholase activity (outer circle) starts from the oxy and met states. A diphenol substrate binds to the met state (for example), followed by the oxidation of the substrate to the first quinone and the formation of the reduced state of the enzyme. Binding of dioxygen leads to the oxy state, which is subsequently attacked by the second diphenol molecule. Oxidation to the second quinone forms the met state again and closes the catalytic cycle.

Alternative reaction mechanisms include a radical mechanism proposed by Kitajima and Morooka[189] and a mechanism involving a Cu(III) intermediate based on measurements of model compounds.[190] On the basis of the crystal structure of the catechol oxidase–PTU inhibitor complex, monodentate binding of the substrate was suggested for catechol oxidase.[180] A radical mechanism, as proposed for the weak catecholase activity found in *Octopus vulgaris* hemocyanin,[191] is also possible for catechol oxidase due to the strong structural relationship between catechol oxidase from *I. batatas* and odg hemocyanin as described above.

The distinct difference between catechol oxidase and tyrosinase has not yet been explained. A lag phase in the monophenolase activity of tyrosinase has been found and studied and is proposed to be a result of temporary inhibition of the met state of tyrosinase by excess of the monophenol substrate (**Figure 25**).[186] Monophenolase activity increases when the diphenol product displaces the monophenol from met tyrosinase and allows the continuation of the catalytic cycle. Catechol oxidase in its isolated form is present exclusively in the met state and is also inhibited by phenol. It was therefore suggested that lack of the oxy state is the reason catechol oxidase lacks cresolase activity. As oxy catechol oxidase also shows no monooxygenase activity, this explanation does not seem entirely satisfying. Another possible reason is that access to Cu_A, which has been proposed to be necessary for the oxygenation of monophenols,[192] is blocked in the crystal structure of catechol oxidase from *I. batatas*.

8.14.2.3   Dioxygenases

8.14.2.3.1   *Quercetin 2,3-dioxygenase (EC 1.13.11.24)*

There is one copper enzyme known, quercetin 2,3-dioxygenase (2,3QD), that acts as a dioxygenase.[193] Dioxygenases incorporate both oxygen atoms of dioxygen into the substrate. Quercetin 2,3-dioxygenase is a type-2 copper-dependent enzyme[194] expressed by *Aspergillus* species when grown on complex aromatic such as rutin or quercetin. It is able to disrupt the O-heteroaromatic ring of flavonols, yielding the corresponding depside (phenolic carboxylic acid ester) and carbon monoxide (**Figure 26**).[195] The difficult breakage of two carbon–carbon bonds makes this reaction chemically challenging. Studies on the 2,3QDs from *Aspergillus flavus*[193] and *Aspergillus niger* DSM 821[194] have shown that the enzyme does not require any additional organic cofactors for catalysis. The *A. flavus* 2,3QD has a molecular mass of 111 kDa, with a sugar content of 27.5%, and a metal content of 2 mol of copper per mol of enzyme.[193] The 2,3QD isolated from *A. niger* DSM 821 is composed of three different subunits, with molecular masses of 63–67, 53–57, and 31–35 kDa, respectively. It has a carbohydrate content of 46–54% and contains 1.0–1.6 mol of copper per mol of enzyme.[194] No information is available on the amino acid sequences of these latter two enzymes, besides their biochemical characterization.

Recently, the crystal structure of 2,3QD from *Aspergillus japonicus* has been determined at a resolution of 1.6 Å.[196] This enzyme is a glycoprotein with a molecular mass of 50 kDa. It is composed of 350 amino acid residues and contains as predicted a single copper ion per monomer. In the crystal structure, the asymmetric unit holds four monomers, which are arranged as two homodimers. At functionally relevant pH values (pH 5.0–7.0), 2,3QD is a homodimer of about 100 kDa with approximately 25% (w/w) of N-linked glycan chains. The monomer of 2,3QD is the enzymatically active unit. It has approximate dimensions of $30 \times 45 \times 50$ Å and is composed of two structurally similar domains positions face to face around a pseudo two-fold symmetry axis (**Figure 27**). The two domains, which are joined by a linker of 60 amino acid residues (residues 1–205), can be superimposed with a rms difference of 1.6 Å for 120 Cα atoms. The N-terminal domain (residues 1–145) shares about 20% sequence identity with the C-terminal domain (206–350). Apart from a salt bridge the interactions are mainly hydrophobic. Each domain is built up by two antiparallel β-sheets, with eight strands forming a
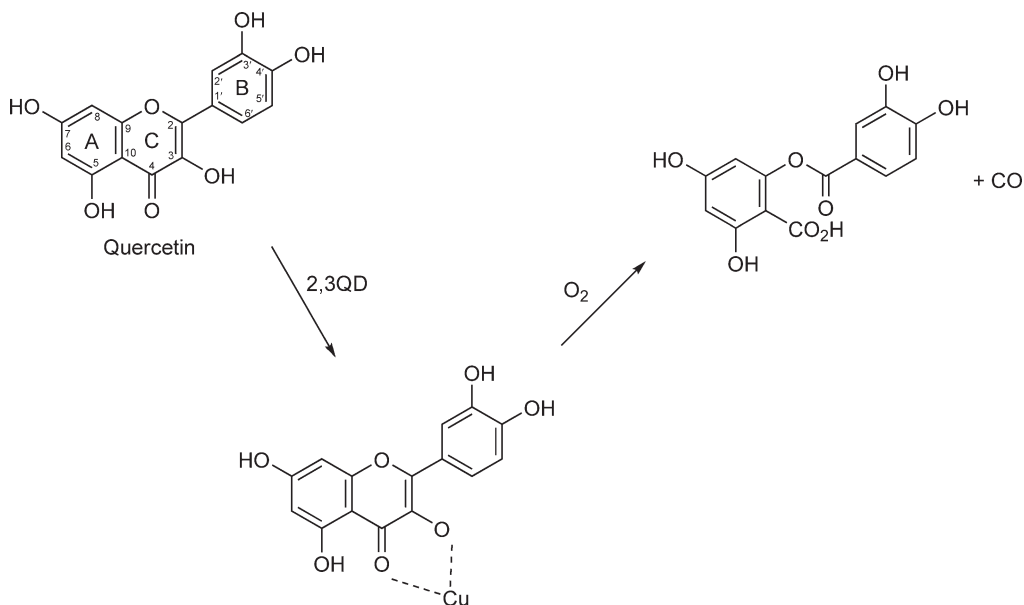
Figure 26 Reaction scheme for quercetin 2,3-dioxygenase. The proposed substrate–Cu complex is shown with quercetin (5,7,3′,4′-tetrahydroxyflavonol). Reproduced from F. Fusetti; K. H. Schroter; R. A. Steiner; P. I. van Noort; T. Pijning; H. J. Rozeboom; K. H. Kalk; M. R. Egmond; B. W. Dijkstra, *Structure* **2002**, *10*, 259–268, with permission from Elsevier Inc.


Figure 27 Ribbon diagram of the monomer of quercetin 2,3-dioxygenase. The N-terminal domain is shown in yellow and the C-terminal domain in brown, and α-helices are displayed in blue. The linker connecting both domains is colored red. The copper is shown as a green sphere and the copper coordinating residues are represented as ball and stick models. Reproduced from F. Fusetti; K. H. Schroter; R. A. Steiner; P. I. van Noort; T. Pijning; H. J. Rozeboom; K. H. Kalk; M. R. Egmond; B. W. Dijkstra, *Structure* **2002**, *10*, 259–268, with permission from Elsevier Inc.

β-sandwich and two short α-helices. In the N-terminal domain, there is an additional β-strand, which builds up part of the catalytic site. In both domains, similar hydrophobic cavities are present. In the N-terminal domain, this cavity represents the catalytic site and contains the mononuclear copper center, located at about 10 Å from the protein surface.

Based on its amino acid sequence and three-dimensional structure, 2,3QD can be classified within the cupin superfamiliy.[197] This superfamiliy includes functionally diverse proteins that are found in archaea, eubacteria, and eukaryota. Structural information shows that they contain a motif of six antiparallel β-strands located within a conserved β-barrel structure.[198] The structure of gemin, which is a 16 kDa Mn-containing oxalate oxidase,[198] can be superimposed on the N-terminal domain of 2,3QD with an rms deviation of 1.8 Å for 91 Cα atoms. One can see in the superposition that the Mn site of germin formed by three histidines, a glutamate, and two water molecules matches with the copper site of 2,3QD.

The copper site is shown in **Figure 28**. The metal center is solvent exposed. The copper ion is mainly coordinated by three histidine residues (His66, His68, and His112) and by a water molecule. The geometry is distorted tetrahedral. There is also an additional coordination in which the three histidines and Glu73 interact with metal and the water molecule is positioned further away from the copper (see the electron density for an alternate side chain conformation of Glu73 in **Figure 28**, which then coordinates one of the carboxylate oxygens of the side chain to the copper). The geometry of this secondary coordination, estimated to present at most 30% of the structure, is trigonal bipyramidal with His66 and Glu73 as the axial ligands. In both coordination arrangements, the Nϵ2 atoms of the histidine side chains are at a distance of about 2.1 Å from the copper(II) ion and coordinate the metal similarly to what has been found in other type-2 copper sites.[199]

2,3QD catalyzes the dioxygenation of quercetin (5,7,3′,4′-tetrahydroxyflavonol) and other flavonoids, resulting in the fixation of two oxygen atoms into the substrate accompanied by a ring-opening step (**Figure 26**). The reactivity of the substrates is greatly influenced by the distribution of the hydroxyl substituents.[193] For example, the absence of C3′–OH doubles the reaction rate, whereas lack of C7–OH and C4′–OH has a drastic negative effect.[193] The C3–hydroxyl and the C4–carbonyl groups as well as the presence of a double bond between C2 and C3 are essential for catalysis.[193,195,200] Since the earliest biochemical characterization of a 2,3QD,[193] it was proposed that the substrates could bind in the active site, chelating the
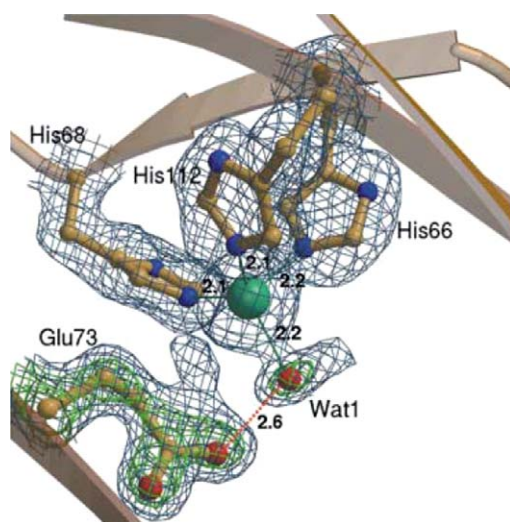
Figure 28   Copper coordination in quercetin 2,3-dioxygenase. The tetrahedral geometry is shown in which three histidine residues (His66, His68, and His112) and one water molecule (Wat1) are coordinated to the copper. The trigonal bipyramidal geometry in which Glu73 is directly participating in the copper coordination is described in the text. The 2Fo–Fc electron density map is contoured at 1σ (blue) and 2σ (green). Reproduced from F. Fusetti; K. H. Schroter; R. A. Steiner; P. I. van Noort; T. Pijning; H. J. Rozeboom; K. H. Kalk; M. R. Egmond; B. W. Dijkstra, *Structure* **2002**, *10*, 259–268, with permission from Elsevier Inc.
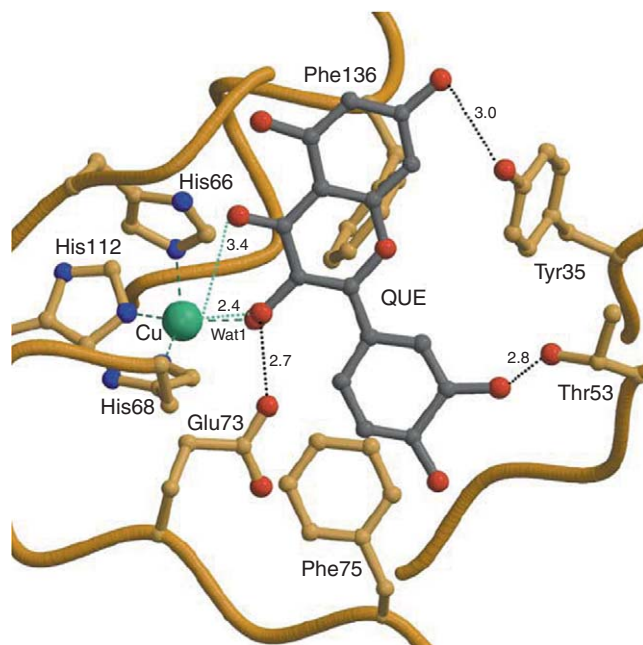
Figure 29 Diagram of the quercetin 2,3-dioxygenase active site with manually docked substrate quercetin (QUE, gray). The copper-binding site and some surrounding side chains are shown as ball and stick diagrams. Reproduced from F. Fusetti; K. H. Schroter; R. A. Steiner; P. I. van Noort; T. Pijning; H. J. Rozeboom; K. H. Kalk; M. R. Egmond; B. W. Dijkstra, *Structure* **2002**, *10*, 259–268, with permission from Elsevier Inc.

copper with the 3-hydroxyl and 4-carbonyl groups. This hypothesis is supported by later biomimetic studies and by the evidence that substrates are able to protect the enzyme from inactivation by metal chelators.[195]

Based on this knowledge, Fusetti *et al.*[196] have modeled a substrate molecule in the catalytic site of 2,3QD (**Figure 29**). Quercetin possesses five hydroxyl groups that probably replace part of the water structure upon binding in the active site. The manual docking was guided by the position of solvent molecules present in the active site of four crystallographically independent molecules. The orientation of the quercetin molecule was adjusted to maximize the number of favorable protein–substrate interactions.

The model shows that the molecule can bind to the copper in a monodentate manner, without requiring major conformational changes of the protein or the substrate structure. Tyr35 and Thr53 are in a favorable position to interact with C7–OH and C3′–OH, in agreement with an enhancing effect of these groups on the oxygenation rate. Substitution at C8 decreases activity; the model shows that there is not enough space for any atom other than hydrogen at this position. Several aromatic side chains, including Tyr35, Phe75, and Phe136, surround the copper center and might stabilize substrate binding. The B ring of quercetin is positioned in a cavity adjacent to the copper site and points toward Gly125. Replacement with any other amino acid would be incompatible with the proposed substrate-binding mode. A comparison between the N- and C-terminal domains shows that Gly125 is replaced by Phe322 in the C-terminal domain. In the model, the B ring of the quercetin occupies the position corresponding to that of the aromatic side chain of Phe322.

Studies with nonenzymatic model systems have suggested that the oxygenolysis of quercetin could be base catalyzed, with substrate activation achieved by deprotonation of the C3–OH group (**Figure 26**). As a result, negative charge develops at the C2 atom, activating the substrate toward electrophilic attack by O_2.[195,200,201] If the enzyme-catalyzed reaction proceeds in a similar way, an amino acid base is expected to assist in substrate deprotonation. In 2,3QD, Glu73 is the only residue in a favorable position to function as the active site base that abstracts the proton from the reactive C3–OH group of the substrate. Site-directed mutagenesis studies showed that a Glu73Gln mutation resulted in an EPR-active enzyme, with more than 1000-fold decreased activity, confirming that this residue is important.[196] Thus, the structural results together with the absence of O_2 bound to the Cu^{2+} ion are consistent with a catalytic pathway in which the substrate is bound to the metal center, and

subsequent proton abstraction, promoted by Glu73, activates the complex for electrophilic attack by O_2. However, an alternative mechanism in which Glu73 modulates the redox potential of the copper and, in this way, may affect the substrate susceptibility to oxygen cannot be excluded at this stage.

8.14.2.4 Superoxide-Decomposing Enzymes

8.14.2.4.1 Cu,Zn-superoxide dismutase (EC 1.15.1.1)

The intracellular form of Cu,Zn-SOD (SOD1) is predominantly found in the cytosol and peroxisomes of eukaryotes,[202] but also in the nucleus and mitochondrial intermembrane space, and protects these cells from the toxic effects of the superoxide radical produced as a by-product of aerobic metabolism.[203] The protective effects of the enzyme appear to be of considerable biological importance. SOD1 seems to be a possible therapeutic agent in diseases that cause a pathological production of oxygen-derived free radicals like all diseases that involve the inflammatory process or impairment of blood flow.[204] However, when mutated, it can also cause disease. Over 100 different mutations have been identified in the *sod1* genes of patients diagnosed with the familiar form of amyotrophic lateral sclerosis (fALS).[205] These mutations result in a highly diverse group of mutant proteins, some of them very similar to and others enormously different from wild-type SOD1. Despite their differences in properties, each member of this diverse set of mutant proteins causes the same clinical disease, presenting a challenge in formulating hypotheses as to what causes SOD1-associated fALS.

Intracellular SOD1 is a homodimeric enzyme of about 16 kDa for one monomer, which contains a copper and a zinc ion per identical subunit. The physiological role of SOD1 is the dismutation of superoxide radicals to molecular oxygen and hydrogen peroxide (McCord and Fridovich[206]), which is then scavenged by other enzymes such as catalase. The catalytic mechanism occurs through a two-step process in which the copper ion is alternatively reduced, with the formation of molecular oxygen, and oxidized back to the resting state with production of hydrogen peroxide.[206,207]

$$SOD1-Cu^{2+} + O_2 \rightarrow SOD1-Cu^+ + O_2 \tag{9}$$

$$SOD1-Cu^+ + O_2 + 2H^+ \rightarrow SOD1-Cu^{2+} + H_2O_2 \tag{10}$$

The observed catalytic rates for this enzyme are very fast, of the order of $2-3 \times 10^9 \, s^{-1} \, mol^{-1} \, l$ and are diffusion controlled.[207]

SOD1 appears as a homodimer in the crystal structure (**Figure 30**).[208] The active site of SOD1 is located at the end of a wide channel bearing numerous charged residues, which can have a relevant role in the catalytic mechanism. It is embedded in a flat β-barrel domain consisting of eight antiparallel β-strands and three external loops. The copper ion in SOD1 is coordinated by four histidines in a distorted square planar geometry, His46 (Nδ1), His48 (Nε2), His63(Nε2), and His120 (Nε2). The zinc ion is coordinated in a distorted tetrahedral arrangement by three histidines and an aspartate, His63 (Nδ1), His71 (Nδ1), His80 (Nδ1), and Asp83 (Oδ1) as first derived from the crystal structure of bovine erythrocyte SOD[61,62] and here shown for human SOD1 (**Figure 31**).[208] The bridging ligand His63 is on the direct line between the copper and zinc atom in a 6 Å distance. A water molecule is located close to the copper ion with a Cu–O distance of 2.7 Å. The copper ion serves as redox partner for the superoxide radical whereas the zinc atom is redox inactive. The zinc atom appears to be not involved in catalysis but seems to have a structurally stabilizing function. Among the residues that form the active site channel Arg143 is the most relevant in terms of catalytic behavior, which is invariant among the SOD enzymes for which the primary structure has been determined. When Arg143, which is positively charged at physiological pH, is substituted by a neutral group or more, by a negative group, the catalytic rates of SOD decrease by one to two orders of magnitude, without any major perturbance of the metal ion ligands.[209,210] A plausible reaction scheme involving Arg143 (Arg141 in bovine SOD) is depicted in **Figure 32** (see Tainer *et al.*[211]). The superoxide molecule replaces the axial water molecule of Cu^{2+} and reduces this to Cu^+. The $Cu–O_2$ complex gets protonated by Arg141 and O_2 is released. The ligand bond between His61 and copper is broken. A second superoxide molecule binds to the copper and this becomes oxidized to Cu^{2+}. A further proton completes the formation of H_2O_2. The Cu–His61 bond is formed again after deprotonation of His61 and peroxide is released.

The reduced form of bovine SOD has been characterized by X-ray crystallography and NMR (Banci *et al.*[212]). In the X-ray structure the imidazolate bridge is maintained but in solution the bridge is broken
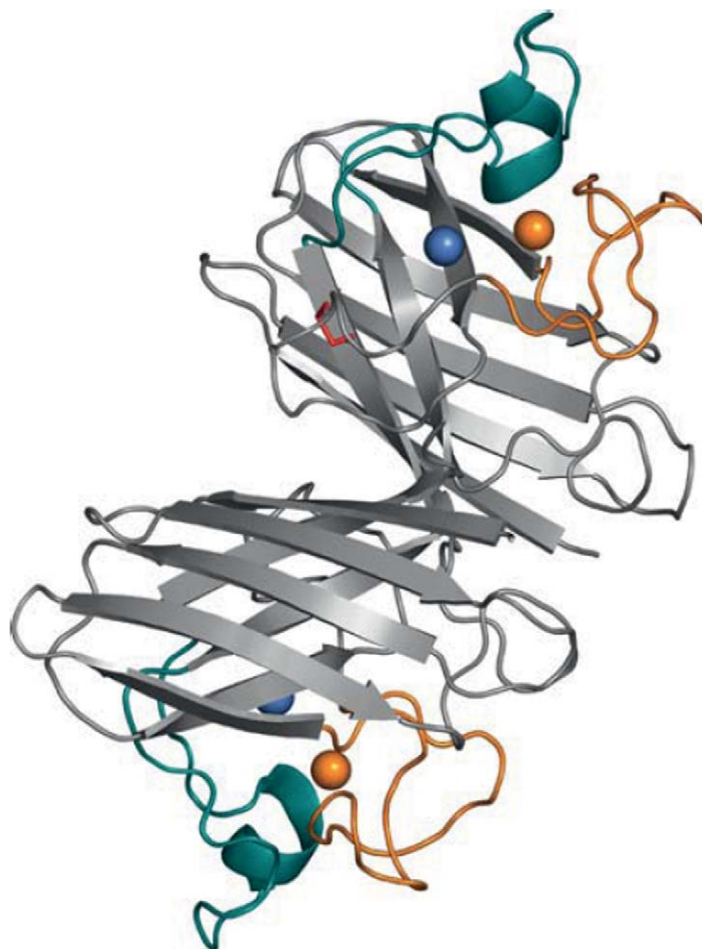
Figure 30 Ribbon diagram of the metal-bound dimeric human SOD1.[208] Copper and zinc ions are shown as blue and orange spheres, respectively. The zinc loop is depicted in orange and the electrostatic loop in teal. Reproduced from J. S. Valentine; P. A. Doucette; S. Z. Potter, *Annu. Rev. Biochem.* **2005**, *74*, 563–593, with permission from Annual Reviews.
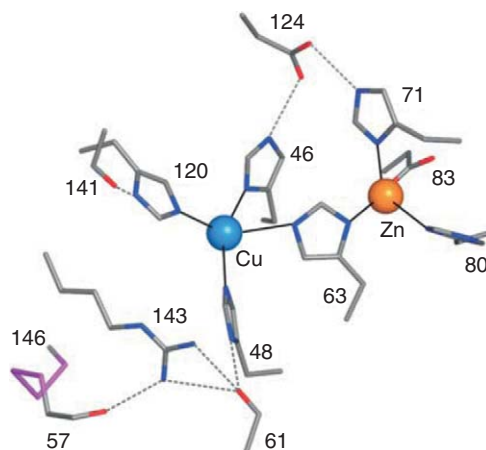
Figure 31 Copper–zinc coordination in human SOD1.[208] Reproduced J. S. Valentine; P. A. Doucette; S. Z. Potter, *Annu. Rev. Biochem.* **2005**, *74*, 563–593, with permission from Annual Reviews.
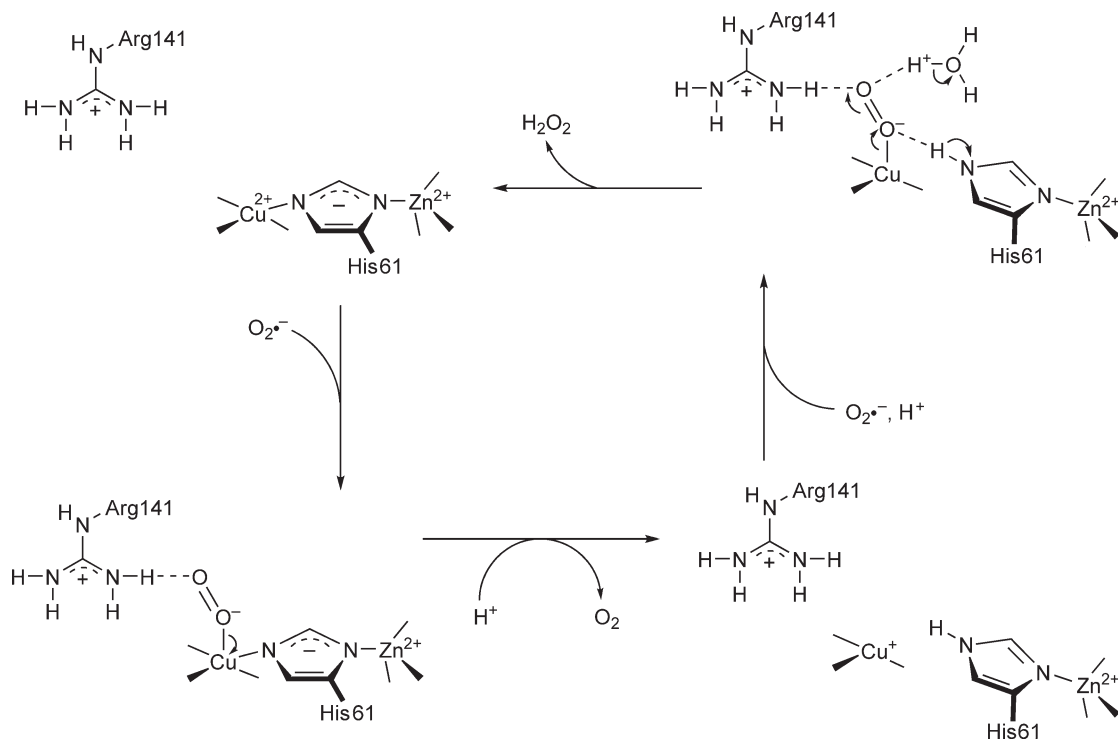
Figure 32 Scheme of the catalytic cycle of Cu,Zn-SOD. Reproduced from J. A. Tainer; E. D. Getzoff; J. S. Richardson; D. C. Richardson, *Nature* **1983**, *306*, 284–287, with permission from Nature Publishing Group.

and the involved histidine is protonated on the side of copper as assumed in the scheme of **Figure 32**. Mutants of human SOD1 have been designed, and functionally and X-ray crystallographically characterized.[213] The mutations were introduced to enhance the electrostatic guidance of the substrate and the corresponding mutants really were significantly more active than the diffusion-limited wild-type enzyme.

8.14.2.5 Four-Electron Reducing Oxidases

The blue oxidases like ascorbate oxidase, laccase, and ceruloplasmin, and the terminal oxidases of aerobic respiratory chains like cytochrome oxidases and quinol oxidases are the only enzymes so far known that catalyze the direct four-electron reduction of molecular oxygen to water. Thereby, the reducing substrates like ascorbate, quinol, Fe^{2+}, and cytochrome c are oxidized in one-electron transfer steps. The substrates of quinol oxidases, ubiquinol, or menaquinol, may be oxidized in two-electron transfer steps. For the two cases the following general reaction formulae can be defined:

$$4S^{n+} + O_2 + 4H^+ \rightarrow 4S^{(n+1)+} + 2H_2O \qquad (11)$$

$$2S^{n+} + O_2 + 4H^+ \rightarrow 2S^{(n+2)+} + 2H_2O \qquad (12)$$

The blue oxidases are soluble extracellular enzymes (see, e.g., Messerschmidt[39]) whereas the terminal respiratory oxidases are membrane bound and use the free energy available from this reaction to pump protons across the membrane. The transmembrane proton and voltage gradient generated by the oxidase and other components of the aerobic respiratory chain is converted directly to more useful forms by a number of membrane-bound energy-conserving systems, such as the ATP synthase and secondary active transport systems (see, e.g., Calhoun *et al.*[214]).

8.14.2.5.1 Multicopper oxidases containing a type-1 copper center and a trinuclear copper site

Laccase, ascorbate oxidase, and ceruloplasmin are the classical members of the multicopper oxidase family also known as blue oxidases. Recently, a small number of bacterial members of this family have been characterized, including CueO from *E. coli*,[215] a spore-coat laccase (CotA) from *Bacillus subtilis*,[216] and phenoxazinone synthase from *Streptomyces antibioticus*.[217] The catalyzed reaction of these enzymes except for phenoxazinone synthase is given in Equation (11). A comprehensive overview of the broad and active research on blue copper oxidases is presented in Messerschmidt.[218] Recent results have been included in a review on the reduction of dioxygen by copper-containing enzymes.[219] The nature and number of the different copper sites in blue oxidases has been described in the sections about the type-1 copper site and the trinuclear copper cluster.

8.14.2.5.1(i) Ascorbate oxidase (EC 1.10.3.3)

Ascorbate oxidase was the first well-structurally characterized blue oxidase. Crystal structures of oxidized native,[70] type-2 depleted,[220] reduced, peroxide, and azide forms[221] of ascorbate oxidase from zucchini squash have been determined. Ascorbate oxidase is found in higher plants. The immonuhistochemical localization of ascorbate oxidase in green zucchini reveals that ascorbate oxidase is distributed ubiquitously over vegetative and reproductive organs in all specimens examined.[222] The *in vivo* role of ascorbate oxidase is still under debate. It might be involved in processes such as fruit ripening, growth promotion, or in susceptibility to disease (for more details see Avigliano and Finazzi-Agro[223]). The best and probably the physiological substrate is ascorbate but catechols and polyphenols are also substrates *in vitro*.[224]

Ascorbate oxidase is a homodimeric enzyme with a molecular mass of 70 kDa and 552 amino acid residues per subunit (zucchini). The three-domain structure and the location of the type-1 and trinuclear copper centers in the ascorbate oxidase monomer as derived from the crystal structure are shown in **Figure 33**. The folding of all three domains is of a similar β-barrel type. The mononuclear copper site is located in domain 3 and the trinuclear copper species is bound between domains 1 and 3. A binding pocket for the reducing substrate that is complementary to an ascorbate molecule is located near the type-1 copper site and accessible from solvent. A broad channel providing access from solvent to the trinuclear copper species, which is the binding and reaction site for the dioxygen is present in ascorbate oxidase. During catalysis an intramolecular electron transfer between the type-1 copper and the trinuclear copper cluster must occur. The region of the molecule between the two centers is depicted in **Figure 34**. The distances between the type-1 copper and the three coppers of the trinuclear center are 12.20, 12.69, and 14.87 Å, respectively. Intramolecular electron transfer from the type-1 copper to the type-3 copper pair of the
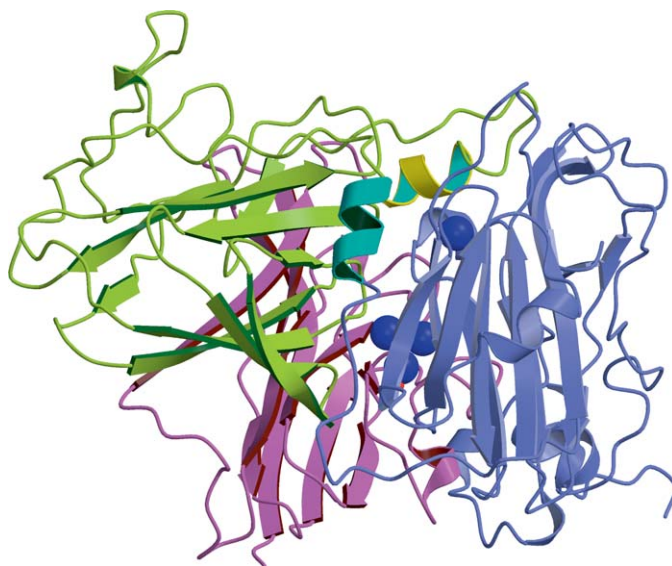
Figure 33 Ribbon diagram of the monomer structure of ascorbate oxidase (PDB-code: 1AOZ); prepared with MOLSCRIPT (P. J. Kraulis, *J. Appl. Crystallogr.* **1991**, *24*, 946–950) and RASTER3D (E. A. Merritt; M. E. P. Murphy, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 869–873).
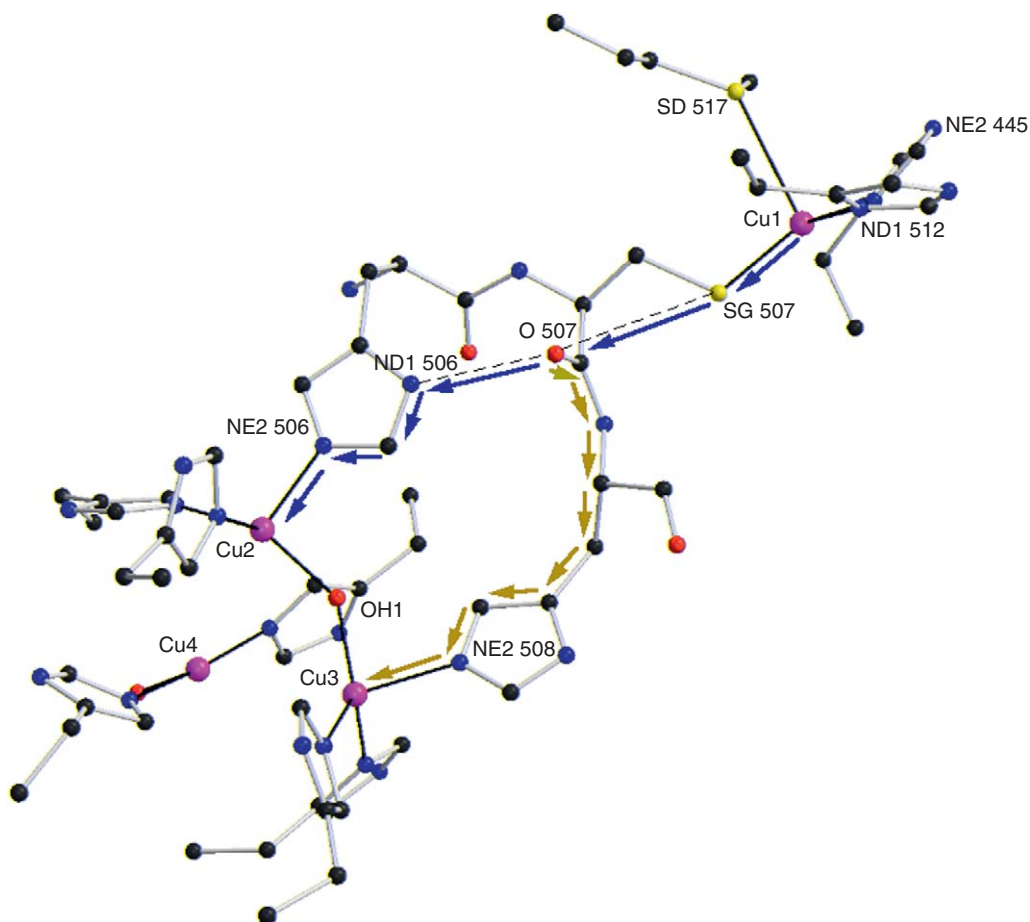
Figure 34 Drawing of the region in the ascorbate oxidase molecule between the type-1 copper center and the trinuclear copper center (PDB-code: 1AOZ); prepared with MOLSCRIPT (P. J. Kraulis, *J. Appl. Crystallogr.* **1991**, *24*, 946–950) and RASTER3D (E. A. Merritt; M. E. P. Murphy, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 869–873).

trinuclear copper site may be through-bond, through-space, or a combination of both. A through-bond pathway is available for both branches each with 11 bonds (see **Figure 34**). The alternative combined through-bond and through-space pathway from the type-1 copper Cu1 to Cu2 of the trinuclear center involves a transfer from the SG atom of Cys507 to the main-chain carbonyl of Cys507 and through the hydrogen bond of this carbonyl to the ND1 atom of His506. Electron transfer processes in blue oxidases have been discussed in detail.[225]

The crystal structure of the reduced form of ascorbate oxidase shows the type-1 copper site geometry virtually unchanged whereas the trinuclear site displays considerable structural changes. The bridging oxygen ligand OH1 is released and the two coppers, Cu2 and Cu3, move toward their respective histidines and become three-coordinated in a trigonal-planar arrangement. The copper–copper distances increase from an average of 3.7 to 5.1 Å for Cu2–Cu3, 4.4 Å for Cu2–Cu4, and 4.1 Å for Cu3–Cu4. In the crystal structure of the peroxide form, the bridging oxygen ligand OH1 is released as well and the peroxide is bound end-on to the copper Cu2 of the trinuclear copper cluster. Solomon and coworkers (see, e.g., Solomon *et al.*[226]) concluded from their spectroscopic data obtained from ascorbate oxidase and laccase that their reoxidation intermediate binds as 1,1-μ hydroperoxide between either Cu2 and Cu4 or Cu3 and Cu4.

A 'two-site ping pong bi bi' mechanism has been deduced for tree laccase from steady-state kinetics.[227] This will be valid for ascorbate oxidase as well because both enzymes are structurally and mechanistically closely related. A reaction scheme for ascorbate oxidase has been proposed based on the available spectroscopic, kinetic, and structural information (**Figure 35**) (Messerschmidt[39]) that should also be valid for laccase or
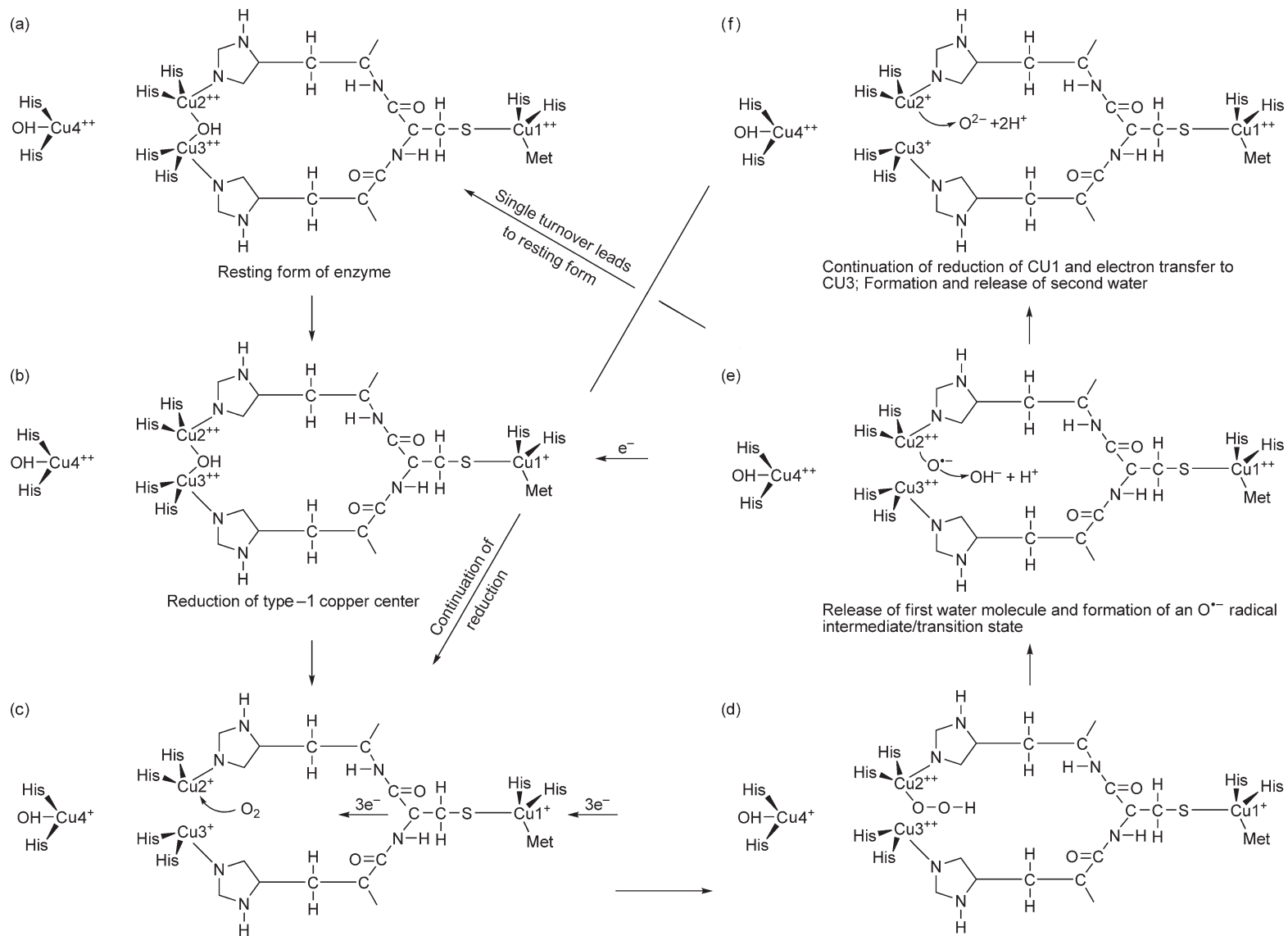
Figure 35 Proposal for the catalytic mechanism of ascorbate oxidase.[39]

ceruloplasmin in the main features. The catalytic cycle starts from the resting form (**Figure 35(a)**), in which all four coppers are oxidized and Cu2 and Cu3 are bridged by an OH⁻ ligand. The first step is the reduction of the type-1 copper Cu1 by the reducing substrate in a one-electron transfer step (**Figure 35(b)**). When ascorbate serves as reducing substrate, the formed semihydroascorbate dismutes to ascorbate and dehydroascorbate. The electrons are transferred through the protein to either Cu2 or Cu3. The fully reduced enzyme requires four electrons to be transferred (**Figure 35(c)**). At this stage, dioxygen may bind to the enzyme at Cu2, probably in the manner shown in the peroxide derivative. Other binding modes of dioxygen species to the trinuclear copper center have been found in crystal structures of CotA[228] and fungal laccases from *Melanocarpus albomyces*[74] and *Trametes versicolor*.[73] They will be discussed in Section 8.14.2.5.1(ii). A transfer of two electrons from the copper pair to dioxygen leads to the formation of a hydroperoxide intermediate (**Figure 35(d)**). A third electron may be transferred from Cu4 to the hydroperoxide intermediate and a fourth electron from the type-1 copper to copper ion Cu2. The O–O bond is broken at this stage and the first water molecule released (**Figure 35(e)**). An oxygen radical has been detected in laccase by EPR. The EPR spectrum indicates that the type-1 copper has been reoxidized and that the EPR signals of the oxygen radical intermediate and type-1 copper are present.[229] The Cu2 is in the reduced state and may facilitate O–O bond breakage and release of water. The catalytic cycle is continued by a further reduction of the type-1 copper center by the reducing substrate. This electron may be transferred to Cu3 of the copper pair. Now the fourth electron may be transferred to the oxygen radical intermediate from copper atom Cu2 and the second water molecule released (**Figure 35(f)**). In the case of only four electron equivalents, the reaction may lead to the resting form and the second water will remain bound as the bridging ligand between Cu2 and Cu3, concomitant with a substantial rearrangement within the trinuclear copper site. If turnover is continued, this will not occur and the trinuclear site will maintain a structure very close to that found in the fully reduced form.

8.14.2.5.1(ii)　Laccase (EC 1.10.3.2)　　Laccase is widely distributed in plants and fungi. Laccase from higher plants, found in various species of the Chinese, Vietnamese, and Japanese lacquer trees has been extensively investigated.[230] The biological function of laccase in these trees is well understood. The laccase of the lacquer trees is found in white latex, which contains phenols. After injury of the tree, these are oxidized by dioxygen to radicals, which spontaneously polymerize, building a protective structure that closes the wound. Laccase is very abundant in fungi. The biological function of fungal laccase has been postulated as playing a role in sporulation, pigment production, lignin degration, and pathogenesis (see, e.g., Smith *et al.*[231]). Primary structures of a tree laccase (*Acer pseudoplatanus*) and numerous fungal laccases have been determined. An amino acid sequence alignment of available laccase sequences with ascorbate oxidase, ceruloplasmin, and blue oxidases-related proteins (see, e.g., Messerschmidt[44]) show that laccases are monomeric with an average polypeptide chain length of about 550 residues. The apparent molecular mass may be up to 140 kDa (tree laccase) due to different carbohydrate contents of the enzymes. All copper ligands as derived from the crystal structures of ascorbate oxidase[70] and ceruloplasmin[42] are conserved except the methionine ligand of the type-1 copper. The methionine may be replaced by a leucine or even phenylalanine in several fungal laccases. These residues cannot be copper ligands.

The X-ray structures of five fungal laccases have recently been reported, from *Coprinus cenereu*,[71] *T. versicolor*,[72,73] *M. albomyces*,[74] *Rigodoporus lignosus*,[75] and *Cerrena maxima*.[232] In addition, the bacterial laccases CotA from *B. subtilis*[228] and CueO from *E. coli*[215] have been characterized by X-ray crystallography. All structures are very similar among each other and to ascorbate oxidase with the three cupredoxin-like domains and the type-1 copper bound to domain 3 and the trinuclear copper center, consisting of the type-3 copper pair and the type-2 copper ion, located between domains 1 and 3. Crystal structures of oxidized, reduced, and peroxide- and azide-bound CotA have been determined.[228] They show that the dioxygen is bound side-on between the copper pair ions in the oxidized form, peroxide bridges the copper pair ions in a zigzag mode, azide binds across the type-3 copper ions so that one terminal nitrogen atom is in an almost identical position to an oxygen atom of the dioxygen in the oxidized structure, and no oxygen atoms are bound in the reduced form.

As previously mentioned, laccase is very closely related to ascorbate oxidase. The principal molecular architecture and arrangement of the mononuclear and trinuclear copper centers are the same. Furthermore, spectroscopic and kinetic properties are similar in many circumstances. Therefore, the catalytic mechanism of the dioxygen reduction should be the same for both. Kinetic studies on fungal and tree laccases have been
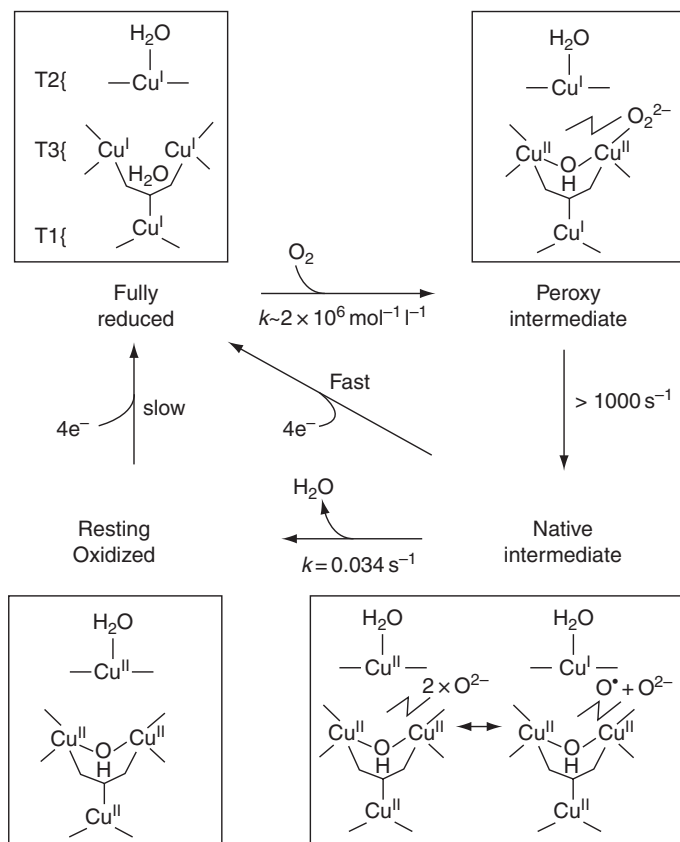
Figure 36 Reaction scheme for laccase as proposed by Solomon and coworkers.[234] Reproduced from S. K. Lee; S. D. George; W. E. Antholine; B. Hedman; K. O. Hodgson; E. I. Solomon, *J. Am. Chem. Soc.* **2002**, *124*, 6180–6193, with permission from American Chemical Society.

reviewed by Reinhammar[233] and the experimental data have been interpreted somewhat differently as in the mechanism shown in **Figure 35**. Solomon and coworkers[178] have intensely studied laccase and ascorbate oxidase with spectroscopic and computational chemical techniques. The presence of two intermediate species termed the peroxy intermediate and the native intermediate could be demonstrated. **Figure 36** summarizes the current working model of the Solomon group of the reaction mechanism,[234] which differs in some points from the scheme in **Figure 35**. The main distinction is the presence of a bridging oxygen ligand between the type-3 copper pair, which is a water in the fully reduced state and a hydroxyl in the other three states. The native intermediate state is split up into two substates. The right one is similar to **Figure 35(f)** with the difference that the type-2 copper (Cu4 in **Figure 35(f)**) is reduced instead of Cu2 from the type-3 copper pair in **Figure 35(f)**. However, in the main features, both reaction schemes are very similar and reflect the spectroscopic, kinetic, and structural known data. Bento *et al.*[219] have presented a reaction scheme that does not involve the fully reduced state in the catalytic cycle and does not contain the spectroscopically proven oxygen radical intermediate. As it is very unlikely that dioxygen binds to the fully oxidized enzyme and the oxygen radical intermediate had not been included, this scheme seems to be very implausible.

The blue oxidases-related enzymes include phenoxazinone synthase from *S. antibioticus*.[235] This enzyme is a copper-containing oxidase that catalyzes the coupling of 2-aminophenols to form the 2-aminophenoxazinone chromophore.[236] This reaction constitutes the final step in the biosynthesis of the potent antineoplastic agent actinomycin. The crystal structure of the oxidized form phenoxazinone synthase from *S. anibioticus* has been determined.[217] It has been solved in his hexameric form. One monomer is very similar to laccase or ascorbate oxidase but it contains a long loop, which connects two domains and stabilizes the hexameric structure. Bound

to this loop is a fifth copper atom that is present as a type-2 copper coordinated by three histidine residues. The trinuclear copper center has one oxygen ligand bound bridging the type-3 copper pair similar to the resting form in ascorbate oxidase.

The gene product of the FET3 gene of *Saccharomyces cerevisiae* is a multicopper oxidase that is required for ferrous iron uptake and exhibits ferroxidase activity.[237] The crystal structure of FET3p reveals its strong similarity to laccase and ascorbate oxidase including the three cupredoxin domain structure and the geometry and arrangement of the type-1 copper and trinuclear copper site.[238]

8.14.2.5.1(iii)   Ceruloplasmin (EC 1.16.3.1)   Ceruloplasmin is exclusively found in the plasma of vertebrates. It is a monomer with a chain length of 1046 amino acid residues (man) with a molecular mass of 132 kDa and a carbohydrate content of 7–8%. Many functions have been ascribed to ceruloplasmin, including mobilization, transport, and homeostasis of copper, ferroxidase, amine oxidase, and possibly superoxide dismutase activity. Ceruloplasmin is an acute-phase protein in the inflammatory response. The protein is deficient in Wilson's disease, an autosomal recessive trait with defect in copper metabolism. Ceruloplasmin catalyzes the oxidation of a great variety of both organic and inorganic substances including amines, dopamine, and seretonin as well as catechol derivatives, aminophenols, and Fe(II) (for reviews see Ryden[239] or Harris *et al.*[240]). The finding that the FET3 gene product of *S. cerevisiae* is a multicopper oxidase and plays a key role in iron metabolism of this eukaryote has underpinned the function of ceruloplasmin in vertebrate iron transport. By virtue of its ferroxidase activity, ceruloplasmin converts Fe(II) into Fe(III), which binds to the iron-binding protein transferrin. Ceruloplasmin is critical for iron egress from some cell types. The transport system responsible for iron release into plasma has not been identified.[241]

The crystal structure of human ceruloplasmin has been determined.[42] The molecule is built up of six cupredoxin domains with a pseudo three-fold symmetry. The type-1 copper site in domain 6 and the trinuclear copper site between domain 1 and 6 are the catalytically functional unit and very similar to the arrangement in ascorbate oxidase and laccase and the reaction schemes for laccase and ascorbate oxidase should also be applicable for ceruloplasmin. The type-1 copper centers in domains 2, 4, and 6 are positioned on an equilateral triangle with the coppers in its corners and a side length of about 18.0 Å. A distance of 18 Å appears within the range for efficient electronic coupling and it is interesting to speculate that the coppers in domain 2 and 4 in human ceruloplasmin increase the possibility of electron capture prior to the transfer of the electron via the domain 6 copper to the oxidase center involving the trinuclear cluster.[242] Two 'labile' metal-binding sites were discovered,[243] one in each of the domains 4 and 6 some 9–10 Å from the respective mononuclear copper sites. These sites contain copper in the crystals used for the X-ray analysis. The coordinating residues for these sites are two glutamates, one aspartate, and one histidine, respectively.

Soaking the crystals for short times with Fe(II) and Fe(III) (1 mmol l^{-1} FeSO$_4$ for 3 h and 1 mmol l^{-1} FeCl$_3$ for 21 h) leads to the removal of the labile copper and the appearance of iron at sites near the outside of the molecule. These are termed the 'holding' sites and four negatively charged residues define their environments. One interpretation of these observations is that the labile sites act as sites of substrate oxidation. For Fe(II) the oxidation mechanism would first involve the cation occupying a labile site (first displacing any Cu(II) present by reducing it to the labile Cu(I)) and then releasing an electron. The oxidized Fe(III) would then translocate to the holding site.

Closely related to ceruloplasmin is hephaestin, which is a transmembrane copper-dependent ferroxidase necessary for effective iron transport from intestinal enterocytes into the circulation.[244] Hephaestin consists of 1185 amino acids with a putative N-terminal signal peptide. It is 50% identical with ceruloplasmin. In contrast to its serum homologue, hephaestin contains additional 86 amino acids at the C-terminus. This segment includes a single predicted transmembrane domain, suggesting that hephaestin is a membrane-bound protein with a large ceruloplasmin-like ectodomain. Comparative structural modeling of the hephaestin ectodomain was based on the known crystal structure of ceruloplasmin and revealed several important facts. Both proteins share the same β-fold and the key structural features for folding and function of ceruloplasmin. All copper-binding sites of ceruloplasmin are also conserved in the structure of hephaestin. The putative iron-binding site with the negatively charged aspartate-rich tract in its vicinity is also conserved in hephaestin.[245] These observations indicate that hephaestin has a ferroxidase activity similar to ceruloplasmin.

8.14.2.5.2 Respiratory oxidases

The heme–copper oxidase superfamily is defined by two criteria: (1) a high degree of amino acid sequence similarity within the largest subunit (subunit I); and (2) a unique bimetallic active site, consisting of a heme and a closely associated copper atom (see **Figure 8**), where dioxygen is reduced to water. There are two main branches of the superfamily, which have distinct substrate specificities: the mitochondrial respiratory oxidases use cytochrome *c* as a substrate and, hence, are called cytochrome *c* oxidases (COX). Bacteria, unlike most mitochondria, contain multiple respiratory oxidases. Many of the prokaryotic respiratory oxidases use membrane-bound quinol (ubiquinol or menaquinol) as a substrate rather than cytochrome *c*. A number of these quinol oxidases have been shown to be members of the heme–copper oxidase superfamily and to pump protons as efficiently as COXs.[246]

8.14.2.5.2(i) Cytochrome c oxidases The eukaryotic COXs all contain 3 mitochondrially encoded subunits (I, II, and III) and up to 10 nucleus-encoded subunits[246] The molecular mass for the protein part of the monomer of bovine heart COX is, for example, 204.005 kDa.[41] Bacterial COXs contain only three or four subunits, (*P. denitrificans* COX, four[40] and *R. sphaeroides* cytochrome *c* oxidase, three[247]).

The crystal structures of bovine heart[41,81] and *P. denitrificans*[40] COXs revealed the molecular spatial structure and its arrangement in the membrane as well as the nature and location of the redox centers in the molecular machine. A schematic representation of subunits I–III with the metal centers and the physiological partner cytochrome *c* is shown in **Figure 37**. The transmembrane part consists of antiparallel transmembrane helices and the membrane-exposed domain of subunit II exhibits the cupredoxin fold. Heme *a*, *a₃*, and Cu_B are found in subunit I and the binuclear copper center Cu_A is in the membrane-exposed domain of subunit II. The geometry of the copper-containing sites is illustrated in **Figures 7 and 8** and features of these redox sites have been discussed in the relevant sections. The arrangement of the metal centers in COXs is depicted in **Figure 38**.

The COX structures support Cu_A as the initial electron acceptor from cytochrome *c*, which is then transferred to heme *a*. Cu_A is closer to heme *a* (19.5 Å, compared to heme a_3, 22.1 Å). Heme *a*, as it is close to heme a_3, would then transfer electrons to the binuclear heme a_3–Cu_B site. Both hemes are perpendicular to the membrane with interplanar angles of 104–108°, the shortest distance between the two hemes is 4.5 Å, the iron-to-iron distance is 13.2 Å. Thus, a direct electron transfer might be possible. The terminal electron acceptor, O_2, is expected to bind between Fe_{a3} and Cu_B.
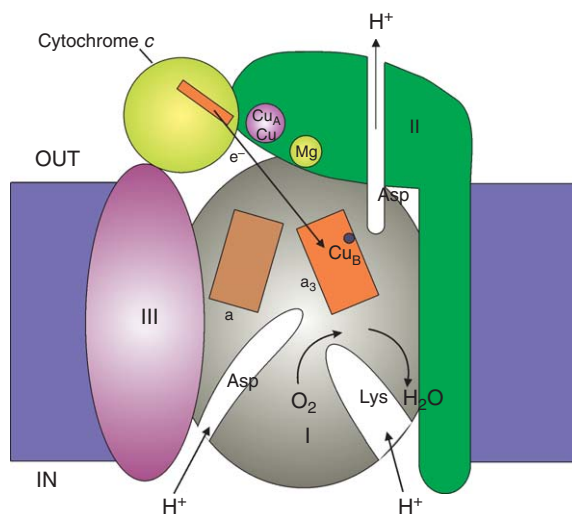
Figure 37 Schematic representation of subunits I–III of cytochrome *c* oxidase. Adapted from R. B. Gennis; S. Ferguson-Miller, *Curr. Biol.* **1996**, *6*, 36–38.
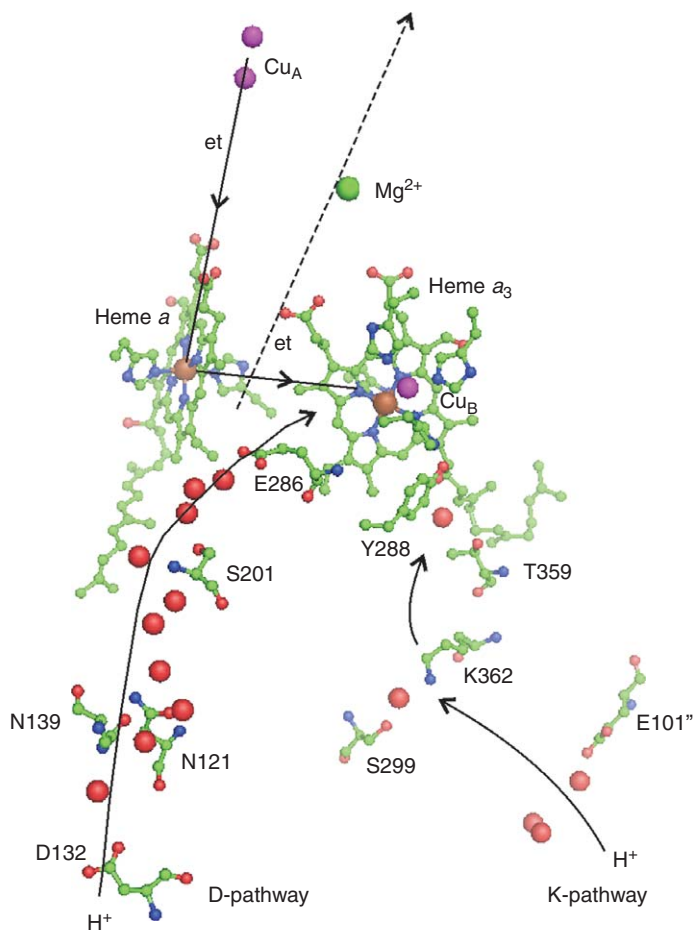
Figure 38 The redox-active cofactors of cytochrome c oxidase and the D and K proton transfer pathways (PDB-code: 1M56), the red spheres are water molecules resolved in the X-ray crystal structures; prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

The X-ray structures deliver some clues how the protons are pumped through the protein complex. Two kinds of protons must be available from the cytoplasmic space: scalar chemical protons for the formation of water from dioxygen and vectorial protons to be pumped across the membrane. For the vectorial protons, two proton transfer pathways have been detected in the COX structures. They are called the D- and K-pathways and connect the proton-input (N) side solution with the catalytic site (**Figure 38**). The output part of the pathway for pumped protons (above the level of hemes a and a_3) is not apparent from analysis of the X-ray structures. However, results from experimental and theoretical calculations indicate that the pathway extends via the region including Arg481 and 482, interacting with the heme propionates.[248] Analyses of the COX crystal structures suggest that the region above the heme propionates at the interface between subunits I and II contains many water molecules, so that proton transfer beyond the propionates may take a large number of alternative routes. There is also a magnesium ion (or manganese) bound in this region, which may have a structural role and also participate in the proton release.[249]

Spectroscopic and chemical studies have provided a reasonably clear picture of the mechanism by which O_2 is reduced to two H_2O during the catalytic cycle.[250] The catalytic cycle (**Figure 39**) is described referring to the different redox states of the catalytic site. The remaining two redox centers, heme a and Cu_A, are not engaged directly in the oxygen chemistry and provide the pathway for electrons to reach the catalytic site. The reaction with O_2 requires that the catalytic site be reduced by two electrons, that is, with the catalytic site in the heme $a_3(Fe^{2+})/Cu_B^+$ state, which we refer to as the R (reduced) state. After O_2 binds to ferrous heme a_3 forming the
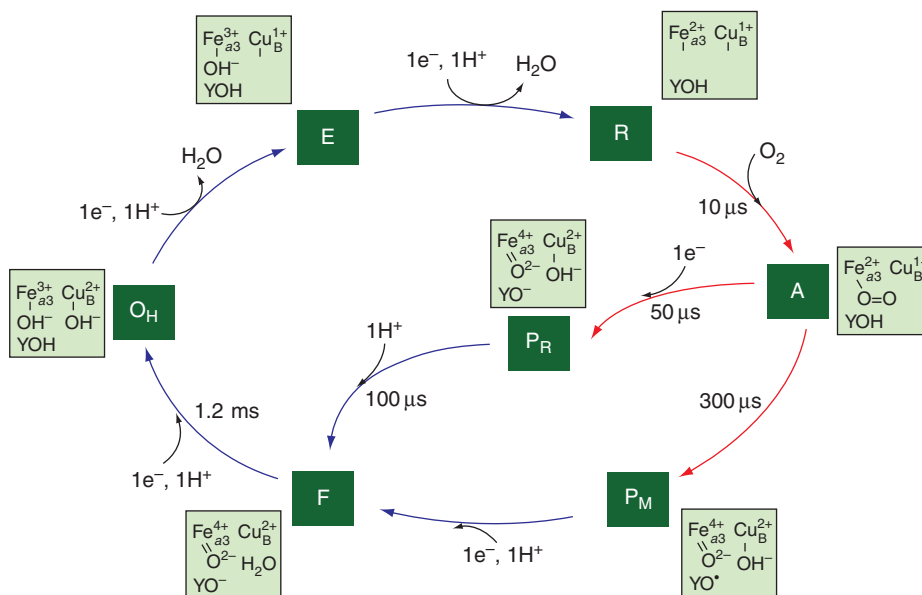
Figure 39 The catalytic reaction of cytochrome *c* oxidase (time constants are those observed with *Rhodobacter sphaeroides*). The outer circle represents the reaction sequence during turnover when electrons are added one by one from cytochrome *c*. The reaction sequence via the P_R state is that observed during reaction of the fully reduced COX with O_2. Reaction steps indicated by blue arrows are linked to proton pumping. Reproduced from G. Branden; R. B. Gennis; P. Brzezinski, *Biochim. Biophys. Acta Bioenerg.* **2006**, *1757*, 1052–1063, with permission from Elsevier Inc.

'A' state (**Figure 39**), the O–O bond is broken in a concerted reaction. To break the O–O bond, four electrons and at least one proton is required. Two of the electrons are donated by the heme a_3 iron (forming the ferryl state, Fe^{4+}) and one from Cu_B (which is oxidized to form Cu_B^{2+}). The source of the additional electron depends on whether heme *a* is oxidized (1), the long route around the circle in **Figure 39**, or reduced (2), the path through P_R, when O_2 binds to the reduced catalytic site.

1. if heme *a* is oxidized when O_2 reacts with the R state of the catalytic site, then the reaction proceeds by oxidizing a nearby amino acid, tentatively identified as Tyr288, which also provides a proton. This is shown in **Figure 39** as the branch going from state A to P_M.

$$\text{YOH heme } a_3 Fe^{2+}\{O\text{–}O\}Cu_B^+ \rightarrow YO^\bullet \text{ heme } a_3\, Fe^{4+} = O^{2-}\, HO^- Cu_B^{2+}$$

$$\text{R state} \qquad\qquad\qquad P_M \text{ state} \tag{13}$$
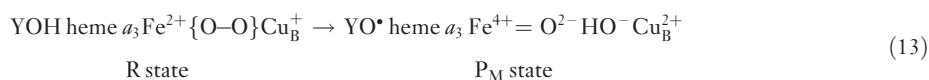
Note that one oxygen atom is bound to the heme iron and one oxygen atom is associated with Cu_B. This reaction is a rapid four-electron reduction of O_2, bypassing any formation of toxic reactive oxygen species (superoxide, peroxide, hydroxide radical) (see Babcock[251]). Even though the $Fe^{4+}=O^{2-}$ state and Tyr radical that are formed at the catalytic site are reactive, they are bound to the COX and are not released. Note that the formation of P_M just rearranges electrons and protons that are already present at the catalytic site and does not require any additional proton or electron input.

The concerted four-electron reduction of O_2 creates a high chemical potential, which is utilized to pump protons across the membrane during each of the following four-electron transfer steps, accompanied by the uptake of four substrate protons, which takes the catalytic site back to the R state (blue arrows, **Figure 39**).

The P_M state has a very high midpoint potential and it is readily reduced. Transfer of an electron into the catalytic site in state P_M, provided from cytochrome $c \rightarrow Cu_A \rightarrow$ heme *a*, probably results in reduction of the Tyr288 radical. Electron transfer to the catalytic site is coupled to a series of proton transfers, which is thought to be the same every time an electron is transferred to the catalytic site in the reaction cycle: two protons are taken up from the *N*-side of the protein and one is released from the *P*-side. One of the protons taken up goes to the catalytic site (substrate proton) and the other proton is pumped.

The state of the catalytic site formed by the reduction of P_M is denoted F. Transfer of an additional electron to the catalytic site is again coupled to the uptake of two protons and the release of one pumped proton. This process results in the formation of the oxidized catalytic site, denoted OH, where the subscript denotes an activated 'high-potential' state in which it is postulated that Cu_B has a very high electrochemical potential.[252] The next two electron transfer reactions convert the O_H state to the E state (Cu_B reduced) and, further to the R state (heme a_3 and Cu_B both reduced). Each of these steps is also thought to be coupled to the uptake of two protons and release of one proton.[252] However, little is known about the nature of the activated O_H state, or the $O_H \rightarrow E$ and $E \rightarrow R$ steps of the reaction. For the purposes of this discussion it is assumed that each of the electron transfer steps to the catalytic site, $O_H \rightarrow E$, $E \rightarrow R$, $P_M \rightarrow F$, and $F \rightarrow O_H$, is associated with proton pumping by the same mechanism, but much more needs to be done experimentally to test this assumption.

2. if heme a is reduced when O_2 reacts with the R state of the catalytic site, then the electron required to break the $O-O$ bond is taken from heme a and not from Tyr288 (**Figure 39**, pathway via P_R). This is the situation when the fully reduced (all four redox centers are reduced) enzyme is reacted with O_2 in the 'flow-flash' reaction commonly used to study the oxidase catalysis. The intermediate that is observed is called the P_R state of the enzyme, which is spectroscopically (UV–vis) identical to P_M, but presumably has a tyrosinate instead of a tyrosine radical at the catalytic site. The ET from heme a to the catalytic site takes about 30–50 μs. After this electron transfer one proton is transferred to the catalytic site, which is observed spectroscopically as the $P_R \rightarrow F$ transition. In addition, the $P_R \rightarrow F$ transition is linked to proton pumping, that is, in this process two protons taken up from the N-side and one proton released on the P-side, but these proton transfer reactions take place with a time constant of 100–200 μs, that is, after the electron transfer from heme a to the catalytic site (**Figure 39**).

In the next transition, $F \rightarrow O_H$, which is observed as part of the reaction of the fully reduced enzyme with O_2, the electron that was originally residing on Cu_A is transferred to the catalytic site (fourth electron), forming the O_H state of the enzyme, as described above. The $F \rightarrow O_H$ transition takes about 1 ms, and the electron and proton transfers are not clearly separable (**Figure 39**). Note that in the reaction of the fully reduced enzyme with O_2 there are only two steps that are linked to proton pumping, $P_R \rightarrow F$ and $F \rightarrow O_H$. If additional reductant is provided to allow the $O_H \rightarrow E$ and $E \rightarrow R$ steps, then two additional protons are pumped and the enzyme is returned to the R state.

8.14.2.5.2(ii)   Quinol oxidases   Quinol oxidases receive their electrons not from a cytochrome but from a quinol (ubiquinol or menaquinol). Therefore, the electron acceptor site differs from that of COXs. A characteristic feature is the lacking of the Cu_A center[253] and different heme types in various combinations like aa_3, ba_3, bb_3, and bo_3.[254] The heme–Cu_B site exhibits very similar properties to that in COXs. The quinol is associated to the membrane and transfers electrons directly to the isolated heme group(s) of subunit I. The dioxygen reduction chemistry at the binuclear redox site will be very similar to that of COXs whereas the spatial proton pathways will be different.

8.14.2.6   Nitrogen Oxide (NO$_x$) Reductases

8.14.2.6.1   Nitrite reductase (EC 1.7.99.3)

Besides dioxygen, nitrogen oxides can serve as electron acceptors in reactions catalyzed by copper enzymes. The copper-containing nitrite reductase (NIR) from denitrifying bacteria such as *Achromobacter*, *Pseudomonas*, or *Rhodobacter* is part of the dissimilatory metabolic pathway of these bacteria. The enzyme catalyzes the one-electron reduction of NO_2^- to NO and water according to Equations (14) and (15).

$$E-Cu^{2+} + e^- + NO_2^- \rightarrow E-Cu^+ - NO_2^- \tag{14}$$

$$E-Cu^+ - NO_2^- + 2H^+ \rightarrow E-Cu^{2+} + NO + H_2O \tag{15}$$
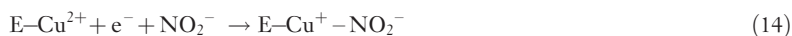
The enzyme is found in the periplasma of Gram-negative bacteria. NIR occurs in solution and in the crystal as a homotrimer (crystal structures have been determined for the *A. cycloclastes*,[255] *A. faecalis*,[256] and *Alcaligenes xylosodidans*[257] NIRs). The trimer shows C3 symmetry. One monomer (339, 378, or 226 amino acid residues, respectively) is
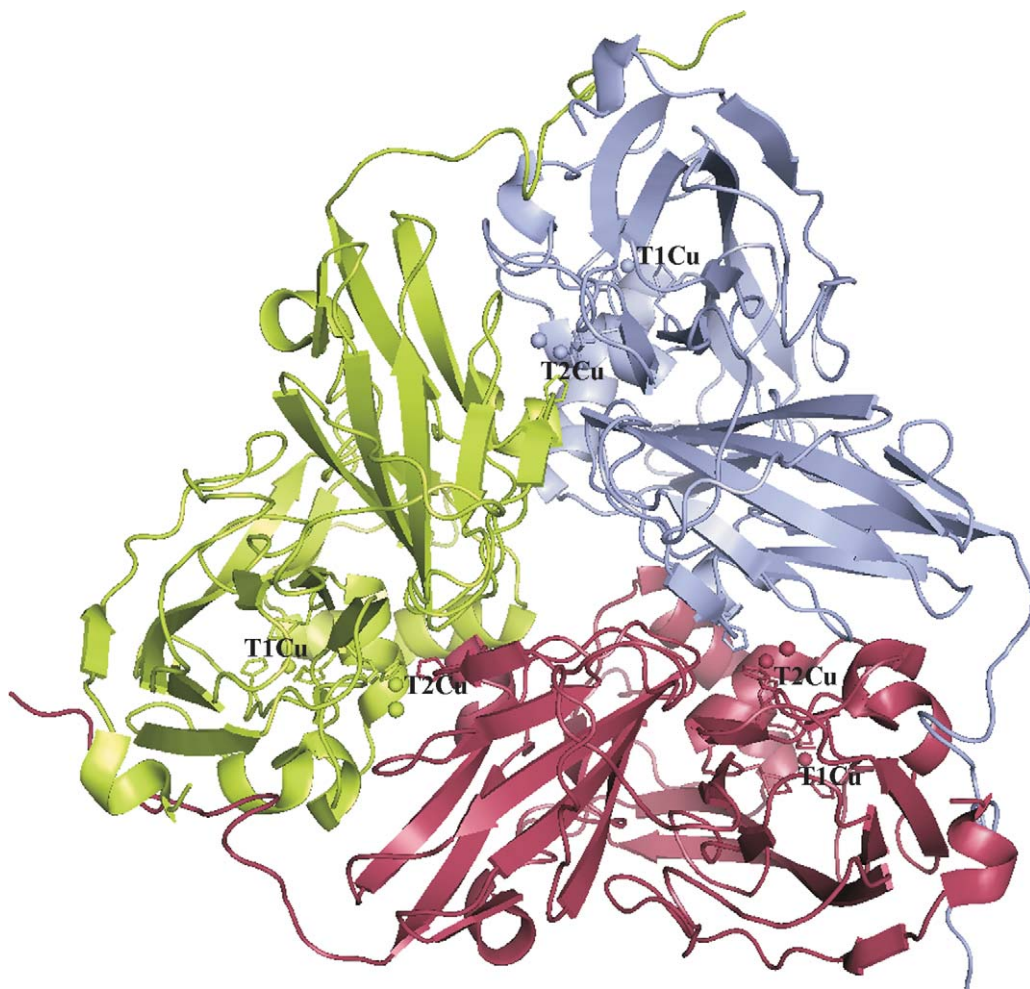
Figure 40 Ribbon diagram of the homotrimer of nitrate reductase from *Alcaligenes xylosoxidans* (PDB-code: 1OE1). The single monomers are colored light blue, lemon, and raspberry. T1Cu, type-1 copper; T2Cu, type-2 copper; prepared with PyMOL (W. L. DeLano, Palo Alto, 2003).

built up of two domains of similar β-barrel fold as found in cupredoxins or in the blue oxidases (**Figure 40**). The enzyme contains two mononuclear copper centers. A type-1 copper site with the canonical ligands (His, Cys, His, and Met) is located in domain 1 and a type-2 copper with three histidines and one water molecule as ligands is bound between domain 1 and domain 2 of the adjacent symmetry-related molecule (**Figure 40**). The type-1 copper and its ligands in nitrite reductase all fall at the type-1 center of ascorbate oxidase or the C-terminal type-1 copper center of ceruloplasmin and all three histidines of the type-2 site as well as the copper fall at the same site as the trinuclear cluster of ascorbate oxidase, laccase, or ceruloplasmin. Sequence comparisons suggest that the type-2 copper of NIR corresponds to one of the type-3 copper pair in the blue oxidases.[258] In the trimer of NIR a six-domian structure is realized, which is reminiscent of the six-domain structure of ceruloplasmin.

The physiological electron donors for NIR are either the cupredoxins pseudoazurin or azurin depending on the organism. Pseudoazurin, for example, reduces the type-1 copper and subsequently the electron is transferred to the type-2 copper site, which is also the binding site for nitrite.[256] Nitrite is bound to the type-2 copper site as demonstrated by electron nucleus double resonance studies on NIR from *Achromobacter xylosoxidans*[259] and in a crystal structure of the complex between nitrite and NIR from *A. cycloclastes*.[43] The crystal structure of the complex shows that nitrite binds asymmetrically with the oxygens toward the copper.
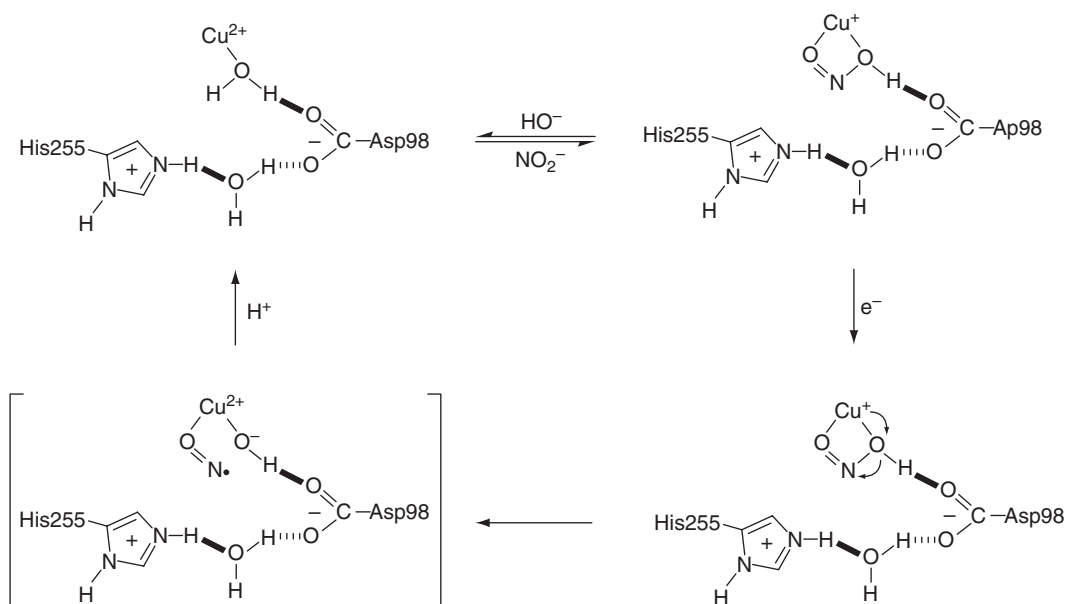
Figure 41   Proposed mechanism of nitrite reductase.[260] Reproduced from E. T. Adman; M. E. P. Murphy, In *Handbook of Metalloproteins*; A. Messerschmidt, R. Huber, T. Poulos, K. Wieghardt, Eds.; John Wiley & Sons: Chichester, 2001; Vol. 2, pp 1381–1390, with permission from John Wiley & Sons.

The combination of biochemical, mutagenesis, and structural results suggest that the mechanism for reduction of nitrite is that shown in **Figure 41**.[260] At the optimum pH, the proton associated with the solvent ligand is balanced between it and the oxygen of Asp98; nitrite binds to oxidized NIR displacing the hydroxyl, which picks up a proton from solvent in the exit channel, NIR is reduced by an ET partner (which could happen at the type-1 site alone in the absence of substrate) with ET to the type-2 copper site occurring most rapidly at the optimum pH. At the type-2 site, the now reduced copper moves an electron to the nitrite oxygen hydrogen bonded to Asp98; the N—O bond is weakened and broken, NO is released, leaving a hydroxyl bound again, and a second proton, probably from locally ordered solvent, restores the solvent ligand. The rate-limiting step is most likely the release of NO, but there is little experimental evidence for that. The formation of N_2O or other nitrosylated products by copper nitrite reductase of *A. cycloclastes* as described by Hulse and colleagues[261] would need a long enough lifetime of the Cu^+ON^+ species. This does not seem to be the case and it is now generally accepted that the copper-containing NIRs reduce nitrite to NO and not to N_2O.

Hasnain and coworkers[262] have structurally and functionally characterized several mutants of NIR from *Alcaligenes xylosoxidans*. These studies establish unequivocally that no direct ET occurs from the physiological electron donor (in this case azurin) to the catalytic type-2 copper center. The mutation of the axial type-1 copper ligand Met144 to Leu increases both the redox potential and catalytic activity, establishing that the rate-determining step of catalysis is the intermolecular ET from azurin to NIR.

8.14.2.6.2   *Nitrous oxide reductase (EC 1.7.99.6)*

Nitrous oxide reductase (N_2OR) catalyzes the last step of the dissimilatory metabolic pathway of denitrifying bacteria (e.g., *Pseudomonas stutzeri*), for a recent review see Eady *et al.*[263] This is a two-electron reduction from N_2O to N_2 and water according to Equation (16)

$$N_2O + 2e^- + 2H^+ \rightarrow N_2 + H_2O \tag{16}$$

This reaction represents difficult chemistry although it is thermodynamically favorable. N_2O is kinetically inert and as a result is a long-lived greenhouse gas.
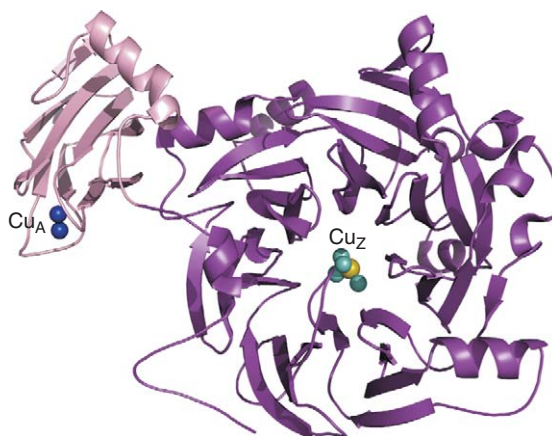
Figure 42 Ribbon diagram of a monomer of nitrous oxide reductase from *Achromobacter cycloclastes* showing two domains, a C-terminal cupredoxin domain (left, in lighter shade) carrying the Cu_A center (dark blue spheres) and the N-terminal β-propeller domain (right, darker shade) with the catalytic Cu_Z center in which copper and inorganic S ions are shown as light blue and dark yellow spheres. Reproduced from R. R. Eady; S. V. Antonyuk; S. S. Hasnain, In *Handbook of Metalloproteins*; A. Messerschmidt, Ed.; John Wiley & Sons: Chichester, 2007; Vol. 4, pp 1–15 with permission from John Wiley & Sons.

When purified from a wide range of bacteria of diverse metabolic groups, N_2ORs have very similar properties. They generally contain only Cu as the metal constituent and are homodimers ($\sim$70 kDa per monomer) with two distinct multinuclear Cu centers per subunit, a binuclear Cu_A ET site (Kroneck *et al.*[77]) (**Figure 7**), and a catalytic Cu_Z center, a novel μ_4-sulfide-bridged tetranuclear Cu cluster (**Figure 9**). As already mentioned, crystal structures of aerobically purified N_2ORs from *Marinobacter hydrocarbonoclasticus* (formerly *P. nautica*),[85,264] *P. denitrificans*,[87] and *A. cycloclastes*[88] have been determined.

The global structures of these enzymes (**Figure 42**) are very similar with each having two domains, an N-terminal seven-bladed β-propeller domain that contains the Cu_Z catalytic center and a C-terminal domain containing the Cu_A center with a cupredoxin fold that is homologous to this domain of COX subunit II. Both copper centers have already been discussed in greater detail above.

During the past years major progress has been made toward the understanding of N_2OR. This includes the discovery of a novel Cu cluster where activation and reduction of this inert gas takes place. The structural results have provided a platform from which one can begin to address the detailed mechanism using an integrated structural biology approach combining site-directed mutagenesis with well-defined kinetic studies, together with high-resolution structures and computational studies. The big challenge remains in terms of how this unique Cu_Z cluster catalyzes the conversion of this potent greenhouse gas to N_2.

Acknowledgment

References

1. R. Lontie, Ed.; *Copper Proteins and Copper Enzymes*; CRC Press: Boca Raton, FL, 1984.
2. A. Messerschmidt; R. Huber; T. Poulos; K. Wieghardt, Eds.; *Handbook of Metalloproteins;* John Wiley & Sons, Ltd: Chichester, 2001; Vol. 2.
3. I. Bertini; A. Sigel; H. Sigel, Eds., *Handbook on Metalloproteins*; Marcel Dekker: New York, 2001.
4. J. S. Valentine; E. B. Gralla, Eds., *Copper-Containing Proteins*; Academic Press: Amsterdam, 2002; Vol. 60.
5. B. Abolmaali; H. V. Taylor; U. Weser, *Struct. Bond.* **1998**, *91*, 92–190.
6. A. Messerschmidt, Copper Metalloenzymes. In *Comprehensive Biological Catalysis*; M. Sinnot, Ed.; Academic Press: London, 1998; Vol. 3, pp 401–426.

7. F. Arnesano; L. Banci, Copper Transporters and Chaperones. In *Handbook of Metalloproteins*; A. Messerschmidt, Ed.; John Wiley & Sons: Chichester, 2007; Vol. 4, pp 1–21.
8. L. A. Finney; T. V. O'Halloran, *Science* **2003**, *300*, 931–936.
9. H. J. Hartmann; L. Morpurgo; A. Desideri; G. Rotilio; U. Weser, *FEBS Lett.* **1983**, *152*, 94–96.
10. M. Vasak; D. W. Hasler, *Curr. Opin. Chem. Biol.* **2000**, *4*, 177–183.
11. E. I. Ochiai, *J. Chem. Educ.* **1986**, *63*, 942–944.
12. F. J. J. R. da Silva; R. J. P. Williams, *The Biological Chemistry of Elements, the Inorganic Chemistry of Life*; Claredon Press: Oxford, 1994.
13. A. C. Rosenzweig; T. V. O'Halloran, *Curr. Opin. Chem. Biol.* **2000**, *4*, 140–147.
14. A. C. Rosenzweig, *Acc. Chem. Res.* **2001**, *34*, 119–128.
15. I. Fridovich, *Science* **1978**, *201*, 875–880.
16. M. M. O. Pena; J. Lee; D. J. Thiele, *J. Nutr.* **1999**, *129*, 1251–1260.
17. J. Camakaris; I. Voskoboinik; J. F. Mercer, *Biochem. Biophys. Res. Commun.* **1999**, *261*, 225–232.
18. R. A. Pufahl; C. P. Singer; K. L. Peariso; S. J. Lin; P. J. Schmidt; C. J. Fahrni; V. C. Culotta; J. E. PennerHahn; T. V. Ohalloran, *Science* **1997**, *278*, 853–856.
19. T. D. Rae; P. J. Schmidt; R. A. Pufahl; V. C. Culotta; T. V. O'Halloran, *Science* **1999**, *284*, 805–808.
20. D. L. Huffman; T. V. O'Halloran, *J. Biol. Chem.* **2000**, *275*, 18611–18614.
21. F. Arnesano; L. Banci; I. Bertini; S. Ciofi-Baffoni; E. Molteni; D. L. Huffman; T. V. O'Halloran, *Genome Res.* **2002**, *12*, 255–271.
22. S. L. Leong; K. J. Barnham; G. Multhaup; R. Cappai, Amyloid Precursor Protein. In *Handbook of Metalloproteins*; A. Messerschmidt, Ed.; John Wiley & Sons: Chichester, 2007; Vol. 4, pp 1–14.
23. P. C. Bull; G. R. Thomas; J. M. Rommens; J. R. Forbes; D. W. Cox, *Nat. Genet.* **1993**, *5*, 327–337.
24. G. J. Brewer, *Proc. Soc. Exp. Biol. Med.* **2000**, *223*, 39–46.
25. G. J. Brewer; R. D. Dick; V. D. Johnson; J. A. Brunberg; K. J. Kluin; J. K. Fink, *J. Lab. Clin. Med.* **1998**, *132*, 264–278.
26. J. F. B. Mercer; J. Livingston; B. Hall; J. A. Paynter; C. Begy; S. Chandrasekharappa; P. Lockhart; A. Grimes; M. Bhave; D. Siemieniak; T. W. Glover, *Nat. Genet.* **1993**, *3*, 20–25.
27. K. Yoshida; K. Furihata; S. Takeda; A. Nakamura; K. Yamamoto; H. Morita; S. Hiyamuta; S. Ikeda; N. Shimizu; N. Yanagisawa, *Nat. Genet.* **1995**, *9*, 267–272.
28. G. J. Brewer, *Curr. Opin. Chem. Biol.* **2003**, *7*, 207–212.
29. R. G. Pearson, *J. Chem. Educ.* **1968**, *45*, 581.
30. M. N. Hughes, *The Inorganic Chemistry of Biological Processes*; John Wiley & Sons: New York, 1990.
31. B. G. Malmstrom, *Eur.J. Biochem.* **1994**, *223*, 711–718.
32. S. J. Lippard; J. M. Berg, *Principles of Bioinorganic Chemistry*; University Science Books: Mill Valley, 1994.
33. S. Mattar; B. Scharf; S. B. H. Kent; K. Rodewald; D. Oesterhelt; M. Engelhard, *J. Biol. Chem.* **1994**, *269*, 14939–14945.
34. B. Reinhammar, *Biochim. Biophys. Acta* **1972**, *275*, 245–259.
35. A. G. Sykes, *Adv. Inorg. Chem.* **1991**, *36*, 377–408.
36. E. T. Adman, *Adv. Protein Chem.* **1991**, *42*, 145–197.
37. G. W. Canters; G. Gilardi, *FEBS Lett.* **1993**, *325*, 39–48.
38. A. Messerschmidt, *Struct. Bond.* **1998**, *90*, 37–68.
39. A. Messerschmidt, *Adv. Bioinorg. Chem.* **1993**, *40*, 121–185.
40. S. Iwata; C. Ostermeier; B. Ludwig; H. Michel, *Nature* **1995**, *376*, 660–669.
41. T. Tsukihara; H. Aoyama; E. Yamashita; T. Tomizaki; H. Yamaguchi; K. Shinzawaltoh; R. Nakashima; R. Yaono; S. Yoshikawa, *Science* **1996**, *272*, 1136–1144.
42. I. Zaitseva; V. Zaitsev; G. Card; K. Moshkov; B. Bax; A. Ralph; P. Lindley, *J. Biol. Inorg. Chem.* **1996**, *1*, 15–23.
43. E. T. Adman; J. W. Godden; S. Turley, *J. Biol. Chem.* **1995**, *270*, 27458–27474.
44. A. Messerschmidt, Spatial Structures of Ascorbate Oxidase, Laccase and Related Proteins: Implications for the Catalytic Mechanism. In *Multi-Copper Oxidases*; A. Messerschmidt, Ed.; World Scientific Publishing: Singapore, 1997; pp 23–79.
45. J. Han; T. M. Loehr; Y. Lu; J. S. Valentine; B. A. Averill; J. Sandersloehr, *J. Am. Chem. Soc.* **1993**, *115*, 4256–4263.
46. Y. Lu; L. B. Lacroix; M. D. Lowery; E. I. Solomon; C. J. Bender; J. Peisach; J. A. Roe; E. B. Gralla; J. S. Valentine, *J. Am. Chem. Soc.* **1993**, *115*, 5907–5918.
47. C. R. Andrew; H. Yeom; J. S. Valentine; B. G. Karlsson; N. Bonander; G. Vanpouderoyen; G. W. Canters; T. M. Loehr; J. Sandersloehr, *J. Am. Chem. Soc.* **1994**, *116*, 11489–11498.
48. B. E. Ramirez; B. G. Malmstrom; J. R. Winkler; H. B. Gray, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11949–11951.
49. H. Nar; A. Messerschmidt; R. Huber; M. Vandekamp; G. W. Canters, *J. Mol. Biol.* **1991**, *221*, 765–772.
50. H. Nar, Röntgenkristallographische Strukturaufklärung von Azurin aus *Pseudomonas aeruginosa*. Technical University, Munich, 1992.
51. J. M. Guss; P. R. Harrowell; M. Murata; V. A. Norris; H. C. Freeman, *J. Mol. Biol.* **1986**, *192*, 361–387.
52. E. Vakoufari; K. S. Wilson; K. Petratos, *FEBS Lett.* **1994**, *347*, 203–206.
53. R. L. Lieberman; D. M. Arciero; A. B. Hooper; A. C. Rosenzweig, *Biochemistry* **2001**, *40*, 5674–5681.
54. P. J. Hart; A. M. Nersissian; R. G. Herrmann; R. M. Nalbandyan; J. S. Valentine; D. Eisenberg, *Protein Sci.* **1996**, *5*, 2175–2183.
55. C. Dennison, *Coord. Chem. Rev.* **2005**, *249*, 3025–3054.
56. N. Ito; S. E. V. Phillips; K. D. S. Yadav; P. F. Knowles, *J. Mol. Biol.* **1994**, *238*, 794–814.
57. M. R. Parsons; M. A. Convery; C. M. Wilmot; K. D. S. Yadav; V. Blakely; A. S. Corner; S. E. V. Phillips; M. J. McPherson; P. F. Knowles, *Structure* **1995**, *3*, 1171–1184.
58. W. F. Novotny; O. Chassande; M. Baker; M. Lazdunski; P. Barbry, *J. Biol. Chem.* **1994**, *269*, 9921–9925.
59. N. J. Blackburn, Chemical and Spectroscopic Studies on Dopamine ß-Hydroxylase and Other Copper Monooxygenases. In *Bioinorganic Chemistry of Copper*; K. D. Karlin, Z. Tyeklar, Eds.; Chapman & Hall: New York, 1993; pp 164–183.
60. S. T. Prigge; A. S. Kolhekar; B. A. Eipper; R. E. Mains; L. M. Amzel, *Science* **1997**, *278*, 1300–1305.
61. J. S. Richardson; K. A. Thomas; B. H. Rubin; D. C. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 1349–1353.
62. J. A. Tainer; E. D. Getzoff; K. M. Beem; J. S. Richardson; D. C. Richardson, *J. Mol. Biol.* **1982**, *160*, 181–217.

63. E. I. Solomon; D. M. Dooley; R. H. Wang; H. B. Gray; M. Cerdonio; F. Mogno; G. L. Romani, *J. Am. Chem. Soc.* **1976**, *98*, 1029–1031.
64. R. S. Himmelwright; N. C. Eickman; C. D. Lubien; K. Lerch; E. I. Solomon, *J. Am. Chem. Soc.* **1980**, *102*, 7339–7344.
65. W. P. J. Gaykema; W. G. J. Hol; J. M. Vereijken; N. M. Soeter; H. J. Bak; J. J. Beintema, *Nature* **1984**, *309*, 23–29.
66. B. Hazes; K. A. Magnus; C. Bonaventura; J. Bonaventura; Z. Dauter; K. H. Kalk; W. G. J. Hol, *Protein Sci.* **1993**, *2*, 597–619.
67. K. A. Magnus; H. Ton-That; J. A. Carpenter, Three-Dimensional Structure of the Oxygenated Form of the Hemocyanin Subunit II of *Limulus Polyphemus* at Atomic Resolution. In *Bioinorganic Chemistry of Copper*; K. D. Karlin, Z. Tyeklar, Eds.; Chapman & Hall: New York, 1993; pp 143–150.
68. Y. Matoba; T. Kumagai; A. Yamamoto; H. Yoshitsu; M. Sugiyama, *J. Biol. Chem.* **2006**, *281*, 8981–8990.
69. M. D. Allendorf; D. J. Spira; E. I. Solomon, *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 3063–3067.
70. A. Messerschmidt; R. Ladenstein; R. Huber; M. Bolognesi; L. Avigliano; R. Petruzzelli; A. Rossi; A. Finazziagro, *J. Mol. Biol.* **1992**, *224*, 179–205.
71. V. Ducros; A. M. Brzozowski; K. S. Wilson; S. H. Brown; P. Ostergaard; P. Schneider; D. S. Yaver; A. H. Pedersen; G. J. Davies, *Nat. Struct. Biol.* **1998**, *5*, 310–316.
72. T. Bertrand; C. Jolivalt; P. Briozzo; E. Caminade; N. Joly; C. Madzak; C. Mougin, *Biochemistry* **2002**, *41*, 7325–7333.
73. K. Piontek; M. Antorini; T. Choinowski, *J. Biol. Chem.* **2002**, *277*, 37663–37669.
74. N. Hakulinen; L. L. Kiiskinen; K. Kruus; M. Saloheimo; A. Paananen; A. Koivula; J. Rouvinen, *Nat. Struct. Biol.* **2002**, *9*, 601–605.
75. S. Garavaglia; M. T. Cambria; M. Miglio; S. Ragusa; V. Lacobazzi; F. Palmieri; C. D'Ambrosio; A. Scaloni; M. Rizzi, *J. Mol. Biol.* **2004**, *342*, 1519–1531.
76. R. Gennis; S. Fergusonmiller, *Science* **1995**, *269*, 1063–1064.
77. P. M. H. Kroneck; W. A. Antholine; J. Riester; W. G. Zumft, *FEBS Lett.* **1988**, *242*, 70–74.
78. P. M. H. Kroneck; W. E. Antholine; D. H. W. Kastrau; G. Buse; G. C. M. Steffens; W. G. Zumft, *FEBS Lett.* **1990**, *268*, 274–276.
79. M. Wilmanns; P. Lappalainen; M. Kelly; E. SauerEriksson; M. Saraste, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11955–11959.
80. M. Svensson-Ek; J. Abramson; G. Larsson; S. Tornroth; P. Brzezinski; S. Iwata, *J. Mol. Biol.* **2002**, *321*, 329–339.
81. T. Tsukihara; H. Aoyama; E. Yamashita; T. Tomizaki; H. Yamaguchi; K. Shinzawaitoh; R. Nakashima; R. Yaono; S. Yoshikawa, *Science* **1995**, *269*, 1069–1074.
82. J. A. Farrar; A. J. Thomson; M. R. Cheesman; D. M. Dooley; W. G. Zumft, *FEBS Lett.* **1991**, *294*, 11–15.
83. J. A. Farrar; W. G. Zumft; A. J. Thomson, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 9891–9896.
84. J. M. Charnock; A. Dreusch; H. Korner; F. Neese; J. Nelson; A. Kannt; H. Michel; C. D. Garner; P. M. H. Kroneck; W. G. Zumft, *Eur. J. Biochem.* **2000**, *267*, 6509.
85. K. Brown; M. Tegoni; M. Prudencio; A. S. Pereira; S. Besson; J. J. Moura; I. Moura; C. Cambillau, *Nat. Struct. Biol.* **2000**, *7*, 191–195.
86. K. Brown; K. Djinovic-Carugo; T. Haltia; I. Cabrito; M. Saraste; J. J. G. Moura; I. Moura; M. Tegoni; C. Cambillau, *J. Biol. Chem.* **2000**, *275*, 41133–41136.
87. T. Haltia; K. Brown; M. Tegoni; C. Cambillau; M. Saraste; K. Mattila; K. Djinovic-Carugo, *Biochem. J.* **2003**, *369*, 77–88.
88. K. Paraskevopoulos; S. V. Antonyuk; R. G. Sawers; R. R. Eady; S. S. Hasnain, *J. Mol. Biol.* **2006**, *362*, 55–65.
89. M. D. Berry; A. V. Juorio; I. A. Paterson, *Prog. Neurobiol.* **1994**, *42*, 375–391.
90. W. S. McIntire; C. Hartmann, Copper-Containing Amine Oxidases. In *Principles and Applications of Quinoproteins*; V. L. Davison, Ed.; Marcel Dekker: New York, 1993; pp 97–171.
91. B. J. Brazeau; B. J. Johnson; C. M. Wilmot, *Arch. Biochem. Biophys.* **2004**, *428*, 22–31.
92. J. L. DuBois; J. P. Klinman, *Arch. Biochem. Biophys.* **2005**, *433*, 255–265.
93. M. Salmi; S. Jalkanen, Enzymatic Control of Leukocyte Trafficking: Role of VAP-1. In *Lymphocyte Activation and Immune Activation IX*; S. Gupta, E. Butcher, W. E. Paul, Eds.; Advances in Experimental Medicine and Biology 512; Springer Verlag: Berlin, Germany, 2002; pp 57–63.
94. S. M. Janes; D. Mu; D. Wemmer; A. J. Smith; S. Kaur; D. Maltby; A. L. Burlingame; J. P. Klinman, *Science* **1990**, *248*, 981–987.
95. M. Mure; S. A. Mills; J. P. Klinman, *Biochemistry* **2002**, *41*, 9269–9278.
96. D. M. Dooley, *J. Biol. Inorg. Chem.* **1999**, *4*, 1–11.
97. M. Kim; T. Okajima; S. Kishishita; M. Yoshimura; A. Kawamori; K. Tanizawa; H. Yamaguchi, *Nat. Struct. Biol.* **2002**, *9*, 591–596.
98. J. M. Murray; C. G. Saysell; C. M. Wilmot; W. S. Tambyrajah; J. Jaeger; P. F. Knowles; S. E. V. Phillips; M. J. McPherson, *Biochemistry* **1999**, *38*, 8217–8227.
99. C. M. Wilmot; J. M. Murray; P. F. Knowles; S. E. V. Phillips, *Science* **1999**, *286*, 1724–1728.
100. C. M. Wilmot; J. M. Murray; G. Alton; M. R. Parsons; M. A. Convery; V. Blakeley; A. S. Corner; M. M. Palcic; P. F. Knowles; M. J. McPherson; S. E. V. Phillips, *Biochemistry* **1997**, *36*, 1608–1620.
101. V. Kumar; D. M. Dooley; H. C. Freeman; J. M. Guss; I. Harvey; M. A. McGuirl; M. C. J. Wilce; V. M. Zubak, *Structure* **1996**, *4*, 943–955.
102. M. C. J. Wilce; D. M. Dooley; H. C. Freeman; J. M. Guss; H. Matsunami; W. S. McIntire; C. E. Ruggiero; K. Tanizawa; H. Yamaguchi, *Biochemistry* **1997**, *36*, 16116–16133.
103. R. B. Li; J. P. Klinman; F. S. Mathews, *Structure* **1998**, *6*, 293–307.
104. A. P. Duff; A. E. Cohen; P. J. Ellis; J. A. Kuchar; D. B. Langley; E. M. Shepard; D. M. Dooley; H. C. Freeman; J. M. Guss, *Biochemistry* **2003**, *42*, 15148–15157.
105. M. Lunelli; M. L. Di Paolo; M. Biadene; V. Calderone; R. Battistutta; M. Scarpa; A. Rigo; G. Zanotti, *J. Mol. Biol.* **2005**, *346*, 991–1004.
106. T. T. Airenne; Y. Nymalm; H. Kidron; D. J. Smith; M. Pihlavisto; M. Salmi; S. Jalkanen; M. S. Johnson; T. A. Salminen, *Protein Sci.* **2005**, *14*, 1964–1974.
107. E. Jakobsson; J. Nilsson; D. Ogg; G. J. Kleywegt, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 1550–1562.
108. B. Schwartz; J. P. Klinman, Mechanism of Biosynthesis of Protein Derived Redox Factors. In *Vitamins and Hormones – Advances in Research and Applications;* G. Liwack, T. Begley, Eds.; Academic Press: New York, 2001; Vol. 61, pp 219–239.
109. J. E. Dove; B. Schwartz; N. K. Williams; J. P. Klinman, *Biochemistry* **2000**, *39*, 3690–3698.
110. B. Schwartz; J. E. Dove; J. P. Klinman, *Biochemistry* **2000**, *39*, 3699–3707.

111. J. P. Klinman, *Chem. Rev.* **1996**, *96*, 2541–2561.
112. D. M. Dooley; M. A. McGuirl; D. E. Brown; P. N. Turowski; W. S. McIntire; P. F. Knowles, *Nature* **1991**, *349*, 262–264.
113. S. Kishishita; T. Okajima; M. Kim; H. Yamaguchi; S. Hirota; S. Suzuki; S. Kuroda; K. Tanizawa; M. Mure, *J. Am. Chem. Soc.* **2003**, *125*, 1041–1055.
114. Q. J. Su; J. P. Klinman, *Biochemistry* **1998**, *37*, 12513–12525.
115. S. A. Mills; Y. Goto; Q. J. Su; J. Plastino; J. P. Klinman, *Biochemistry* **2002**, *41*, 10577–10584.
116. F. Buffoni, *Pharmacol. Rev.* **1966**, *18*, 1163–1199.
117. B. O. Elmore; J. A. Bollinger; D. M. Dooley, *J. Biol. Inorg. Chem.* **2002**, *7*, 565–579.
118. H. M. Kagan, Characterization and Regulation of Lysyl Oxidase. In *Regulation of Matrix Accumulation*; R. P. Mecham, Ed.; Biology of Extracellular Matrix; Academic Press: Orlando, 1986; pp. 321–398.
119. S. N. Gacheru; P. C. Trackman; M. A. Shah; C. Y. Ogara; P. Spacciapoli; F. T. Greenaway; H. M. Kagan, *J. Biol. Chem.* **1990**, *265*, 19022–19027.
120. E. D. Harris; Wa. Gonnerma; J. E. Savage; B. L. Odell, *Biochim. Biophys. Acta* **1974**, *341*, 332–344.
121. H. M. Kagan; N. A. Hewitt; L. L. Salcedo; C. Franzbla, *Biochim. Biophys. Acta* **1974**, *365*, 223–234.
122. P. C. Trackman; A. M. Pratt; A. Wolanski; S. S. Tang; G. D. Offner; R. F. Troxler; H. M. Kagan, *Biochemistry* **1990**, *29*, 4863–4870.
123. S. X. Wang; M. Mure; K. F. Medzihradszky; A. L. Burlingame; D. E. Brown; D. M. Dooley; A. J. Smith; H. M. Kagan; J. P. Klinman, *Science* **1996**, *273*, 1078–1084.
124. A. P. Duff; A. E. Cohen; P. J. Ellis; K. Hilmer; D. B. Langley; D. M. Dooley; H. C. Freeman; J. M. Guss, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 1073–1084.
125. J. W. Whittacker, Galactose Oxidase. In *Copper-Containing Molecules*; J. S. Valentine, E. B. Gralla, Eds.; Advances in Protein Chemistry; Academic Press: New York 2002; Vol. 60; pp ; pp 1–49.
126. J. W. Whittaker, *Chem. Rev.* **2003**, *103*, 2347–2363.
127. M. J. McPherson; M. R. Parsons; R. K. Spooner; C. M. Wilmot, Prokaryotic Copper Amine Oxidases. In *Handbook of Metalloproteins*; A. Messerschmidt, R. Huber, T. Poulos, K. Wieghardt, Eds.; John Wiley & Sons: Chichester, 2001; Vol. 2, pp 1272–1283.
128. G. Avigad; C. Asensio; B. L. Horecker; D. Amaral, *J. Biol. Chem.* **1962**, *237*, 2736–2743.
129. M. M. Whittaker; J. W. Whittaker, *J. Biol. Chem.* **1988**, *263*, 6074–6080.
130. P. F. Knowles; N. Ito, Galactose Oxidase. In *Perspectives on Bioinorganic Chemistry*; R. W. Hay, J. R. Dilworth, Eds.; JAI Press: London, 1993, Vol. 2, pp 208–243.
131. N. Ito; S. E. V. Phillips; C. Stevens; Z. B. Ogel; M. J. McPherson; J. N. Keen; K. D. S. Yadav; P. F. Knowles, *Nature* **1991**, *350*, 87–90.
132. G. T. Babcock; M. K. Eldeeb; P. O. Sandusky; M. M. Whittaker; J. W. Whittaker, *J. Am. Chem. Soc.* **1992**, *114*, 3727–3734.
133. M. M. Whittaker; Y. Y. Chuang; J. W. Whittaker, *J. Am. Chem. Soc.* **1993**, *115*, 10029–10035.
134. J. W. Whittaker, Free Radical-Coupled Copper Active Site of Galactose Oxidase. In *Metalloenzymes Involving Amino Acid-Residue and Related Radicals*; H. Sigel, A. Sigel, Eds.; Marcel Dekker: New York, 1994, Vol. 30, pp 315–360.
135. M. M. Whittaker; J. W. Whittaker, *J. Biol. Chem.* **2003**, *278*, 22090–22101.
136. R. M. Wachter; B. P. Branchaud, *Biochemistry* **1996**, *35*, 14425–14435.
137. P. F. Knowles; R. D. Brown; S. H. Koenig; S. Wang; R. A. Scott; M. A. McGuirl; D. E. Brown; D. M. Dooley, *Inorg. Chem.* **1995**, *34*, 3895–3902.
138. M. M. Whittaker; J. W. Whittaker, *Biophys. J.* **1993**, *64*, 762–772.
139. M. M. Whittaker; D. P. Ballou; J. W. Whittaker, *Biochemistry* **1998**, *37*, 8426–8436.
140. M. M. Whittaker; P. J. Kersten; N. Nakamura; J. SandersLoehr; E. S. Schweizer; J. W. Whittaker, *J. Biol. Chem.* **1996**, *271*, 681–687.
141. M. M. Whittaker; P. J. Kersten; D. Cullen; J. W. Whittaker, *J. Biol. Chem.* **1999**, *274*, 36226–36232.
142. H. S. Mason, *Annu. Rev. Biochem.* **1965**, *34*, 595.
143. L. C. Stewart; J. P. Klinman, *Annu. Rev. Biochem.* **1988**, *57*, 551–592.
144. A. F. Bradbury; M. D. A. Finnie; D. G. Smyth, *Nature* **1982**, *298*, 686–688.
145. J. Bell; R. El Meskini; D. D'Amato; R. E. Mains; B. A. Eipper, *Biochemistry* **2003**, *42*, 7133–7142.
146. B. A. Eipper; A. S. W. Quon; R. E. Mains; J. S. Boswell; N. J. Blackburn, *Biochemistry* **1995**, *34*, 2857–2865.
147. W. A. Francisco; N. J. Blackburn; J. P. Klinman, *Biochemistry* **2003**, *42*, 1813–1819.
148. S. Jaron; N. J. Blackburn, *Biochemistry* **2001**, *40*, 6867–6875.
149. S. T. Prigge; A. S. Kolhekar; B. A. Eipper; R. E. Mains; L. M. Amzel, *Nat. Struct. Biol.* **1999**, *6*, 976–983.
150. S. T. Prigge; B. A. Eipper; R. E. Mains; L. M. Amzel, *Science* **2004**, *304*, 864–867.
151. A. Saxena; P. Hensley; J. C. Osborne; P. J. Fleming, *J. Biol. Chem.* **1985**, *260*, 3386–3392.
152. J. G. Robertson; G. W. Adams; K. F. Medzihradszky; A. L. Burlingame; J. J. Villafranca, *Biochemistry* **1994**, *33*, 11563–11575.
153. E. M. McHugh; R. McGee; P. J. Fleming, *J. Biol. Chem.* **1985**, *260*, 4409–4417.
154. E. J. Lewis; L. P. Asnani, *J. Biol. Chem.* **1992**, *267*, 494–500.
155. L. C. Stewart; J. P. Klinman, *Biochemistry* **1987**, *26*, 5302–5309.
156. M. C. Brenner; C. J. Murray; J. P. Klinman, *Biochemistry* **1989**, *28*, 4656–4664.
157. B. J. Reedy; N. J. Blackburn, *J. Am. Chem. Soc.* **1994**, *116*, 1924–1931.
158. J. P. Klinman, *J. Biol. Chem.* **2006**, *281*, 3013–3016.
159. J. P. Evans; K. Ahn; J. P. Klinman, *J. Biol. Chem.* **2003**, *278*, 49691–49698.
160. P. Chen; E. I. Solomon, *J. Am. Chem. Soc.* **2004**, *126*, 4991–5000.
161. M. J. Knapp; J. P. Klinman, *Eur.J. Biochem.* **2002**, *269*, 3113–3121.
162. W. A. Francisco; M. J. Knapp; N. J. Blackburn; J. P. Klinman, *J. Am. Chem. Soc.* **2002**, *124*, 8194–8195.
163. W. A. Francisco; G. Wille; A. J. Smith; D. J. Merkler; J. P. Klinman, *J. Am. Chem. Soc.* **2004**, *126*, 13168–13169.
164. S. M. Miller; J. P. Klinman, *Biochemistry* **1983**, *22*, 3091–3096.
165. R. S. Hanson; T. E. Hanson, *Microbiol. Rev.* **1996**, *60*, 439.
166. S. Park; K. W. Brown; J. C. Thomas, *Waste Manag. Res.* **2002**, *20*, 434–444.

167. J. P. Sullivan; D. Dickinson; H. A. Chase, *Crit. Rev. Microbiol.* **1998**, *24*, 335–373.

168. M. Merkx; D. A. Kopp; M. H. Sazinsky; J. L. Blazyk; J. Muller; S. J. Lippard, *Angew. Chem. Int. Ed.* **2001**, *40*, 2782–2807.

169. D. J. Arp; L. A. Sayavedra-Soto; N. G. Hommes, *Arch. Microbiol.* **2002**, *178*, 250–255.

170. A. C. Rosenzweig; C. A. Frederick; S. J. Lippard; P. Nordlund, *Nature* **1993**, *366*, 537–543.

171. R. L. Lieberman; A. C. Rosenzweig, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 147–164.

172. J. D. Semrau; A. Chistoserdov; J. Lebron; A. Costello; J. Davagnino; E. Kenna; A. J. Holmes; R. Finch; J. C. Murrell; M. E. Lidstrom, *J. Bacteriol.* **1995**, *177*, 3071–3079.

173. S. Stolyar; A. M. Costello; T. L. Peeples; M. E. Lidstrom, *Microbiol.-SGM* **1999**, *145*, 1235–1244.

174. S. S. F. Yu; K. H. C. Chen; M. Y. H. Tseng; Y. S. Wang; C. F. Tseng; Y. J. Chen; D. S. Huang; S. I. Chan, *J. Bacteriol.* **2003**, *185*, 5915–5924.

175. R. L. Lieberman; D. B. Shrestha; P. E. Doan; B. M. Hoffman; T. L. Stemmler; A. C. Rosenzweig, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3820–3825.

176. R. L. Lieberman; A. C. Rosenzweig, *Nature* **2005**, *434*, 177–182.

177. R. L. Lieberman; K. C. Kondapalli; D. B. Shrestha; A. S. Hakemian; S. M. Smith; J. Telser; J. Kuzelka; R. Gupta; A. S. Borovik; S. J. Lippard; B. M. Hoffman; A. C. Rosenzweig; T. L. Stemmler, *Inorg. Chem.* **2006**, *45*, 8372–8381.

178. E. I. Solomon; U. M. Sundaram; T. E. Machonkin, *Chem. Rev.* **1996**, *96*, 2563–2605.

179. W. S. Oetting, *Pigm. Cell Res.* **2000**, *13*, 320–325.

180. T. Klabunde; C. Eicken; J. C. Sacchettini; B. Krebs, *Nat. Struct. Biol.* **1998**, *5*, 1084–1090.

181. K. Lerch; M. Huber; H. J. Schneider; R. Drexel; B. Linzen, *J. Inorg. Biochem.* **1986**, *26*, 213–217.

182. G. Muller; S. Ruppert; E. Schmid; G. Schutz, *EMBO J.* **1988**, *7*, 2723–2730.

183. A. Volbeda; W. G. J. Hol, *J. Mol. Biol.* **1989**, *206*, 531–546.

184. M. E. Cuff; K. I. Miller; K. E. van Holde; W. A. Hendrickson, *J. Mol. Biol.* **1998**, *278*, 855–870.

185. A. Volbeda; W. G. J. Hol, *J. Mol. Biol.* **1989**, *209*, 249–279.

186. A. Sanchezferrer; J. N. Rodriguezlopez; F. Garciacanovas; F. Garciacarmona, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **1995**, *1247*, 1–11.

187. C. Eicken; B. Krebs; J. C. Sacchettini, *Curr. Opin. Struct. Biol.* **1999**, *9*, 677–683.

188. K. D. Karlin; S. Kaderli; A. D. Zuberbuhler, *Acc. Chem. Res.* **1997**, *30*, 139–147.

189. N. Kitajima; Y. Morooka, *Chem. Rev.* **1994**, *94*, 737–757.

190. P. L. Holland; W. B. Tolman, *Coord. Chem. Rev.* **1999**, *192*, 855–869.

191. B. Salvato; M. Santamaria; M. Beltramini; G. Alzuet; L. Casella, *Biochemistry* **1998**, *37*, 14065–14077.

192. H. Decker; F. Tuczek, *Trends Biochem. Sci.* **2000**, *25*, 392–397.

193. T. Oka; F. J. Simpson; Hg. Krishnam, *Can.J. Microbiol.* **1972**, *18*, 493–508.

194. H. K. Hund; J. Breuer; F. Lingens; J. Huttermann; R. Kappl; S. Fetzner, *Eur.J. Biochem.* **1999**, *263*, 871–878.

195. G. Speier, Copper Dioxygenation Chemistry Relevant to Quercitin Dioxygenase. In *Bioinorganic Chemistry of Copper*; K. D. Karlin, Z. Tyeklar, Eds.; Chapman & Hall: New York, 1993; pp 382–394.

196. F. Fusetti; K. H. Schroter; R. A. Steiner; P. I. van Noort; T. Pijning; H. J. Rozeboom; K. H. Kalk; M. R. Egmond; B. W. Dijkstra, *Structure* **2002**, *10*, 259–268.

197. J. M. Dunwell; S. Khuri; P. J. Gane, *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 153–179.

198. E. J. Woo; J. M. Dunwell; P. W. Goodenough; A. C. Marvier; R. W. Pickersgill, *Nat. Struct. Biol.* **2000**, *7*, 1036–1040.

199. S. Karlin; Z. Y. Zhu; K. D. Karlin, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 14225–14230.

200. L. Barhacs; J. Kaizer; G. Speier, *J. Mol. Catal. A: Chem.* **2001**, *172*, 117–125.

201. E. Balogh-Hergovich; J. Kaizer; G. Speier, *J. Mol. Catal. A: Chem.* **2000**, *159*, 215–224.

202. I. Fridovich, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1986**, *58*, 61–97.

203. J. P. Phillips; S. D. Campbell; D. Michaud; M. Charbonneau; A. J. Hilliker, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 2761–2765.

204. B. A. Omar; S. C. Flores; J. M. McCord, *Adv. Pharmacol.* **1992**, *23*, 109–161.

205. J. S. Valentine; P. A. Doucette; S. Z. Potter, *Annu. Rev. Biochem.* **2005**, *74*, 563–593.

206. J. M. McCord; I. Fridovic, *J. Biol. Chem.* **1969**, *244*, 6049.

207. J. A. Fee; C. Bull, *J. Biol. Chem.* **1986**, *261*, 13000–13005.

208. R. W. Strange; S. Antonyuk; M. A. Hough; P. A. Doucette; J. A. Rodriguez; P. J. Hart; L. J. Hayward; J. S. Valentine; S. S. Hasnain, *J. Mol. Biol.* **2003**, *328*, 877–891.

209. W. F. Beyer; I. Fridovich; G. T. Mullenbach; R. Hallewell, *J. Biol. Chem.* **1987**, *262*, 11182–11187.

210. L. Banci; I. Bertini; C. Luchinat; R. A. Hallewell, *J. Am. Chem. Soc.* **1988**, *110*, 3629–3633.

211. J. A. Tainer; E. D. Getzoff; J. S. Richardson; D. C. Richardson, *Nature* **1983**, *306*, 284–287.

212. L. Banci; I. Bertini; B. Bruni; P. Carloni; C. Luchinat; S. Mangani; P. L. Orioli; M. Piccioli; W. Ripniewski; K. S. Wilson, *Biochem. Biophys. Res. Commun.* **1994**, *202*, 1088–1095.

213. E. D. Getzoff; D. E. Cabelli; C. L. Fisher; H. E. Parge; M. S. Viezzoli; L. Banci; R. A. Hallewell, *Nature* **1992**, *358*, 347–351.

214. M. W. Calhoun; J. W. Thomas; R. B. Gennis, *Trends Biochem. Sci.* **1994**, *19*, 325–330.

215. S. A. Roberts; A. Weichsel; G. Grass; K. Thakali; J. T. Hazzard; G. Tollin; C. Rensing; W. R. Montfort, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2766–2771.

216. F. J. Enguita; L. O. Martins; A. O. Henriques; M. A. Carrondo, *J. Biol. Chem.* **2003**, *278*, 19416–19425.

217. A. W. Smith; A. Camara-Artigas; M. T. Wang; J. P. Allen; W. A. Francisco, *Biochemistry* **2006**, *45*, 4378–4387.

218. A. Messerschmidt, Ed., *Multi-Copper Oxidases*; World Scientific Publishing: Singapore, 1997.

219. I. Bento; M. A. Carrondo; P. F. Lindley, *J. Biol. Inorg. Chem.* **2006**, *11*, 539–547.

220. A. Messerschmidt; W. Steigemann; R. Huber; G. Lang; P. M. H. Kroneck, *Eur.J. Biochem.* **1992**, *209*, 597–602.

221. A. Messerschmidt; H. Luecke; R. Huber, *J. Mol. Biol.* **1993**, *230*, 997–1014.

222. G. Chichiricco; M. P. Ceru; A. Dalessandro; A. Oratore; L. Avigliano, *Plant Sci.* **1989**, *64*, 61–66.

223. L. Avigliano; A. Finazzi-Agro, Biological Function and Enzyme Kinetics of Ascorbate Oxidase. In *Multi-Copper Oxidases*; A. Messerschmidt, Ed.; World Scientific Publishing: Singapore, 1997; pp 251–284.

224. A. Marchesini; P. Capelletti; L. Canonica; B. Danieli; S. Tollari, *Biochim. Biophys. Acta* **1977**, *484*, 290–300.

225. O. Farver; I. Pecht, Electron Transfer Reactions in Multi-Copper Oxidases. In *Multi-Copper Oxidases*; A. Messerschmidt, Ed.; World Scientific Publishing: Singapore, 1997; pp 355–389.
226. E. I. Solomon; T. E. Machokin; U. M. Sundaram, Spectroscopy of Multi-Copper Oxidases. In *Multi-Copper Oxidases*; A. Messerschmidt, Ed.; World Scientific Publishing: Singapore, 1997; pp 102–127.
227. L. C. Petersen; H. Degn, *Biochim. Biophys. Acta* **1978**, *526*, 85–92.
228. I. Bento; L. O. Martins; G. G. Lopes; M. A. Carrondo; P. F. Lindley, *Dalton Trans.* **2005**, 3507–3513.
229. L. E. Andreasson; R. Branden; B. Reinhammar, *Biochim. Biophys. Acta* **1976**, *438*, 370–379.
230. B. Reinhammar; B. G. Malmstrom, ''Blue'' Copper-Containing Oxidases. In *Copper Proteins, Metal Ions in Biology*; T. G. Spiro, Ed.; John Wiley & Sons: New York, 1981; Vol. 3, pp 109–149.
231. M. Smith; C. F. Thurston; D. A. Wood, Fungal Laccases: Role in Delignification and Possible Industrial Applications. In *Multi-Copper Oxidases*; A. Messerschmidt, Ed.; World Scientific Publishing: Singapore, 1997; pp 201–224.
232. A. V. Lyashenko; N. E. Zhukhlistova; A. G. Gabdoulkhakov; Y. N. Zhukova; W. Voelter; V. N. Zaitsev; I. Bento; E. V. Stepanova; G. S. Kachalova; O. V. Koroleva; E. A. Cherkashyn; V. I. Tishkov; V. S. Lamzin; K. Schirwitz; E. Y. Morgunova; C. Betzel; P. F. Lindley; A. M. Mikhailov, *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2006**, *62*, 954–957.
233. B. Reinhammar, Kinetic Studies on Polyporus and Tree Laccases. In *Multi-Copper Oxidases*; A. Messerschmidt, Ed.; World Scientific Publishing: Singapore, 1997; pp 167–200.
234. S. K. Lee; S. D. George; W. E. Antholine; B. Hedman; K. O. Hodgson; E. I. Solomon, *J. Am. Chem. Soc.* **2002**, *124*, 6180–6193.
235. H. A. Choy; G. H. Jones, *Arch. Biochem. Biophys.* **1981**, *211*, 55–65.
236. C. E. Barry; P. G. Nayar; T. P. Begley, *Biochemistry* **1989**, *28*, 6323–6333.
237. C. Askwith; D. Eide; A. Vanho; P. S. Bernard; L. T. Li; S. Daviskaplan; D. M. Sipe; J. Kaplan, *Cell* **1994**, *76*, 403–410.
238. A. B. Taylor; C. S. Stoj; L. Ziegler; D. J. Kosman; P. J. Hart, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15459–15464.
239. L. Ryden, Ceruloplasmin. In *Copper Proteins and Copper Enzymes*; R. Lontie, Ed.; CRC Press: Boca Raton, FL, 1984; Vol. 3, pp 37–100.
240. Z. L. Harris; H. Morita; J. D. Gitlin, The Biology of Human Ceruloplasmin. In *Multi-Copper Oxidases*; A. Messerschmidt, Ed.; World Scientific Publishing: Singapore, 1997; pp 285–304.
241. J. Kaplan; T. V. Ohalloran, *Science* **1996**, *271*, 1510–1512.
242. P. F. Lindley; I. Zaitseva; V. Zaitsev; G. Card; K. Moshkov; B. Bax, The Structure of Human Ceruloplasmin at 3.1 Å Resolution. In *Multi-Copper Oxidases*; A. Messerschmidt, Ed.; World Scientific Publishing: Singapore, 1997; pp 81–102.
243. P. F. Lindley; G. Card; I. Zaitseva; V. Zaitsev; B. Reinhammar; E. SelinLindgren; K. Yoshida, *J. Biol. Inorg. Chem.* **1997**, *2*, 454–463.
244. J. Petrak; D. Vyoral, *Int.J. Biochem. Cell Biol.* **2005**, *37*, 1173–1178.
245. B. A. Syed; N. J. Beaumont; A. Patel; C. E. Naylor; H. K. Bayele; C. L. Joannou; P. S. N. Rowe; R. W. Evans; S. K. S. Srai, *Protein Eng.* **2002**, *15*, 205–214.
246. M. Saraste, *Q. Rev. Biophys.* **1990**, *23*, 331–366.
247. J. P. Hosler; J. Fetter; M. M. J. Tecklenburg; M. Espe; C. Lerma; S. Fergusonmiller, *J. Biol. Chem.* **1992**, *267*, 24264–24272.
248. H. Michel, *Biochemistry* **1999**, *38*, 15129–15140.
249. B. Schmidt; J. McCracken; S. Ferguson-Miller, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15539–15542.
250. G. Branden; R. B. Gennis; P. Brzezinski, *Biochim. Biophys. Acta Bioenerg.* **2006**, *1757*, 1052–1063.
251. G. T. Babcock, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12971–12973.
252. D. Bloch; I. Belevich; A. Jasaitis; C. Ribacka; A. Puustinen; M. I. Verkhovsky; M. Wikstrom, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 529–533.
253. J. Vanderoost; A. P. N. Deboer; J. W. L. Degier; W. G. Zumft; A. H. Stouthamer; R. J. M. Vanspanning, *FEMS Microbiol. Lett.* **1994**, *121*, 1–9.
254. J. A. Garciahorsman; B. Barquera; J. Rumbley; J. X. Ma; R. B. Gennis, *J. Bacteriol.* **1994**, *176*, 5587–5600.
255. J. W. Godden; S. Turley; D. C. Teller; E. T. Adman; M. Y. Liu; W. J. Payne; J. Legall, *Science* **1991**, *253*, 438–442.
256. M. Kukimoto; M. Nishiyama; M. E. P. Murphy; S. Turley; E. T. Adman; S. Horinouchi; T. Beppu, *Biochemistry* **1994**, *33*, 5246–5252.
257. M. J. Ellis; F. E. Dodd; G. Sawers; R. R. Eady; S. S. Hasnain, *J. Mol. Biol.* **2003**, *328*, 429–438.
258. F. F. Fenderson; S. Kumar; E. T. Adman; M. Y. Liu; W. J. Payne; J. Legall, *Biochemistry* **1991**, *30*, 7180–7185.
259. B. D. Howes; Z. H. L. Abraham; D. J. Lowe; T. Bruser; R. R. Eady; B. E. Smith, *Biochemistry* **1994**, *33*, 3171–3177.
260. E. T. Adman; M. E. P. Murphy, Copper Nitrite Reductase. In *Handbook of Metalloproteins*;A. Messerschmidt, R. Huber, T. Poulos, K. Wieghardt, Eds.; John Wiley & Sons: Chichester, 2001; Vol. 2, pp 1381–1390.
261. C. L. Hulse; B. A. Averill; J. M. Tiedje, *J. Am. Chem. Soc.* **1989**, *111*, 2322–2323.
262. M. A. Hough; M. J. Ellis; S. Antonyuk; R. W. Strange; G. Sawers; R. R. Eady; S. S. Hasnain, *J. Mol. Biol.* **2005**, *350*, 300–309.
263. R. R. Eady; S. V. Antonyuk; S. S. Hasnain, Nitrous Oxide Reductase. In *Handbook of Metalloproteins*; A. Messerschmidt, Ed.; John Wiley & Sons: Chichester, 2007; Vol. 4, pp 1–15.
264. M. Prudencio; A. S. Pereira; P. Tavares; S. Besson; I. Cabrito; K. Brown; B. Samyn; B. Devreese; J. Van Beeumen; F. Rusnak; G. Fauque; J. J. G. Moura; M. Tegoni; C. Cambillau; I. Moura, *Biochemistry* **2000**, *39*, 3899–3907.

Biographical Sketch

Albrecht Messerschmidt is a research group leader at the Max-Planck-Institute of Biochemistry, Martinsried, Germany and extraordinary professor at the Biological Faculty of the University of Constance, Germany.

He obtained his Diploma in Crystallography and his Ph.D. both at the Humboldt-University, Berlin. After working on X-ray crystallography of biologically important steroid molecules at the Central Institute for Molecular Biology at the Academy of Sciences of the GDR in Berlin-Buch, he moved in 1983 to the Department of Structural Research at the Max-Planck-Institute of Biochemistry in Martinsried, headed by the Nobel laureate Robert Huber.

His main research interest is the determination of X-ray structures of metalloproteins and metalloenzymes, PLP-dependent enzymes, and, recently, structural proteomics of protein kinases and adapter proteins involved in signal transduction. His group was initially in the department of Robert Huber and since 2005 it has moved to the Department of Proteomics and Signal Transduction headed by Matthias Mann.

Albrecht Messerschmidt is the editor of the three volumes of the *Handbook of Metalloproteins* published by Wiley, Chichester; editor of *Multicopper Oxidases* from World Scientific Publishing, Singapore; and the author of the textbook *X-Ray Crystallography of Biomacromolecules* published by Wiley-VCH, Weinheim.

8.15 Mechanisms of Metal-Dependent Hydrolases in Metabolism

Marcy Hernick, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Carol Fierke, University of Michigan, Ann Arbor, MI, USA

8.15.1 Introduction

The metal-dependent hydrolase class of enzymes uses catalytic metal ion(s) along with key active-site side chains to catalyze the hydrolysis of a wide variety of biologically important substrates, including carbohydrates, peptides, proteins, nucleotides, phosphodiesters, and xenobiotics. Interest in understanding the mechanisms of

these enzymes is based on the significance of the specific pathways and metabolites that are involved in the reactions, which has also made several metallohydrolases targets for drug development. Metal-dependent hydrolases use a combination of amino acid side chains and one or more transition metal ions to participate in proton transfer reactions and stabilization of intermediates that are necessary to efficiently facilitate hydrolytic reactions. Herein, we review the chemical and physical properties that make divalent metal ions suitable cofactors for metal-dependent hydrolases, and describe the roles that metal ions are able to fulfill in catalysis. These roles are discussed in the context of specific enzyme mechanisms, as several metal-dependent hydrolases have been well characterized using a variety of biochemical and biophysical techniques to probe the role(s) of specific groups in catalysis. These studies show that the most commonly used metal ion cofactor(s) for these reactions is Zn^{2+} followed by Fe^{2+}, Mn^{2+}, and Ni^{2+}, and that metal ions accelerate substrate hydrolysis by providing and/or enhancing the reactivity of the water nucleophile and stabilizing reaction intermediates. Additionally, this class of enzymes often employs amino acid side chains to function as general acid–base catalysts (GABC) to facilitate proton transfer reactions and/or to serve as electrostatic catalysts to stabilize reaction intermediates.

8.15.2   Metal Ion Cofactors

Metal ions play a variety of essential biological functions such as serving as cofactors for catalysis, structural stabilization, transport of molecules (i.e., oxygen), communication, and electron transfer. The properties of divalent transition metal ions make them well suited to participate in both hydrolytic and redox reactions. This review focuses on examining the mechanisms of enzymes that utilize metal ion cofactors to catalyze hydrolytic reactions, broadly defined as the metallohydrolases class of enzymes, including a discussion of the properties that make divalent metal ions suitable to function as cofactors for these reactions. Special emphasis is given to the metal ion Zn^{2+}, which emerges as the prototypical metal ion for the metallohydrolases owing to its: (1) ability to serve as a strong Lewis acid, (2) flexible coordination number (four to six ligands) and geometries (tetrahedral, octahedral), (3) rapid ligand exchange rates, and (4) insensitivity to redox reactions (d^{10} electron configuration). Although Zn^{2+} appears to be the best suited and most widely used cofactor for metallohydrolases, it is not alone in its ability to participate as an effective cofactor in these reactions. For example, Ni^{2+} (urease, URE), Mn^{2+} (arginase), and Fe^{2+} (peptide deformylase, PDF) all serve as native cofactors for metal-dependent hydrolases, whereas Co^{2+} is also known to participate as a competent cofactor in lieu of native divalent metal ion cofactors. For an in-depth discussion of the reactions using Cu, refer to Chapter 8.14. These findings demonstrate that several divalent metal ions have the physical and chemical properties necessary for supporting hydrolysis reactions. If a number of metal ions are effective at serving as cofactors for these enzymes, why and how does nature select a single metal ion, most commonly Zn^{2+}, as a cofactor for a specific metallohydrolase? The selection of specific metal ion(s) by enzymes for participation in hydrolysis reactions, as well as its effectiveness as a catalyst, is not simple and is dictated by a number of factors. Ultimately, the selection of a specific metal ion from those with similar desired properties may be determined by the relative availability of different metal ions, as dictated by factors such as overall abundance, water solubility, affinity (K_D^{eff}), effects of pH, oxygen sensitivity, and uptake. The properties of metal ions are the subject of several textbooks and is briefly summarized in the next section.[1–3]

8.15.2.1   Properties

8.15.2.1.1   Lewis acidity
Foremost, metal ions that catalyze hydrolysis reactions must be strong Lewis acids (electron acceptors). This property enables metal ions to facilitate hydrolytic reactions by enhancing the electrophilicity of the organic substrate and/or nucleophilicity of the water molecule, and allows hydrolysis reactions to proceed at, or around, neutral pH. Lewis acidity can be observed as a lowering of the pK_a values of bound ligands, such as a bound water molecule in hydrolysis reactions. The pK_a of a water molecule decreases substantially upon coordination by a metal ion (**Table 1**), allowing the water molecule to become deprotonated near neutral pH. Coordination of more than one metal ion to a protic ligand, such as the case with binuclear metallohydrolases,

Table 1 Select properties of metal ions

Metal ion	pK_a[a]	I_2 (eV)[b]	Ionic radii (pm)	Exchange rate (s^{-1})[c]	CN	Preferred geometry
Ca^{2+}	13.4	11.87	100	3×10^8	6	
Mg^{2+}		15.04	72	6×10^5	6	
Mn^{2+}	11.1	15.64	83	2×10^7	6	Octahedral > other
Fe^{2+}		16.18	78[d], 61[e]	4×10^6	6	Octahedral > 5-coordinate > tetrahedral
Fe^{3+}	2.2		65[d], 55[e]	2×10^2	6	Octahedral > other
Co^{2+}		17.06	65	3×10^6	6	Octahedral > tetrahedral > other
Ni^{2+}		18.17	69	4×10^4	6	Octahedral > other
Cu^{1+}					2	Linear > trigonal, tetrahedral
Cu^{2+}	10.7	20.29	73	1×10^9	4	Square planar > 5-coordinate > tetrahedral
Zn^{2+}	10	17.96	74	2×10^7	4	Tetrahedral > trigonal bipyramidal, octahedral

[a] pK_a of metal–water at 25 °C.
[b] Ionization potential of M to M^{2+}.
[c] Exchange rate of water from metal ion at 25 °C.
[d] High spin.
[e] Low spin.
Adapted from J. J. R. Frausto da Silva; R. J. P. Williams, *The Biological Chemistry of the Elements. The Inorganic Chemistry of Life*, 2nd ed.; Oxford University Press: New York, 2001. S. J. Lippard; J. M. Berg, *Principles of Bioinorganic Chemistry*; University Science Books: Mill Valley, CA, 1994. W. Kaim; B. Schwederski, *Bioinorganic Chemistry: Inorganic Elements in the Chemistry of Life*, 1st ed.; John Wiley & Sons Ltd: Chichester, 1994.

produces an even greater lowering of the pK_a. Although trivalent metal ions are more effective at lowering the pK_a values of protic ligands compared to their divalent analogues (i.e., Fe^{3+} versus Fe^{2+}; **Table 1**), divalent metal ions are more commonly used as cofactors for hydrolysis reactions. The Lewis acidity of divalent metal ions, as measured by ionization potential (I_2; **Table 1**), increases according to the order: $Ca^{2+} < Mg^{2+} < Mn^{2+} < Fe^{2+} < Co^{2+} < Cd^{2+} < Zn^{2+} < Ni^{2+} < Cu^{2+}$. Consequently, if Lewis acidity alone were the dominant factor in cofactor selection, one would expect metallohydrolases to be best served by a Cu^{2+} cofactor. However, Cu^{2+} is not typically found to be a hydrolytic cofactor, and substitution of known metal-dependent hydrolases with Cu^{2+} often leads to significantly lower reactivity. Therefore, as Lewis acidity is an important property for metal-dependent hydrolase cofactors, these findings suggest that factors other than Lewis acidity also contribute to the ability of metal ions to catalyze hydrolysis reactions and to the selection of the *in vivo* metal ion.

8.15.2.1.2 Metal ligation
Metal ion selection and catalytic efficiency are also influenced by preferences in metal ligation, including coordination number (CN), geometries, and ligand type (donor atom). The electronic configuration of the metal ion is an important determinant in coordination number and geometry of metal ions. Preferred coordination numbers and geometries for commonly used metal ions are shown in **Table 1** and **Figure 1**. There are differences amongst metal ions with the same oxidation state. For example, Zn^{2+} has a slight preference for tetrahedral geometry (four ligands), whereas Co^{2+}, Fe^{2+}, and Mn^{2+} favor higher coordination numbers and an octahedral geometry (six ligands). There are also differences observed for one metal at different oxidation states; Cu^{1+} prefers a linear geometry (two ligands), whereas Cu^{2+} preferentially adopts a square planar geometry (four ligands). Consequently, the ability of metal ions to adopt the geometric constraints dictated by the metal binding site throughout the course of the reaction (substrate binding, catalysis, product release) will be important determinants for metal selection and catalytic efficiency.

The overall thermodynamic stability of complexes formed between transition metal ions and ligands can vary greatly depending on the nature of the ligand–donor atom, and in general follows the order: O ligands < N ligands < S ligands. However, the relative thermodynamic stability of complexes for transition metals with the same donor atom is largely independent of the donor atom and follows the Irving–Williams series:
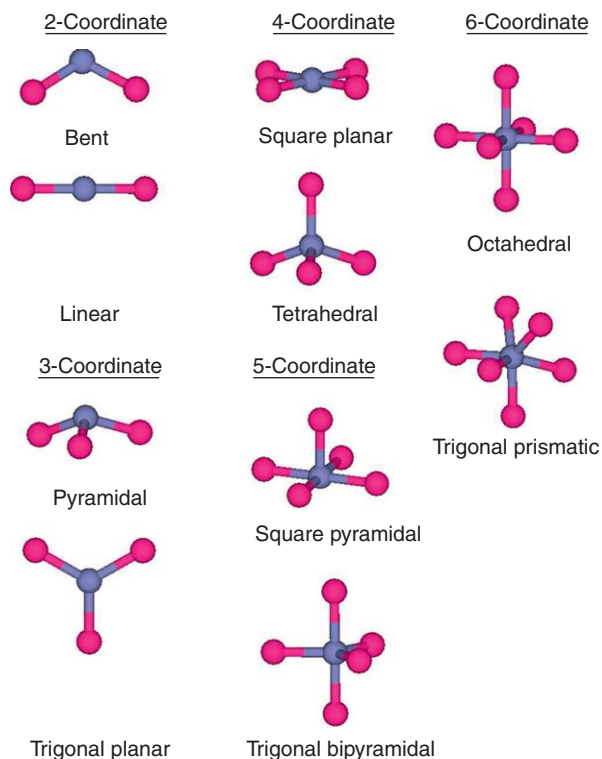
Figure 1 Commonly observed coordination geometries for metal ions with two to six ligands. The figure was created by using the Accelrys DS Visualizer program.

$Ca^{2+} < Mg^{2+} < Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}$, which is related to the decrease in ionic radii across the series that leads to stronger metal ligand bonds. Preferences in metal ligation defined by differences in ligand–donor atoms are also described by the hard–soft acid–base (HSAB) concept. This concept focuses on the polarizability of the metal ion, which is approximated using the ionic potential (electrostatic binding) as defined by z/r (z = charge, r = ionic radius) and ionization potential (I_n) of M to M^{n+} (electron acceptor power). Polarizability reflects the ability of ions to experience a z shift in their electron shell through interaction with a coordination partner. The term 'hard' refers to small ions that are fairly nonpolarizable species (e.g., Mg^{2+}, Ca^{2+}, Mn^{2+}, Fe^{3+}), whereas 'soft' refers to larger, more polarizable species (e.g., Cd^{2+}, Cu^+). In terms of metal–ligand coordination, species prefer to interact with partners of the same type – 'hard' metal ions prefer 'hard' ligands (i.e., O ligands) and form a highly ionic bond, whereas 'soft' metal ions prefer 'soft' ligands (i.e., S ligands) and form a partially covalent bond. Metal ions of intermediate hardness (e.g., Zn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}) prefer to coordinate intermediate ligands (i.e., N ligands), and appear to be the preferred cofactors for metal-dependent hydrolases. Consequently, the most common ligands for catalytic metal ions observed for the metal-dependent hydrolases are the side chains of His > Asp/Glu ~ Cys for protein ligands and water is the most commonly observed nonprotein ligand. Metal ligation may also be used to distinguish between structural and catalytic metal ion binding sites. For example, structural zinc-binding sites, such as zinc fingers, typically contain four protein ligands (Cys > His > Asp/Glu), whereas catalytic zinc-binding sites, which also contain four ligands, typically have three protein ligands (His > Asp/Glu, Cys) and a water molecule as the fourth ligand. The presence of a water ligand is typically the distinguishing feature of catalytic metal ions, with the exception of enzymes that catalyze electron transfer reactions. Finally, in situations where there is a cavity of a specific size, it is possible that the ionic radii of the metal ion will be a determinant in cofactor selection (**Table 1**; $Mg^{2+} < Cu^{2+} < Zn^{2+} < Co^{2+} < Ni^{2+} < Fe^{2+} < Mn^{2+} < Cd^{2+} < Ca^{2+}$). However, typically other factors such as the nature of the ligand–donor atoms and stereochemistry are proposed to play larger roles in metal ion selection.

8.15.2.1.3   Ligand exchange

Metal ions that serve as cofactors for hydrolytic reactions are often required to facilitate substrate(s) binding and product release, and thus rapid ligand exchange rates are a desirable property. The exchange rates of water from metal ions are listed in **Table 1**. The exchange rates or lability of other metal–ligand complexes follow the trend observed for the metal–water complexes: $Fe^{3+} < Ni^{2+} < Mg^{2+} < Co^{2+}$, $Fe^{2+} < Zn^{2+}$, $Mn^{2+} < Ca^{2+} < Cu^{2+}$. These rates decrease further upon binding by a multidentate chelating ligand, and when the metal binding site is found in protein cores that are inaccessible to solvent. These data indicate that the exchange of water from divalent metal ions proceeds rapidly enough to facilitate the observed catalytic reactions in metallohydrolases.

8.15.2.1.4   Electronic structure

Several transition metals have more than one biologically relevant oxidation state due to partial occupancy of the $3d$ orbitals, including manganese (2+, 4+), iron (2+, 3+), cobalt (2+, 3+), and copper (1+, 2+). Differences in the oxidation state, as well as spin state (i.e., the way electrons are packed into $3d$ orbitals), of a given metal ion alter properties such as Lewis acidity, ionic radii, polarizability, and ligation preferences (**Table 1**). Changes in ligand–donor atoms and stereochemistry around the metal ion can also greatly affect the electron transfer potential of the metal ion. Consequently, the redox status of the environment can alter the properties of redox-sensitive metal ions (e.g., Fe, Cu), thereby affecting enzyme activity. Only metal ions with a full shell of electrons (d^{10} configuration), such as Zn^{2+}, will be able to tolerate changes in the redox environment without altering catalytic efficiency. This may be an important feature for enzymes that perform critical functions, including enzymes that are necessary during oxidative stress.

8.15.2.2   Catalytic Roles

The chemical and physical properties mentioned above make divalent metal ions well suited to participate as a catalyst in hydrolysis reactions. Specifically, these properties enable metal ion cofactors to assume one or more of the following roles in catalysis (**Figure 2**): (1) facilitate substrate binding (water and organic substrate);
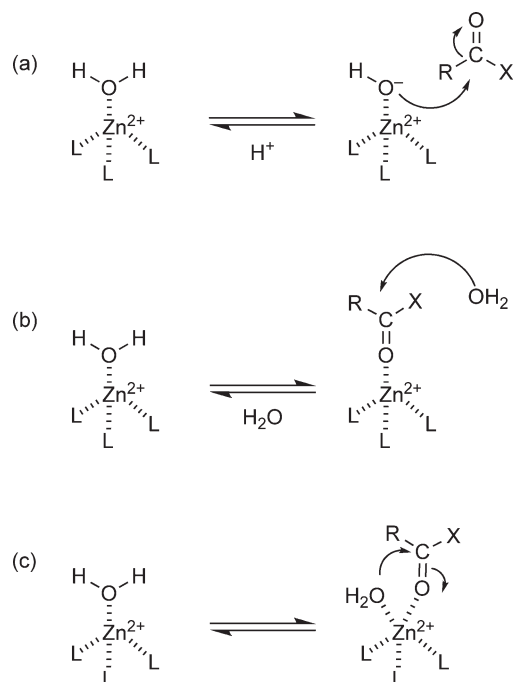
Figure 2   Common roles for metal ions in catalysis. (a) Lewis acid, (b) electrostatic catalyst, and (c) template effect.

(2) gathering/template effects; (3) function as an electrostatic catalyst (carbonyl polarization and transition state stabilization); (4) function as a Lewis acid to lower the pK_a of metal–water; and (5) stabilize the formation of the leaving group. (1) Substrate binding: Organic substrates for metallohydrolases typically contain carbonyl groups (e.g., amide bond) or other heteroatom centers (e.g., phosphodiesters) that are common ligands for metal ions. Consequently, these substrates are often able to coordinate with the catalytic metal ion, which serves to enhance the affinity of this substrate for the enzyme and properly align the substrate for attack by water (**Figure 2(b)**). The loss of a water molecule is often a pre- or corequisite for binding the organic substrate, and therefore the coordinated water ligand(s) in the resting enzyme must be readily exchangeable. For metal ions with flexible coordination numbers and geometry, such as Zn^{2+}, prior dissociation of water may not be required for organic substrate binding. (2) Gathering/template effects: Metal ions with higher coordination numbers (i.e., 5–6) are able to coordinate the water and organic substrates simultaneously for the reaction, thereby placing the two substrates in close proximity to one another (**Figure 2(c)**). This effectively allows the reaction to further approach that of an intramolecular reaction, thereby facilitating catalysis. (3) Electrostatic catalyst: Since the organic substrates for these enzymes are often relatively weak electrophiles, catalytic metal ions often participate by using their Lewis acidity to polarize the carbonyl group on the substrate (through coordination with the metal ion), thereby enhancing its electrophilicity and/or using their ionic charge to stabilize the negatively charged oxyanion intermediate and flanking transition states (**Figure 2(b)**). (4) Metal–water: The strong Lewis acidity of metal ions also serves to lower the pK_a of coordinated water molecule(s), thereby increasing the amount of hydroxide present at neutral pH and/or facilitating removal of a proton by a general base (**Figure 2(a)**). (5) Leaving group stabilization: Finally, in instances where negatively charged products are generated, the positively charged metal ion can function to effectively stabilize these leaving groups and facilitate catalysis. The chemical properties of metal ions make them well suited to serve as cofactors in hydrolytic reactions, as illustrated by the variety of catalytic roles that are described here.

We clarify some general points regarding metal ligation by metal-dependent hydrolases as they pertain to the subsequent discussion. In general, enzymes that require metal ion cofactors are characterized by a loss of catalytic activity upon incubation with chelating agents, although this does not distinguish between metal ion cofactors that play structural and catalytic roles. This review focuses on enzymes that use metal ion cofactors that serve catalytic roles. A catalytic metal ion is most easily identified by observation of direct coordination with the substrate. Metal-dependent hydrolases are broadly defined as being either mononuclear (one metal ion) or binuclear (two metal ions) in nature, referring to the number of metal ions required for maximal activity. Mononuclear metallohydrolases are typically both activated and inhibited by divalent metal ions. Occupancy of the catalytic metal ion binding site by the appropriate transition metal ion, most commonly Zn^{2+}, leads to an increase in catalytic activity, whereas occupancy of a neighboring inhibitory metal ion binding site results in decreased catalytic activity. This inhibition is usually attributed to the finding that catalytically essential side chains, as well as the nucleophilic water molecule, are coordinated with the inhibitory metal ion. Consequently, we focus our attention on enzymes that have been characterized as mononuclear in nature using experiments that examine catalytic activity as a function of metal ion:enzyme ratio (i.e., metal titration experiments), and on results obtained through experiments carried out under stoichiometric metal:enzyme conditions. For binuclear metallohydrolases, the nature of these metal centers produces an active site with altered properties compared to their mononuclear counterparts, and therefore catalysis by these enzymes occurs through alternative reaction mechanisms. A distinguishing feature of the metal centers found in these enzymes is the presence of bridging ligands, such as a water molecule and/or carboxylate side chain. Interpretation of metal titration experiments with binuclear metalloenzymes is much more complex, as inherent differences in metal affinity may exist between the two sites and mixed metal ligation can occur. The identification of ligands for the metal center is of great interest in the characterization of metallohydrolases, and is often probed using a variety of biochemical and biophysical approaches, often done in conjunction with sequence alignment results. Biochemically, mutation of a side chain (to Ala) that is a ligand for a catalytically essential metal ion typically leads to a $\sim 10^3$- to 10^4-fold loss of activity and a significantly decreased zinc affinity. Biophysical characterization of the catalytic metal centers is often carried out using experiments such as extended X-ray absorption fine structure (EXAFS), which provides information regarding N/O versus S/Cl ligation. Together results from biochemical and biophysical studies can often accurately predict the enzyme side chains that are responsible for metal coordination; however, ultimately, metal ligands can only be unambiguously identified using high-resolution

crystal structures. Therefore, we focus our discussion on metal centers confirmed through structural studies. The metal ion centers of metallohydrolases are primary targets for the development of therapeutic inhibitors, as enzyme inhibitors typically contain a group that coordinates the catalytic metal ion(s) to both inhibit enzyme activity, and to provide enhanced potency.[4-10] Consequently, establishing the identity of the native cofactor for a specific enzyme is important for the development of inhibitors as effective therapeutic agents and therefore the focus of many studies.

8.15.3   Mononuclear Metallohydrolases

Metal-dependent hydrolases are a group of enzymes involved in the metabolism of structurally diverse compounds, including carbohydrates, proteins, nucleotides, and other biologically important molecules. Since several of these enzymes do not fall into larger enzyme families, we have chosen to group enzymes together according to the class of molecule(s) metabolized for our discussion. Metal-dependent hydrolases are broadly defined as being either mononuclear or binuclear in nature, referring to the number of metal ions required for maximal activity. We start with a review of mononuclear metal-dependent hydrolases – enzymes that use a single metal ion cofactor for catalysis.

8.15.3.1   Carbohydrate Metabolism

Metal-dependent deacetylases (MDDs) are involved in the metabolism of several important carbohydrates (**Figure 3**), including the enzymes UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase (LpxC; lipopolysaccharide (LPS) biosynthesis), N-acetylglucosamine-6-phosphate deacetylase (NagA; peptidoglycan and techoic acid biosynthesis) and N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase (MshB; mycothiol biosynthesis). Interestingly, these structurally diverse enzymes act upon the same N-acetyl-glucosamine moiety of distinct substrates with seemingly remarkable specificity. Since these enzymes are involved in biosynthetic pathways that produce essential biological molecules, all three enzymes are targets for drug development. Consequently, information regarding the catalytic mechanisms and recognition properties of these enzymes will be critical for the development of inhibitors that can function as antibiotics.

8.15.3.1.1   *UDP-3-*O*-(*R*-3-hydroxymyristoyl)-*N*-acetylglucosamine deacetylase*
LPS molecules make up the outer membranes of Gram-negative bacteria, and serve as a barrier to prevent the entry of hydrophobic and negatively charged molecules.[11] Lipid A, also known as endotoxin, is the hydrophobic anchor of LPS and also the portion of LPS responsible for stimulating the immune system in septic shock.[11] Lipid A is essential for the viability of Gram-negative bacteria, and is synthesized from UDP-N-acetylglucosamine in a 10-step pathway.[11] LpxC catalyzes the committed, and second overall, step in the biosynthesis of lipid A – the hydrolysis of UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine to form UDP-3-O-(R-3-hydroxymyristoyl)-glucosamine and acetate (**Figure 3(a)**).[12] Consequently, LpxC is a target for the development of antibiotics for the treatment of Gram-negative bacterial infections.

The high-resolution structure of LpxC has been solved by both X-ray crystallography[13] and NMR spectroscopy,[14] revealing that the overall fold of LpxC is unrelated to any other metal-dependent hydrolase (**Figure 4(a)**) and that this enzyme contains a unique zinc-binding motif. Topologically, the active site of LpxC is located at the interface of two homologous domains, each consisting of two α-helices sandwiched by a five-stranded β-sheet, and contains a binuclear zinc metal center. Zn_1 coordinates His79, His238, Asp242, and a water molecule (bridging), whereas Zn_2 coordinates Glu78, His265, a water molecule (bridging), and a myristate/palmitate ligand. LpxC activity is inhibited by incubation with chelating agents (ethylenediamine-tetraacetic acid, EDTA; dipicolinic acid, DPA), and restored by the addition of divalent metal ions (Zn^{2+}, Co^{2+}, Ni^{2+}, and Mn^{2+}) demonstrating that at least one metal ion is important for catalysis.[15] Metal titration experiments with Zn^{2+} indicate that LpxC is maximally active with one metal ion and the addition of excess metal ions inhibits catalytic activity, confirming that LpxC is a mononuclear metallohydrolase.[15] Mutagenesis,[16] EXAFS,[17] and structural studies[13] indicate that Zn_1 represents the catalytic metal ion binding
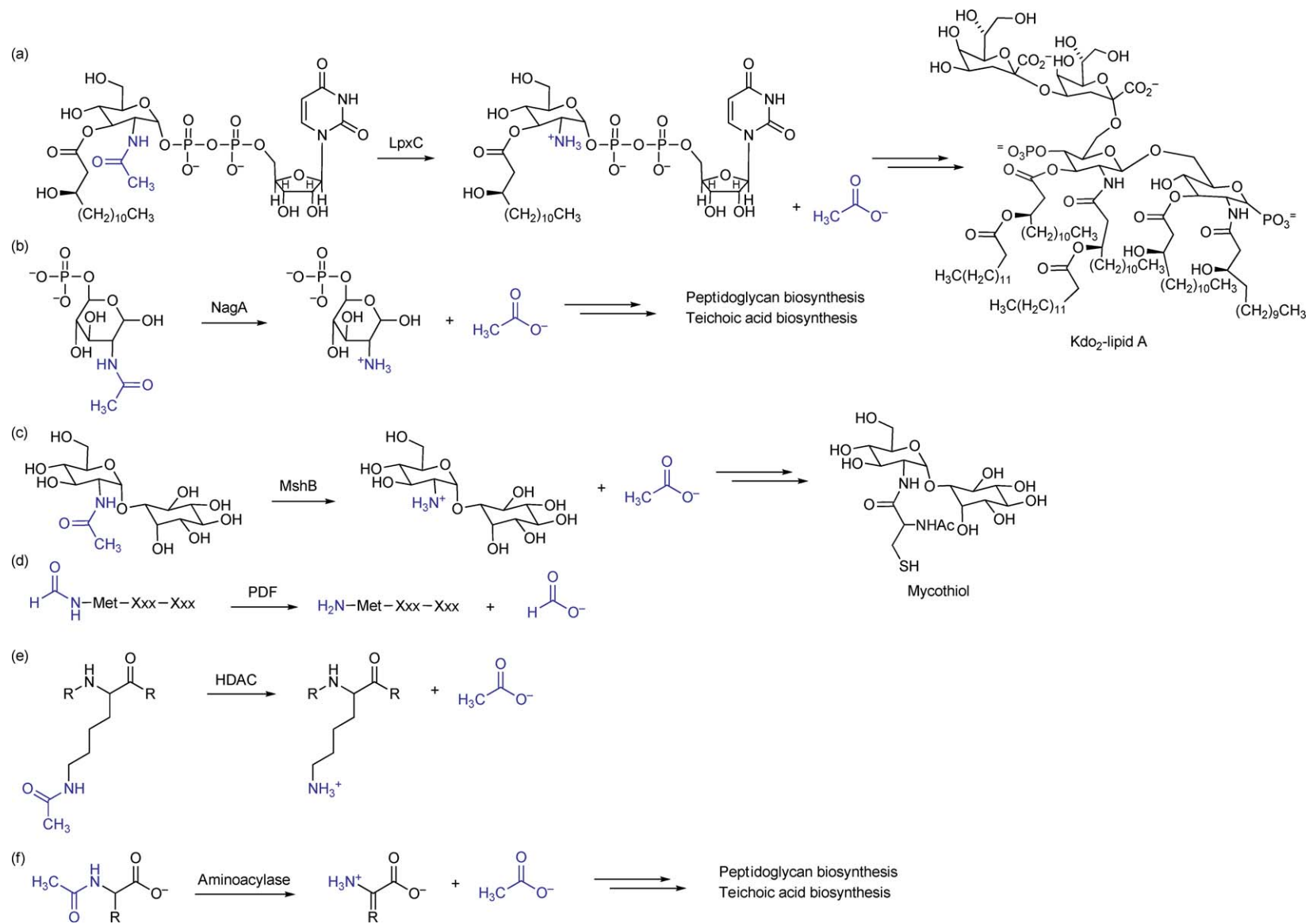
Figure 3 Hydrolytic reactions catalyzed by various metal-dependent deacetylases with the leaving group shown in blue.
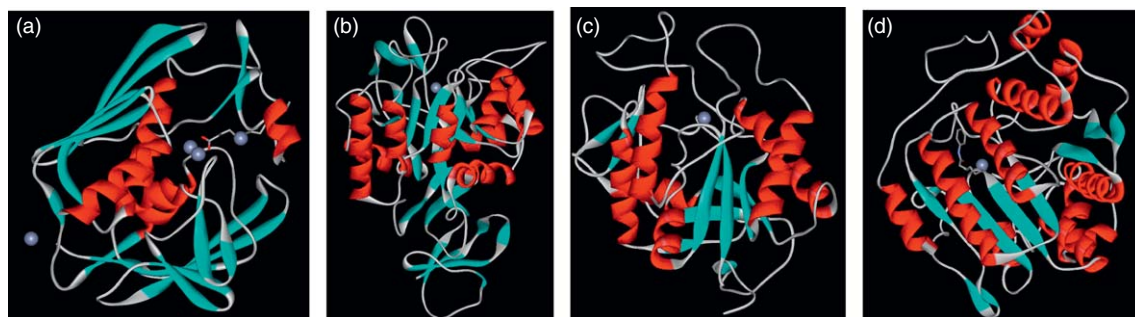
Figure 4 Representative structures of various folds for metal-dependent deacetylases. For clarity, only a monomer is shown for each enzyme with the metal ions shown as gray spheres. (a) LpxC (PDB 1P42), (b) NagA (PDB 2P50), (c) MshB (PDB 1Q74), and (d) HDAC (PDB 1C3S).

site (tetrahedral geometry), whereas Zn_2 represents the inhibitory metal binding site (in EcLpxC the side chain of Cys63 is also a ligand for the inhibitory metal ion[18]). The finding that LpxC isolated from *Escherichia coli* (Ec) copurifies with Zn^{2+} led to its classification as a zinc-dependent deacetylase. However, recently it has been suggested that LpxC from *E. coli* (EcLpxC) may be a Fe^{2+}-dependent enzyme *in vivo* (M. Hernick, S. G. Gattis, and C. A. Fierke, unpublished results). This is supported by the findings that the activity of Fe^{2+}-EcLpxC > Zn^{2+}-EcLpxC and EcLpxC copurifies with Fe^{2+} under anaerobic conditions. Metal ion titration experiments indicate that Fe^{2+}-EcLpxC also functions as a mononuclear enzyme, and therefore probably proceeds through a similar reaction mechanism.

The high-resolution structures of LpxC also identify several side chains in close proximity to the catalytic zinc ion that could participate in the catalytic reaction through GABC (Glu78, Asp246, and His265) either through a single bifunctional GABC[13,14] or a GABC pair mechanism[19,20] (**Figure 5**). The mechanism of LpxC from *E. coli* and *A. aeolicus* (Aa) has been investigated using biochemical and biophysical studies. The LpxC-catalyzed reaction exhibits a bell-shaped dependence on pH (k_{cat}/K_M), indicating that at least two ionizations are important for catalytic activity.[20,21] Mutagenesis and kinetic experiments were used to identify the sources for the observed ionizations.[20–22] Results from these experiments indicate that the pK_{a1} of ∼6, describing a favorable deprotonation, represents ionization of Glu78. Furthermore, solvent isotope effect experiments suggest that Glu78 functions as a general base catalyst (GBC) in the reaction, consistent with the assignment of pK_{a1} as Glu78. The assignment of the pK_{a2} of ≥8 reflecting a favorable protonation, which often describes a GAC, is more challenging. One proposed source for pK_{a2} was the side chain of His265 for the following reasons: it is located near the catalytic zinc ion, loss of this side chain leads to a >10^3-fold decrease in activity, and the pK_a of this side chain is ∼7.5 (determined using NMR spectroscopy for an AaLpxC mutant[23] and product binding studies for EcLpxC[22]), in close agreement with the observed value of 7.9 for pK_{a2} in the AaLpxC pH profile. However, mutation of His265 and other ionizable active-site side chains (e.g., Thr191, Lys239, Asp246) to Ala does not substantially alter the pH rate profile suggesting that none of these active-site side chains are responsible for pK_{a2} in the WT enzyme. Substitution of the enzyme with various divalent metal ions alters the value of pK_{a2}, suggesting that pK_{a2} reflects ionization of the metal-bound water or a group that forms a hydrogen bond with the metal–water. Functional studies of the E78A/H265A double mutant indicate that Glu78 and His265 interact in an anticooperative manner to bind products and have partially additive effects on activity, which are consistent with Glu78 and His265 acting as a GABC pair to facilitate hydrolysis.[20,22] The side chain of Asp246 is part of a His–Asp charge relay with His265, and therefore the importance of Asp246 for catalytic activity is probably mediated through His265. The crystal structure of a LpxC•cacodylate complex[20] suggests that the side chains of Glu78 and His265 form a hydrogen bond with the same oxygen atom of the tetrahedral intermediate, consistent with the GABC pair proposal. This structure also suggests a role for Thr191 in stabilization of the oxyanion intermediate and flanking transition states in the hydrolytic reaction, which is further supported by results from mutagenesis experiments. Together these findings suggest the mechanism shown in **Figure 5**, wherein Glu78 acts as a GBC to activate the metal-bound water for attack of the carbonyl group on the substrate. The resulting oxyanion intermediate is stabilized by the catalytic zinc ion, and the side
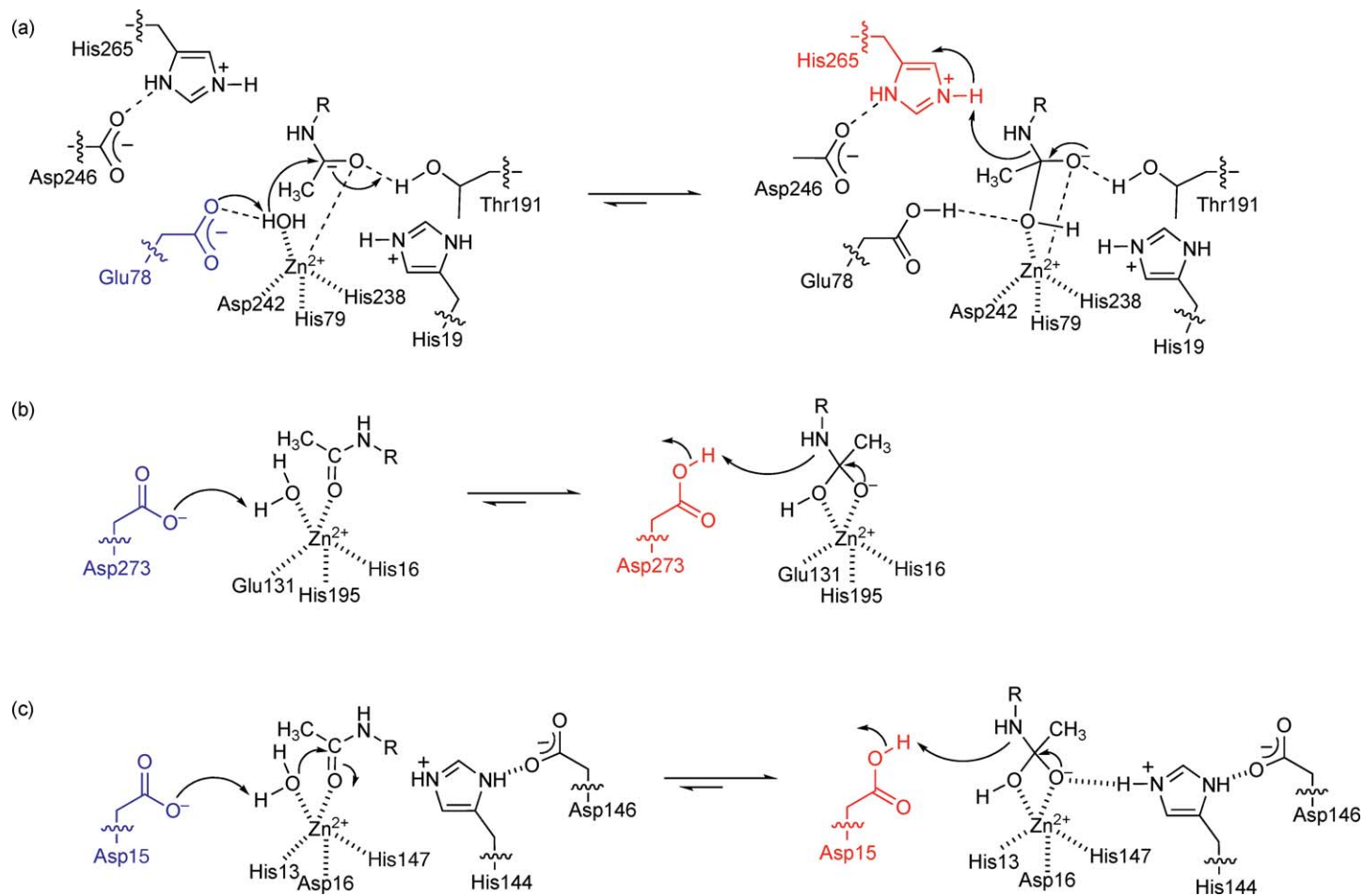
Figure 5 Proposed catalytic mechanisms for (a) LpxC, (b) NagA and (c) MshB with general base catalyst shown in blue and general acid catalyst shown in red.

chain of T191. Protonation of the amine leaving group by His265 facilitates breakdown of the tetrahedral intermediate and generation of the reaction products. A recent theoretical examination of the LpxC catalyzed mechanism has been done using DFT (density functional theory) calculations and the resulting data support the proposal that LpxC uses a GABC pair mechanism.[24]

8.15.3.1.2   N-*acetylglucosamine-6-phosphate deacetylase*

The cell wall surrounding Gram-negative and Gram-positive bacteria is made up of peptidoglycan, a well-established target for antibiotics (β-lactams, cephalosporins). The enzyme NagA catalyzes the deacetylation of *N*-acetyl-D-glucosamine-6-phosphate to form D-glucosamine-6-phosphate and acetate (**Figure 3(b)**), an important step in cell wall recycling and *N*-acetyl-glucosamine metabolism.[25] *N*-Acetyl-glucosamine is an essential metabolite in several critical microbial pathways, including peptidoglycan, teichoic acid, and LPS biosynthesis. Therefore, inhibitors of NagA have the potential to function as antibiotics for the treatment of both Gram-positive and Gram-negative bacterial infections.

NagA is a member of the amidohydrolase superfamily of enzymes, which also includes the enzymes URE, phosphotriesterase (PTE), adenosine deaminase (ADA), and cytosine deaminase (CDA).[26,27] Structurally, members of this enzyme superfamily share a common $(\beta/\alpha)_8$-barrel fold (**Figure 4(b)**) containing a mono-nuclear or binuclear metal ion center in the active site. These enzymes also share a common HXH signature motif and catalytic Asp side chain that are required for metal binding and catalysis, respectively. Occupancy of the metal centers varies amongst members of this superfamily with either Zn^{2+}, Fe^{2+}, or Ni^{2+} found at the M_α, M_β or $M_\alpha M_\beta$ sites, where M_α and M_β reflect the more buried and solvent accessible metal ions, respectively. Crystal structures of NagA from *E. coli*,[28,29] *Bacillus subtilis*[30], and *Thermatoga maritima* (Tm) have been solved, revealing diversity amongst the metal ion centers of the NagA active sites.[26] NagA from *E. coli* (EcNagA) contains a mononuclear metal ion center (M_β) that is occupied by a bound Zn^{2+}, whereas NagA from *B. subtilis* (BsNagA) contains a binuclear center ($M_\alpha M_\beta$) with two bound Fe^{2+} ions. Although the side-chain ligands for the binuclear metal center are conserved in the *T. maritima* NagA (TmNagA), this center is occupied by a single Fe^{2+} ion (M_β). EcNagA activity is abolished by treatment with chelating agents (1,10-phenanthroline, DPA, EDTA), and restored upon the addition of divalent metal ions, confirming that the bound zinc ion observed in the crystal structure is essential for catalysis.[28,31] Furthermore, metal titration experiments of apo-NagA with Zn^{2+} confirm that both EcNagA and TmNagA are maximally active with one metal ion; addition of two to three equivalents of Zn^{2+} to apo-NagA does not alter activity relative to that observed for the 1:1 complex indicating that a single metal ion is required for catalysis.[29,31] The metal requirements for BsNagA activity have not been probed; however, if BsNagA does have different preferences with respect to the number and identity of metal ions bound at the metal center, the reactions catalyzed by these enzymes would be required to proceed through different mechanisms. These findings highlight the fact that the metal cofactor used by a given metallohydrolase can vary depending on the species and, potentially, the growth conditions; this result has been observed for other enzymes, including PDF,[32] methionine aminopeptidase (MetAP2),[33] glyoxalase I,[34] and possibly LpxC (see above) and histone deacetylases (HDACs).[35]

Mechanistic studies to date have focused on the mononuclear EcNagA and are reviewed here.[31] Since TmNagA is also a mononuclear enzyme and the bound iron shares a similar coordination sphere (Glu115, His176, His197, and a water molecule) to EcNagA, the reactions catalyzed by these enzymes likely proceed using a similar mechanism. Characterization of EcNagA as a zinc-dependent deacetylase is based on the findings that Zn^{2+} is the most efficient cofactor and this enzyme copurifies with Zn^{2+} under aerobic conditions. The catalytic zinc ion in EcNagA coordinates with the side chains of Glu131, His195 and His216, and a water molecule (a known zinc-binding motif) with a tetrahedral geometry. NagA is also activated by other metal ions according to the order (k_{cat}/K_M): Zn^{2+}, $Co^{2+} > Mn^{2+}$, $Cd^{2+} > Fe^{2+} > Ni^{2+}$. A potential direct interaction between the catalytic metal ion and the carbonyl group on the substrate was probed by comparing the activity of Zn(II)- and Cd(II)-NagA with the substrates *N*-acetyl-D-glucosamine-6-phosphate and *N*-thioacetyl-D-glucosamine-6-phosphate. Hydrolysis of the thioacetyl substrate is ~10-fold faster with Cd(II)-NagA, whereas hydrolysis of the acetyl substrate is faster with Zn(II)-NagA. These findings are consistent with a direct coordination of substrate carbonyl to the metal ion as expected based on the HSAB preferences of these metal ions. The pH dependence of NagA activity (k_{cat}/K_M) is bell-shaped with pK_a values of 6.4 (representing two ionizations; pK_{a1a} and pK_{a1b}) and 9.3 (pK_{a2}). The sources of these ionizations were probed using a

combination of mutagenesis and kinetic experiments with various substrate analogues. Replacement of the phosphate group at the sixth-position of the substrate with sulfate eliminates two of the ionizations (pK_{a1a} and pK_{a2}) observed in the pH profile for WT. Consequently, it is proposed that pK_{a1a} reflects protonation of the phosphate group on the substrate and pK_{a2} reflects ionization of an enzyme side chain that interacts with the substrate, such as Lys139 or Tyr223. The source of pK_{a1b} describing a favorable deprotonation is likely the zinc–water or an active-site side chain functioning as a GBC (i.e., Asp273). Mutagenesis experiments indicate that the side chains of His143 (6000-fold decrease), Asp273 (>5000-fold), and His251 (400-fold decrease) are all important for catalytic activity. Significant efforts have been made to decipher the steps described by k_{cat}/K_M and k_{cat}. In contrast to LpxC, there is no solvent isotope effect observed ($[k_{cat}/K_M]^H/[k_{cat}/K_M]^D = 1.02$, $[k_{cat}]^H/([k_{cat}]^D = 1.1$) for NagA, suggesting that proton transfer is not important for the rate determining step of the reaction. Hydrolysis of a trifluoroacetyl substrate analogue is faster (both k_{cat}/K_M and k_{cat}) compared to the natural substrate acetyl analogue, consistent with these steps reflecting chemistry. Similarly, solvent viscosity experiments indicate that k_{cat} and k_{cat}/K_M do not reflect product dissociation or substrate association, respectively. Together these results suggest that the reaction is limited by either formation or breakdown of the tetrahedral intermediate. Structural data may help to clarify the catalytic roles of side chains shown to be critical for activity.[29] Importantly, Asp273 shares a hydrogen bond with the zinc-bound water, consistent with a role of this side chain as a GABC. The interaction of His143 with a phosphonate analogue (tetrahedral intermediate mimic) suggests that this side chain is positioned appropriately to polarize the substrate and/or stabilize the tetrahedral intermediate. Consequently, NagA is proposed to function through the mechanism shown in **Figure 5(b)**,[31] wherein the organic substrate binds to the catalytic metal ion thereby polarizing the carbonyl group. Next, Asp273 functions as a GBC to activate the zinc–water for nucleophilic attack of the carbonyl group. The resulting oxyanion intermediate is stabilized by the catalytic zinc ion and His143, prior to protonation of the amine leaving group by the now protonated Asp273. Breakdown of the tetrahedral intermediate is followed by release of products to regenerate the active form of the enzyme.

8.15.3.1.3   *N-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase*

Mycothiol (MSH) is an intracellular reducing agent used by mycobacteria (e.g., *Mycobacterium tuberculosis*) to protect against oxidative damage, a role similar to that fulfilled by glutathione in eukaryotes and other bacteria; therefore, MSH is essential for growth of pathogenic mycobacteria.[36–38] Additionally, MSH is involved in the metabolism of xenobiotics, suggesting that this molecule may modulate bacterial sensitivity to antibiotics.[39] MSH is synthesized from 1-L-inositol-1-phosphate and UDP-*N*-acetylglucosamine in five steps.[40] MshB catalyzes the committed, and third overall, step in MshB – the conversion of *N*-acetyl-glucosamine-inositol to form glucosamine-inositol and acetate (**Figure 3(c)**).[37] Consequently, MshB is a target for the development of antibiotics for the treatment of mycobacterial infections such as tuberculosis and leprosy.

The crystal structure of MshB from *M. tuberculosis* has been solved, revealing an overall fold (α/β Rossman fold) similar to lactate dehydrogenase and unlike previously known zinc hydrolase enzymes (**Figure 4(c)**).[41] The active site of MshB contains one bound zinc ion that coordinates with the side chains of His13, Asp16, His147, and two water molecules. This bound zinc ion was confirmed to be important for catalytic activity as MshB activity is inhibited by treatment with a metal chelator (1,10-phenanthroline) and restored upon addition of divalent metal ions (Zn^{2+}, Co^{2+}, Ni^{2+}, and Mn^{2+}) to apo-MshB.[42] The finding that MshB copurifies with one equivalent of zinc (aerobic conditions) has led to its characterization as a mononuclear zinc metalloenzyme. Mechanistic inferences for MshB have been proposed based on the similarity of the MshB active site to other known metallohydrolases, including both single bifunctional GABC and GABC pair mechanisms (**Figure 5**).[19,41] In the first mechanism (**Figure 5(c)**), the conserved Asp15 side chain serves as a bifunctional GABC, whereas stabilization of the tetrahedral intermediate is afforded by the catalytic zinc ion and His144 (similar to NagA mechanism). The second mechanism involves the use of a GABC pair, wherein the conserved Asp15 functions as a GBC the conserved His144/Asp146 charge relay functions as a GAC, and stabilization of the tetrahedral intermediate is provided by the catalytic zinc ion (similar to LpxC). Additional studies are needed to distinguish between these two possible mechanisms.

Recently, the crystal structure of BcZBP (*Bacillus cereus* zinc-binding protein, ZBP Bc1534 gene), a putative virulence gene from *B. cereus*, with homologues in *B. anthracis*, has been solved and shown to be structurally homologous to MshB.[43] The active site of Bc1534 is strikingly similar to both MshB and LpxC, and contains a

bound zinc ion that coordinates with His12, Asp15, His113, and an acetate molecule. Although the function and the substrate of this protein have yet to be identified, BcZBP was shown to have N-acetylglucosamine deacetylase activity, which suggests that it may be another MDD.[43] These findings may imply that MshB and BcZBP belong to a larger unidentified superfamily of evolutionarily related enzymes.

8.15.3.2 Protein Metabolism

Metal-dependent hydrolases are involved in the metabolism of proteins through the hydrolysis of 'backbone' amide bonds (Note: here the term 'backbone' is used to refer to non-side chain reactions), and the substrates for these enzymes include proteins (metalloproteases), peptides (PDF), and amino acids (D-aminoacylase). There are both mononuclear and binuclear metallohydrolases involved in protein metabolism; this section reviews the mechanisms of the mononuclear enzymes. Biologically, these are the most functionally diverse group of metallohydrolases, playing important roles in cell growth and differentiation, inflammation, bone remodeling, hormone processing, cardiovascular function, and protein maturation. Consequently, metallohydrolases that metabolize protein and peptide substrates are targets for the treatment of cancer, cardiovascular disease, inflammation, and infectious diseases.

8.15.3.2.1 Carboxypeptidase A

Metalloproteases catalyze the hydrolysis of amide bonds in peptides and/or proteins (**Figure 6(a)**), and members of this family include the enzymes carboxypeptidase A (CPA), thermolysin, matrix metalloproteases (MMPs), angiotensin converting enzyme (ACE), anthrax lethal factor, and snake venom protease.[44–47] These enzymes are broadly defined as either endopeptidases or exopeptidases depending on the reaction cleavage site. Exopeptidases (e.g., CPA) cleave a terminal peptide bond, whereas endopeptidases (e.g., thermolysin) cleave an internal peptide bond. Metabolically, metalloproteases play a role in the degradation of proteins and are involved in modulation of cell growth, inflammation, immunity, and hormone processing. Consequently, several of these enzymes are targets for drug development. Structurally, the metalloproteases belong to the zinc α,β-hydrolase superfamily of enzymes, which also includes aminopeptidases, aminoacylases, desucciny-lases, and ornithine deacetylase.[48,49] Numerous structures have been solved for members of this superfamily, indicating that the overall fold of these proteins is characterized by an eight-stranded β-sheet that is sandwiched by six α-helices (**Figure 7(a)**); however, the active sites of these proteins do not overlay and there are differences in side-chain identities. The catalytic zinc ion of the metalloproteases typically coordinates with a His_2-Glu-H_2O or His_3-H_2O metal polyhedra and described by a HEXXH + (H/E) sequence motif, wherein the metal ligands are the two His and a His/Glu, whereas the first Glu is a catalytically important residue (GABC).[46]

   CPA is the prototypical zinc metalloprotease, and has been extensively characterized using mutagenesis, kinetics experiments, and biophysical studies that have previously been reviewed.[45,47,50] CPA activity is inhibited by metal-chelating agents (EDTA), which is reversed following the addition of divalent metal ions (e.g., Zn^{2+}, Co^{2+}), confirming that the bound metal ion is catalytically important.[51,52] In contrast to EcLpxC, Fe^{2+}-CPA is actually less active than Zn^{2+}-CPA.[51] Metal titration experiments indicate that CPA is maximally active with a single bound zinc ion; addition of excess zinc ions inhibits catalytic activity.[53,54] These findings coupled with the finding that CPA copurifies with Zn^{2+} led to its characterization as a zinc-dependent enzyme. The zinc-binding site of CPA has been extensively characterized using a variety of biophysical techniques, including X-ray crystallography and XAFS studies.[55,56] These studies indicate that the zinc ion in the resting enzyme coordinates with His69, Glu72, His196, and a water molecule in a tetrahedral geometry. Mutagenesis and kinetic studies have been used to probe the mechanism of CPA.[45,50] Similar to LpxC and NagA, the activity of CPA exhibits a bell-shaped dependence on pH with pK_a values of ~6–6.5 and 9, whereas mutagenesis experiments indicate that the side chains of Glu270 and Arg127 are essential for catalysis. These studies also suggest that Glu270 is responsible for pK_{a1} and the zinc–water reflects pK_{a2}. Consequently, CPA is proposed to function through the mechanism outlined in **Figure 8(a)**,[45,50] and uses a single bifunctional GABC that resembles the mechanism described above for NagA. The amide bond of the protein substrate coordinates with the catalytic zinc ion, which serves to polarize the carbonyl group and enhance the electrophilicity of the substrate. Next, Glu270 acts as a GBC to activate the zinc-bound water for attack on the carbonyl group of the
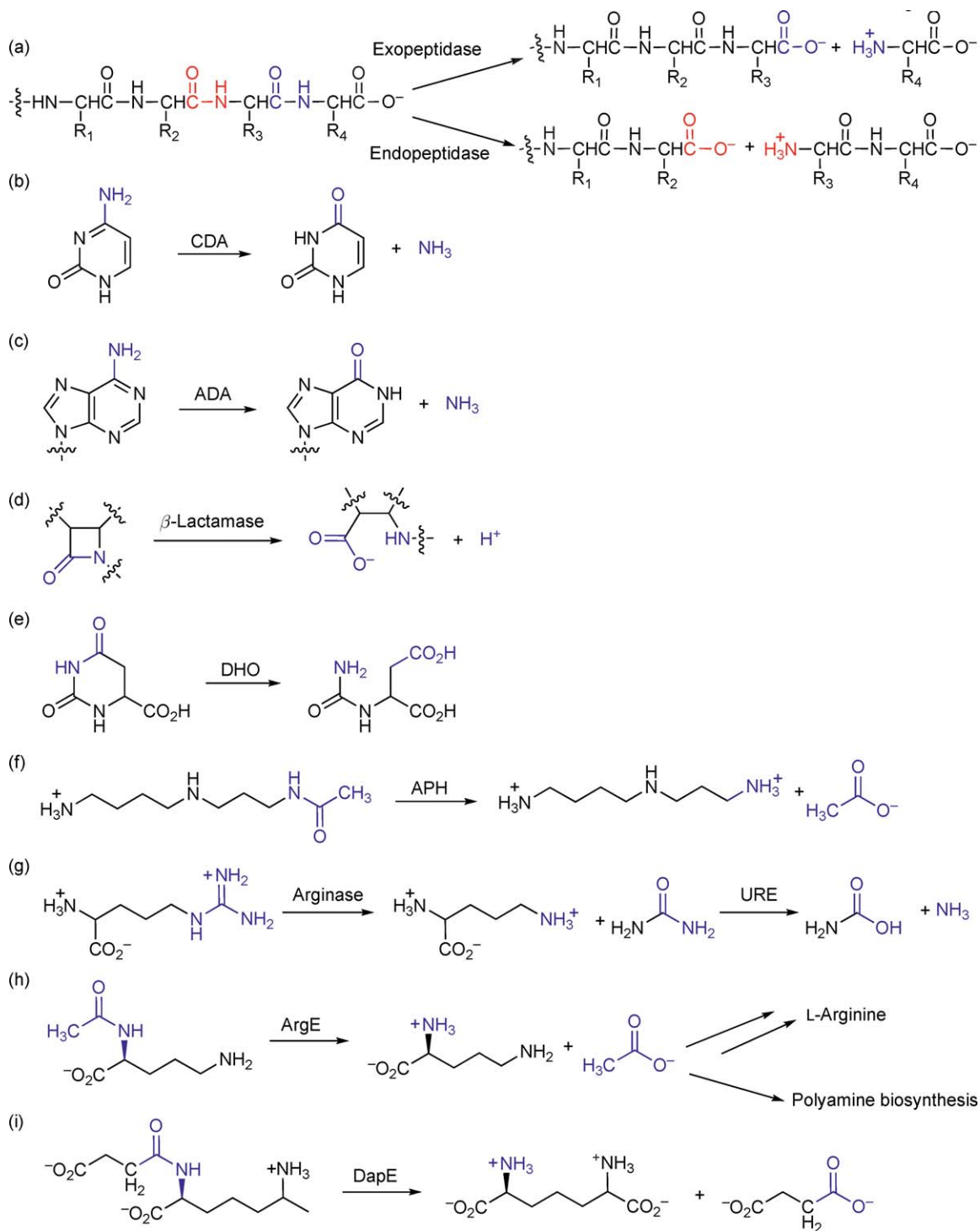
Figure 6 Hydrolytic reactions catalyzed by various metallohydrolases with the leaving group shown in blue and red.

substrate, while the resulting oxyanion tetrahedral intermediate and flanking transition states are stabilized by the catalytic zinc ion and positively charged side chain of Arg127. Breakdown of the tetrahedral intermediate is facilitated by protonation of the amine leaving group by the now protonated Glu270 functioning as a GAC to regenerate the active enzyme following product release. Other metalloproteases (ie., thermolysin, MMPs) are
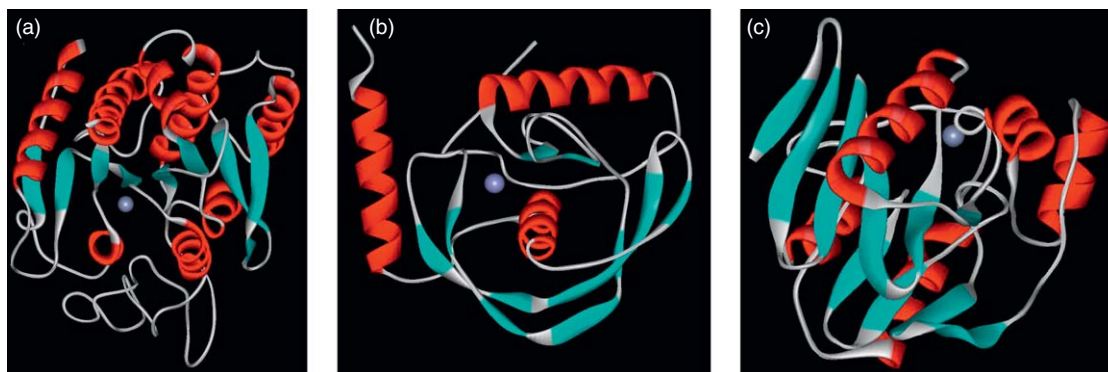
Figure 7 Representative structures of folds of metallohydrolases involved in peptide and β-lactam metabolism. For clarity, only a monomer is shown for each enzyme with the metal ions shown as gray spheres. (a) CPA (PDB 5CPA), (b) PDF (PDB 1BSZ) and (c) β-lactamase (PDB 1BMC).
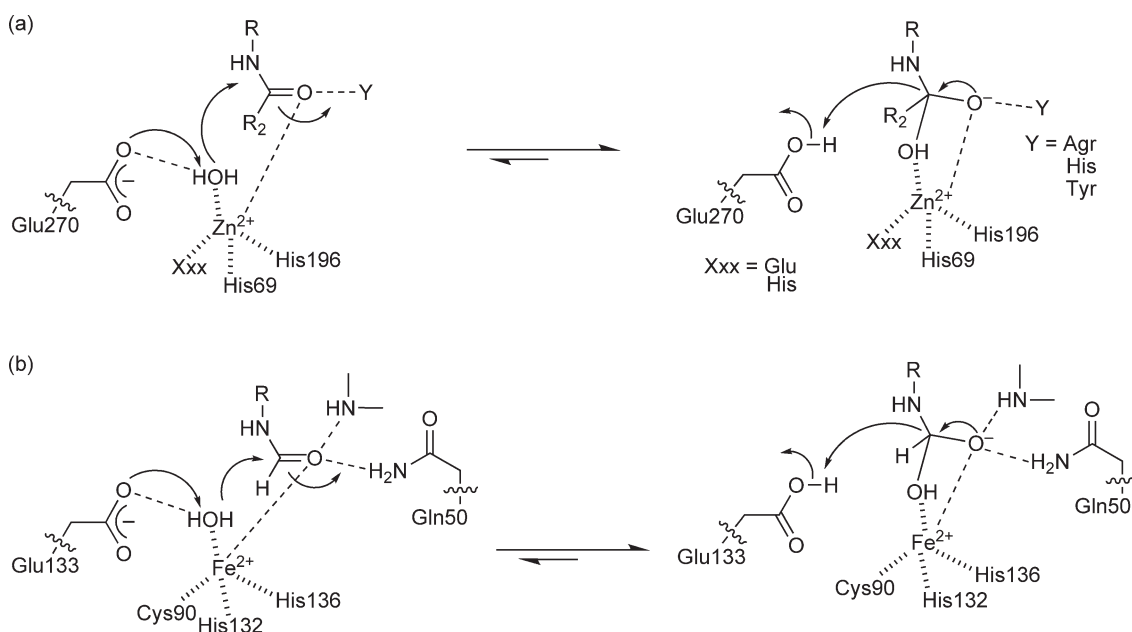
Figure 8 Proposed catalytic mechanisms for enzymes involved in protein/peptide metabolism using a single bifunctional general acid–base catalyst. (a) CPA and (b) PDF.

proposed to follow a similar CPA-like mechanism using a single bifunctional GABC (Glu). The major deviation for the mechanisms of these enzymes from the CPA mechanism is the identity of the side chain (i.e., His versus Arg) that assists the zinc ion as the electrostatic catalyst in the reaction.

8.15.3.2.2   Peptide deformylase

PDF catalyzes the removal of the N-terminal formyl group from nascent polypeptides in bacteria to generate the mature free N-terminal polypeptide and formate (**Figure 3(d)**).[57] To date, several bacterial PDFs have been identified. Although eukaryotes do not use formyl-methionine in protein translation, a human homologue (HsPDF) has been identified. This HsPDF contains a mutation that renders this protein much less active than its bacterial counterpart, suggesting that the human enzyme is an 'evolutionary remnant' with no functional role.[58] Consequently, PDF is an excellent proposed drug target for the development of antibiotics. Mechanistically, EcPDF is the most widely studied of these enzymes and will be the focus of this review.

The crystal structure of PDF from *E. coli* (EcPDF) has been solved and indicates an overall fold distinct from those of other known metallohydrolases (**Figure 7(b)**).[59] Structurally, there is a center α-helix that contains the active site and a metal ion (bound by a conserved HEXXH motif), which is surrounded by three β-sheets and two α-helices. PDF was originally proposed to be a zinc-dependent enzyme; however, later studies have shown EcPDF, as well as PDF from other organisms, is an Fe^{2+}-dependent metallohydrolase.[60] Kinetic experiments indicate that EcPDF activity follows: Fe^{2+}, Co^{2+}, $Ni^{2+} > Zn^{2+}$, whereas metal ion affinity (K_D^{metal}) for EcPDF is of the order $Zn^{2+} < Fe^{2+}$, Ni^{2+}. PDF copurifies with Fe^{2+} under anaerobic conditions; the activity of Fe^{2+}-PDF is not stable under aerobic conditions and PDF is inactive with Fe^{3+}. Consequently, misidentification of PDF as a zinc-dependent enzyme is attributed to the aerobic instability of Fe^{2+}, with the oxidation of Fe^{2+} to Fe^{3+} leading to cofactor dissociation and replacement by Zn^{2+}. Interestingly, PDF from *Borrelia burgdorferi*, an Fe^{2+}-limited organism, has been shown to be a zinc-dependent enzyme.[32] This reiterates the fact that enzymes from different organisms can have different preferences for metal ion cofactors, and may allow organisms to adapt to their specific environments.

Crystal structures of PDF with various metal ions (Zn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}) bound to the active site have been solved, and indicate that the catalytic metal ion coordinates with the side chains of His132, His136, Cys90, and a water molecule in tetrahedral geometry.[59,61] Although the overall fold of PDF (**Figure 7(b)**) differs from the metalloproteases, the reaction that is catalyzed and the active-site architecture of PDF are strikingly similar to these enzymes suggesting that they may have similar mechanisms. The mechanism of PDF has been probed kinetically.[58,60,62,63] The PDF-catalyzed reaction has a bell-shaped dependence on pH, indicating that at least two ionizations are important for maximal activity. Mutagenesis, kinetic, and spectroscopy experiments indicate that the pK_{a1} of 5.2 (favorable deprotonation) represents ionization of Glu133, which is essential for catalysis ($\sim10^7$-fold decrease in activity for E133A mutant). Studies also suggest that the pK_{a2} of ~11.6 (favorable protonation) reflects ionization of the metal–water in the WT enzyme; the pK_a of metal–water in the E133A mutant (Co(II)-PDF) is 6.5. These pK_a assignments (pK_{a1} = GBC, pK_{a2} = metal–water) are similar to those proposed for LpxC described above. Crystal structures of PDF•formate complexes, which should mimic the oxyanion intermediate, highlight important coordination differences for PDF substituted with different metal ions.[64] The formate molecule in PDF substituted with Fe^{2+} and Co^{2+} coordinates with the active-site metal ion through a bidentate interaction, whereas the formate in Zn^{2+}-PDF coordinates with the metal ion through a monodentate interaction. These findings suggest that there is an increase in coordination number and an alteration in the geometry of the catalytic metal ion during the course of the reaction. The suggestion that Fe and Co are better suited to undergo these changes may account for the observed differences in catalytic activity observed for PDF substituted with these metal ions. Together these data have led to the proposed mechanism shown in **Figure 8(b)**.[63] Following substrate binding and coordination with the Fe^{2+} ion, Glu133 serves as a GBC to activate the metal–water for attack on the substrate carbonyl group. The resulting oxyanion intermediate is stabilized by the iron ion through a bidentate interaction, as well as through interactions with the residues Leu91 (backbone) and Gln50 (side chain). Proton transfer from Glu133 to the nitrogen atom facilitates breakdown of the tetrahedral intermediate and subsequent product formation.

8.15.3.2.3   D-aminoacylase

Aminoacylases catalyze the hydrolysis of an *N*-acetyl group from *N*-α-acyl amino acids (**Figure 3(f)**). Although the biological role(s) of aminoacylases has been elucidated, these enzymes are believed to function in xenobiotic metabolism.[65] There is also commercial interest in developing aminoacylases for the preparation of chiral amino acids. D- and L-aminoacylases do not share sequence or structural homology, and therefore the reactions catalyzed by these enzymes likely occur through different mechanisms. Crystal structures of mononuclear (D-aminoacylase)[66] and binuclear (aminoacylase-1, ACY1)[67] aminoacylases indicate that these enzymes possess different overall folds and belong to different superfamilies, the amidohydrolase and zinc α,β-hydrolase superfamilies, respectively. This section will focus on the mononuclear enzyme D-aminoacylase.

The crystal structure of D-aminoacylase from *Alcaligenes faecalis* DA1 has been solved, indicating that it is structurally homologous to the amidohydrolase superfamily of enzymes (e.g., NagA, URE).[66] Similar to TmNagA, D-aminoacylase contains a binuclear metal center although only a single metal ion is required for catalysis.[66,68] Since D-aminoacylase copurifies with Zn^{2+}, it is presumed that this is the native cofactor for the enzyme. The ligands for the catalytic zinc ion are Cys96, His220, His250, and acetate (presumably a water

molecule in the resting enzyme), as determined using both the crystal structure and mutagenesis studies. Metal titration experiments confirm that D-aminoacylase is a mononuclear metallohydrolase.[66] Consequently, the second zinc ion observed in the crystal structure presumably binds at an inhibitory site, as observed for other metallohydrolases.[68] The catalytic importance of conserved active-site side chains have been examined kinetically.[69,70] Mutation of the conserved Asp366 to Ala abolishes enzyme activity, confirming that this side chain is essential for catalysis. Furthermore, since this side chain is located adjacent to the zinc–water in the crystal structure it is proposed to function as a GBC to activate the metal–water in the reaction mechanism. Mutagenesis experiments also indicate that the conserved side chains of His67 and His69, which are ligands for the inhibitory zinc ion, are important for catalysis. Consequently, D-aminoacylase is proposed to follow a CPA-like mechanism (**Figure 8(a)**).[66] However, it is also possible that these essential His side chains could function to stabilize the oxyanion intermediate (similar to NagA, **Figure 5(b)**) or as GABC to assist Asp266 (similar to LpxC, **Figure 5(a)**).[19] The fact that D-aminoacylase belongs to the amidohydrolase superfamily may suggest that this enzyme is more likely to function using a mechanism similar to that shown in **Figure 5(b)**.

8.15.3.3 Amine Metabolism

Metal-dependent hydrolases are also involved in the metabolism of aliphatic amines, including lysine and various polyamine molecules (i.e., putrescine, spermine, and spermidine). Sequence homology suggests that the enzymes that catalyze the deacetylation of these molecules belong to a larger superfamily of enzymes that includes the HDACs, acetylpolyamine amidohydrolases (APHs) and acetoin utilization proteins.[71] Consequently, our review of the HDACs and APHs has been grouped together. The substrates for these enzymes play important biological roles in gene expression, as well as cell growth, proliferation, and signaling and are therefore targets for drug development.

8.15.3.3.1 Histone deacetylase

HDACs catalyze the removal of an acetyl group from an ϵ-acetylated lysine side chain (**Figure 3(e)**). In eukaryotic cells, the lysine side chains on histone molecules are substrates for the HDACs.[72] Since modulation of histone acetylation plays an important role in regulating gene transcription, HDACs are targets for the development of anticancer agents. Although the role(s) of bacterial HDACs, also known as HDAC-like proteins (HDLPs), remain to be elucidated, these enzymes have also been targeted for antibiotic development.[73] Human HDACs are divided into four different classes of enzymes.[74] Class I (HDAC 1, 2, 3, and 8), class II (HDAC 4, 5, 6, 7, 9, and 10), and class IV (HDAC 11) are metal-dependent enzymes, and will be the focus of this section. Class III HDACs are NAD^+-dependent enzymes, whose mechanisms are discussed elsewhere. Bacterial HDACs from all four classes have been identified.[73,74]

Crystal structures of class I HDACs (human HDAC8[75] and Aa HDLP[76]) and bacterial class IIb HDACs (*Bordetella/Alcaligenes* HDLP[77]) reveal an overall α,β-fold with an eight-stranded parallel β-sheet sandwiched between 13-α-helices (**Figure 4(d)**) that is quite similar to the Mn^{2+}-dependent enzyme arginase (discussed below). The active sites of these HDACs contain a bound zinc ion that coordinates with the side chains of Asp178, His180, Asp267, and a water molecule. The ligand side chains that coordinate the catalytic metal ion are conserved across the metal-dependent HDACs, and describe a unique His-Asp_2-H_2O zinc-binding site. Recently, a crystal structure of the first eukaryotic class IIa HDAC (catalytic domain of HDAC7)[78] was determined revealing a similar overall fold to the class I and IIb HDACs and a second conserved zinc-binding site (Cys533, Cys535, His541, Cys618; tetrahedral geometry) that is unique to the class IIa enzymes. This second binding site is located at the entrance of the active site near the hydrophobic capping group of the bound inhibitor, suggesting that this site may be important for substrate recognition, protein–protein interactions, and regulation of activity.[78] Treatment of HDACs with metal chelators (EDTA) inhibits enzyme activity, demonstrating that the bound metal ion is essential for catalytic activity.[35,76,79] The finding that HDACs copurify with Zn^{2+} has led to their characterization as zinc-dependent enzymes. Metal titration experiments indicate that HDAC is maximally active with a single metal ion and inhibited in the presence of excess Zn^{2+}, confirming that HDAC is a mononuclear enzyme.[35] Recently, HDAC8 has been shown to copurify with Fe^{2+} under anaerobic conditions suggesting that HDAC8, and possibly other HDACs, may be an Fe^{2+}-dependent enzyme.[35]

Crystal structures indicate that the HDAC active site contains two His/Asp charge relays (His142/Asp176 and His143/Asp183) and a conserved Tyr306 in close proximity to the catalytic zinc ion. Both His/Asp charge relays are conserved in class I HDACs, whereas one charge relay is lost with a conserved Asp to Asn substitution (His143/Asn183) in the class II and IV HDACs. The activity (k_{cat}, k_{cat}/K_M) of HDAC has a bell-shaped dependence on pH, similar to the metallohydrolases previously described.[80] The pK_a values (k_{cat}/K_M) for these ionizations are ∼7 and ∼9, with some variation observed for the different isozymes. Although there are several groups in the active site that could be responsible for these ionizations, including the zinc–water, His142, His143, and Tyr306, the sources of these ionizations have yet to be assigned. Mutagenesis experiments indicate that the side chains of His142, Asp176, and Tyr306 are all important for catalysis to different degrees.[76,79,81,82] These findings are consistent with the proposal that hydrolysis by HDAC follows the GABC-pair mechanism shown in **Figure 9(a)**.[76] The substrate coordination with the catalytic metal ion leads to polarization of the carbonyl group on the substrate. Next, His142 acts as a GBC to activate the metal–water for nucleophilic attack of the substrate, and the resulting intermediate is stabilized by the catalytic zinc ion and the side chain of Tyr306. Breakdown of the tetrahedral intermediate is facilitated by protonation of the amine by His143 acting as a GAC and is followed by product release. However, a more recent theoretical investigation of the HDLP mechanism using DFT calculations suggests that His142 and His143 are both singly protonated species in the enzyme–substrate complex, and that the reaction proceeds through an alternative mechanism using a single bifunctional GABC (**Figure 9(b)**).[83] In this mechanism, the side chain of His143 serves as a GBC to activate the metal–water for attack on the carbonyl group on the substrate (rate-determining step), whereas His142 functions to stabilize the water through a hydrogen bonding interaction. The resulting tetrahedral intermediate is stabilized by the side chain of Tyr306. Finally, the now protonated His143 serves as a GAC to facilitate breakdown of the tetrahedral intermediate. Results from recent kinetics experiments support this latter mechanism (S. L. Gantt and C. A. Fierke, unpublished results).

8.15.3.3.2   *Acetylpolyamine amidohydrolases*

Polyamines are a group of aliphatic amines that are derived from ornithine, including spermidine, spermine, cadaverine, and putrescine, which are important for normal cell growth.[84,85] The APHs are a group of enzymes that catalyze the removal of an acetyl group from the corresponding acetylated-polyamine analogues. An example of the deacetylation of *N*-acetyl-spermidine is shown in **Figure 6(f)**. Although there are currently no APH structures available, sequence homology suggests that these enzymes belong to a superfamily of enzymes that also includes the HDACs.[71] Consequently, information regarding HDACs has been used to offer possible insights into the catalytic mechanisms of these enzymes.

Metal titration experiments on APH from *Mycoplana ramose* indicate that APH is maximally active with a single metal ion, and is inhibited by excess zinc, demonstrating that APH is a mononuclear enzyme.[86] Although this enzyme can also use Co^{2+} as a cofactor,[86] the finding that APH copurifies with Zn^{2+} suggests that Zn^{2+} is the native cofactor for this enzyme. The mechanism for APH hydrolysis has not been probed using mutagenesis and kinetic studies; consequently, sequencing data has been used to suggest possible mechanisms. Sequence alignment of APH with CPA (no overall statistically significant sequence similarity) suggests that APH may contain a conserved HEXXH + E motif. This led to the proposal that the catalytic zinc ion coordinates with a His_2-Glu-H_2O motif, and that the reaction proceeds using CPA-like mechanism (**Figure 8(a)**).[86] However, sequence alignment of APH with HDACs, which belong to the same enzyme superfamily, reveals that the metal ligands (Asp178, His180, and Asp267) and catalytic side chains (His142, His143, Asp176, and Tyr306) of HDAC are conserved in APH. More specifically, the sequence of APH is in better agreement with the class II HDACs (one conserved His/Asp dyad). This led to the suggestion that the catalytic zinc ion coordinates with a His-Asp_2-H_2O motif and that the reaction proceeds through a GABC pair mechanism similar to the HDACs (**Figure 9(a)**).[19] This example illustrates the limitations of using sequence data alone for the identification of metal binding sites, catalytically important side chains, and reaction mechanisms.

8.15.3.4   Nucleotide Metabolism

Deamination of purine and pyrimidine bases (**Figures 6(b)** and **6(c)**) is an important reaction in nucleotide salvage pathways and RNA editing.[87,88] In nucleosides, two of the most characterized reactions involve the conversion of adenosine and cytidine to form inosine and uridine (with the elimination of one molecule of
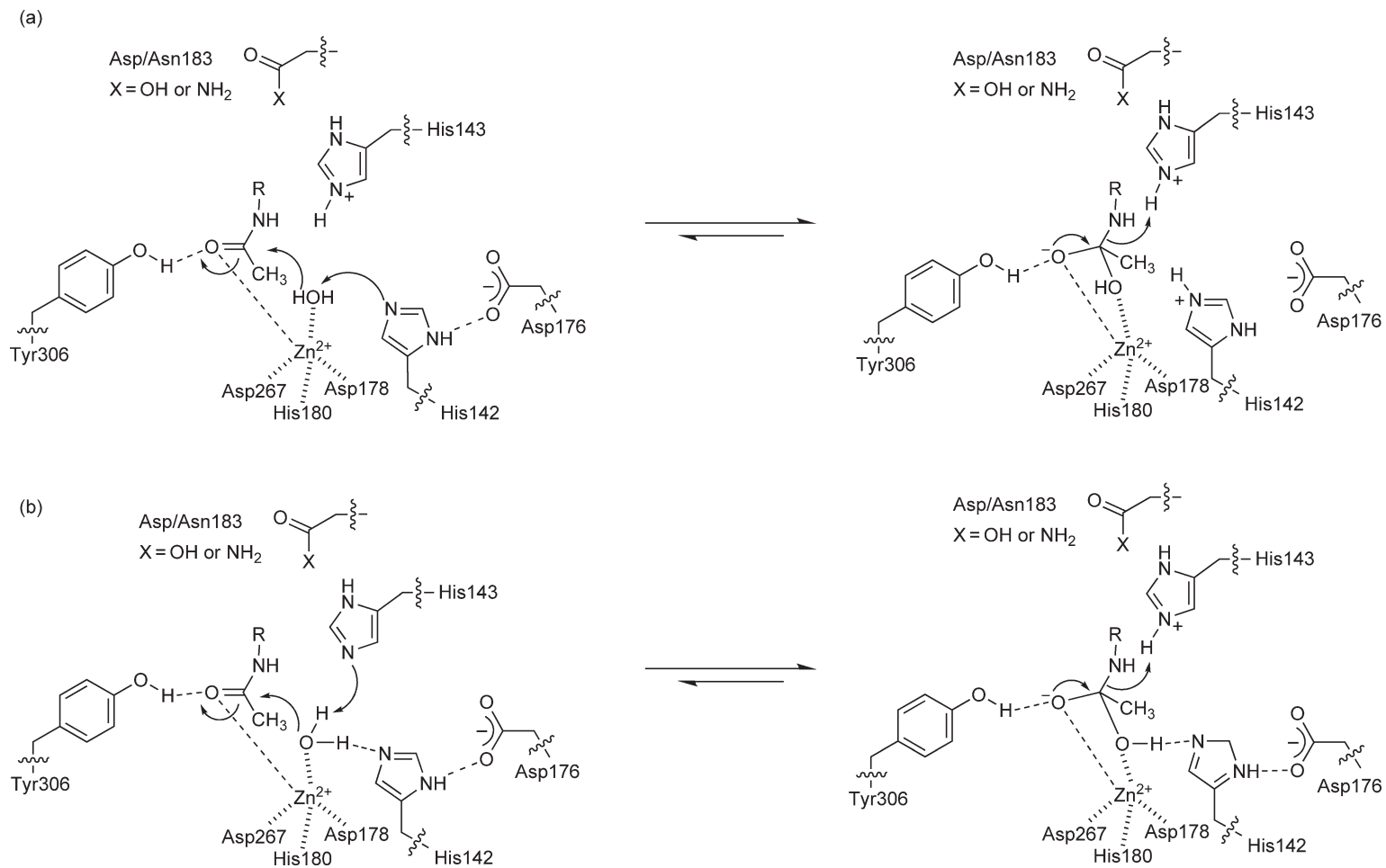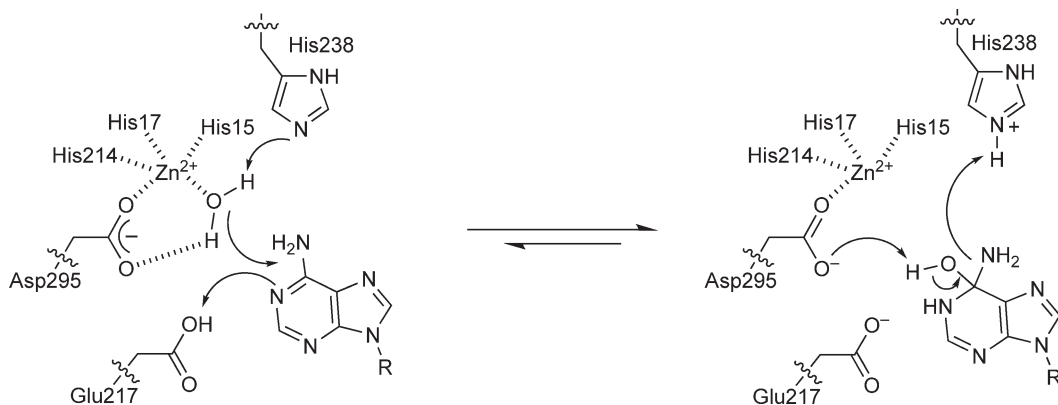
Figure 9 Proposed catalytic mechanisms for histone deacetylases using a (a) GABC pair mechanism or (b) single bifunctional GABC mechanism.

ammonia), respectively. Deamination of these bases can occur either in the free nucleobase, nucleoside, or nucleotide molecules, in reactions that are catalyzed by different enzymes. Since the substrates for these enzymes are aromatic molecules, these hydrolytic reactions are nucleophilic aromatic substitution reactions, a distinguishing feature from the reactions that have been already discussed. Two of these enzymes, ADA and CDA belong to the amidohydrolase superfamily, and are the focus of this review.

8.15.3.4.1 Adenosine deaminase

ADA catalyzes the conversion of adenosine to form inosine and ammonia, whereas CDA catalyzes the hydrolysis of cytosine to form uracil and ammonia (**Figures 6(b)** and **6(c)**). Crystal structures of ADA[89] and CDA[90] illustrate the overall folds of these enzymes have similar $(\beta/\alpha)_8$-barrels, indicating that both enzymes belong to the amidohydrolase superfamily.[26] The active sites of these enzymes contain a bound single metal ion located at the M_α site with a His_3-Asp-H_2O coordination. The ADA active site contains a bound Zn^{2+} that coordinates with the side chains of His15, His 17, His214, Asp295, and a water molecule, whereas the CDA active site contains a bound Fe^{2+} that coordinates with His61, His63, His214, Asp313, and a water molecule. In both enzymes, His246 is appropriately positioned for an octahedral metal geometry, but it is too far (3.7 Å) to suggest a direct coordination. Treatment with metal-chelating agents inhibits deaminase activity, indicating that the bound metal ion is essential for catalysis in both enzymes.[91,92] Metal titration experiments indicate that both enzymes are mononuclear enzymes that use a single metal ion for catalysis, and that the natural cofactors for ADA and CDA are Zn^{2+} and Fe^{2+}, respectively.[91,92] This difference in ADA and CDA metal ligation is another example of where a common structural scaffold uses alternative cofactors to catalyze similar reactions.

Crystal structures of ADA complexed with a transition state analogue 6-hydroxyl-1,6-dihydropurine riboside (HDPR)[89,93] or 1-deazaadenosine[93,94] suggested several roles for side chains in the reaction mechanism that have subsequently been probed using mutagenesis and kinetic studies.[95–97] Although the ADA reaction has a bell-shaped dependence on pH, the magnitude of change over the range of the pH profile is more modest than that observed for the other enzymes. Mutagenesis experiments indicate that the side chains of His238, Asp295, Asp296, and Glu217 (~3200-fold decrease in activity) are important for catalysis. Structural data are more consistent with Asp296 functioning to stabilize substrate binding, rather than serving a direct role in the chemical mechanism. The side chains of His238 and Asp295 share hydrogen bonds with the hydroxyl group in the transition state analogue, suggesting that they are important for formation, stabilization, and/or orientation of the metal-hydroxide nucleophile. The findings that Glu217 is located adjacent to N-1 in the crystal structure with the transition state analogue and that the mutation to Gln decreases activity (4800-fold, comparable to loss of side chain) suggest that this side chain catalyzes a proton transfer reaction in the mechanism. In the proposed mechanism (**Figure 10**),[26] His238 is proposed to function as a GBC (or as electrostatic catalyst) to activate the metal–water for attack of the aromatic ring, which is also facilitated by protonation of the N-1 ring nitrogen by

Figure 10 Proposed catalytic mechanism for adenosine deaminase.

Glu217 (conserved HxxE motif). In contrast to the metal-dependent hydrolases that have already been discussed, the resulting carbinol tetrahedral intermediate for this reaction is not an oxyanion and therefore may not require the stabilization observed for these other enzymes. Deprotonation of the hydroxyl group on the carbinol intermediate by Asp295 and protonation of the amine group by His238 facilitates breakdown of this tetrahedral intermediate. Proton transfer from Glu217 to Asp295 regenerates the active enzyme following product release.

8.15.3.5 β-Lactam Metabolism – *Bacillus cereus* β-Lactamase

The β-lactam antibiotics (e.g., pencillins, cephalosporins, and carbapenems) are widely used for the treatment of infectious diseases. Resistance to these agents is common and can be attributed to hydrolysis of the amide bond in the β-lactam moiety in a reaction catalyzed by β-lactamase (βl) enzymes (**Figure 6(d)**).[98] Consequently, an understanding of these enzymes is warranted as a means to circumvent antibiotic resistance. βls are broadly categorized as either serine βls (classes A, C, and D) or metallo-βl (class B).[99,100] Metallo-β-lactamases fall into three different subclasses B1, B2, and B3, and may require one or two metal ions for optimal activity.[99,100] Structures indicate that the active sites of metallo-βl contain two zinc-binding sites, although the affinities of these sites and the metal ion requirements for catalysis vary. It should be mentioned that the exact mechanisms for these enzymes remain a subject of controversy. Readers are referred to a recent review of the metallo-βls for a more detailed description of the properties and characteristics of specific enzymes.[100] This controversy may in part be explained by members of the metallo-β-lactamases superfamily of proteins having different metal ion requirements and reaction mechanisms. A brief overview of studies on the mononuclear metallo-βls, in particular, the enzyme from Bc, is presented here. For additional information on β-lactamases, see also Chapter 8.13.

Crystal structures of representatives from all subclasses of metallo-βls have been solved, illuminating a similar overall $\alpha\beta\beta\alpha$-fold (**Figure 7(c)**) and signature HXHXDH motif that is important for metal coordination and catalytic activity.[100] The active site of βl from *B. cereus*[101] (Bcβl) contains two zinc-binding sites, although it is suggested that only one zinc ion is necessary for maximal activity, and that the affinity of Zn for the catalytic site is ~1000-fold tighter than the second binding site.[102,103] The Zn_1 site (catalytic) coordinates with the side chains of His116, His118, His196, and a water molecule with tetrahedral geometry, whereas Zn_2 site (inhibitory) coordinates with Asp120, Cys221, His263, and two water molecules with trigonal bipyramidal geometry. Bcβl activity is inhibited by metal-chelating agents, which is restored following the addition of Zn^{2+}, consistent with its characterization as a zinc-dependent metallohydrolase.[102,103] Mutagenesis studies indicate that the conserved Asp120 is important for catalytic activity.[104] The Zn(II)-Bcβl catalyzed reaction exhibits a bell-shaped dependence on pH that has been described by pK_{a1} value of 5.6 (represents two ionizations) and pK_{a2} of 9.5.[105,106] There is no significant solvent isotope effect observed for the reaction ($[k_{cat}]^H/[k_{cat}]^D = 1.5$, $[k_{cat}/K_M]^H/[k_{cat}/K_M]^D = 1.3$), suggesting that GBC does not occur in the steps described by these parameters. The value of pK_{a1} is highly dependent on the identity of the metal bound to the active site and varies from 5.6 (Zn^{2+}- Bcβl) to 8.7 (Cd^{2+}-Bcβl), suggesting that pK_{a1} reflects ionization of a group near the metal ion, such as the metal–water. Since Bcβl activity increases with increasing pH for this ionization, a metal-hydroxide is proposed to be the nucleophile for the reaction. Of the metal ions examined (Zn^{2+}, Co^{2+}, Mn^{2+}, Cd^{2+}), only Zn^{2+} can stabilize a metal-hydroxide at neutral pH, which may explain why this is the preferred cofactor for this βl. The value of pK_{a2} does not vary significantly with changes in metal ion or the substrate, suggesting that this ionization reflects a group on the enzyme. It has been suggested that this ionization may be that of Lys224, which is positioned to interact with a carboxylate group. On the basis of these studies, a mechanism for the hydrolysis reaction catalyzed by Bcβl and other mononuclear βls is proposed and shown in **Figure 11(a)**.[105] The catalytic zinc ion functions as a Lewis acid to lower the pK_a of the metal–water. This metal-hydroxide attacks the carbonyl group on the substrate to form a tetrahedral intermediate. The side chain of Asp120 functions as a GBC to deprotonate the OH of this intermediate to form a second dianionic tetrahedral intermediate that is stabilized by binding to the catalytic zinc ion. Finally, Asp120 serves as a GAC to protonate the ring nitrogen and facilitate opening of the lactam ring to generate the product.
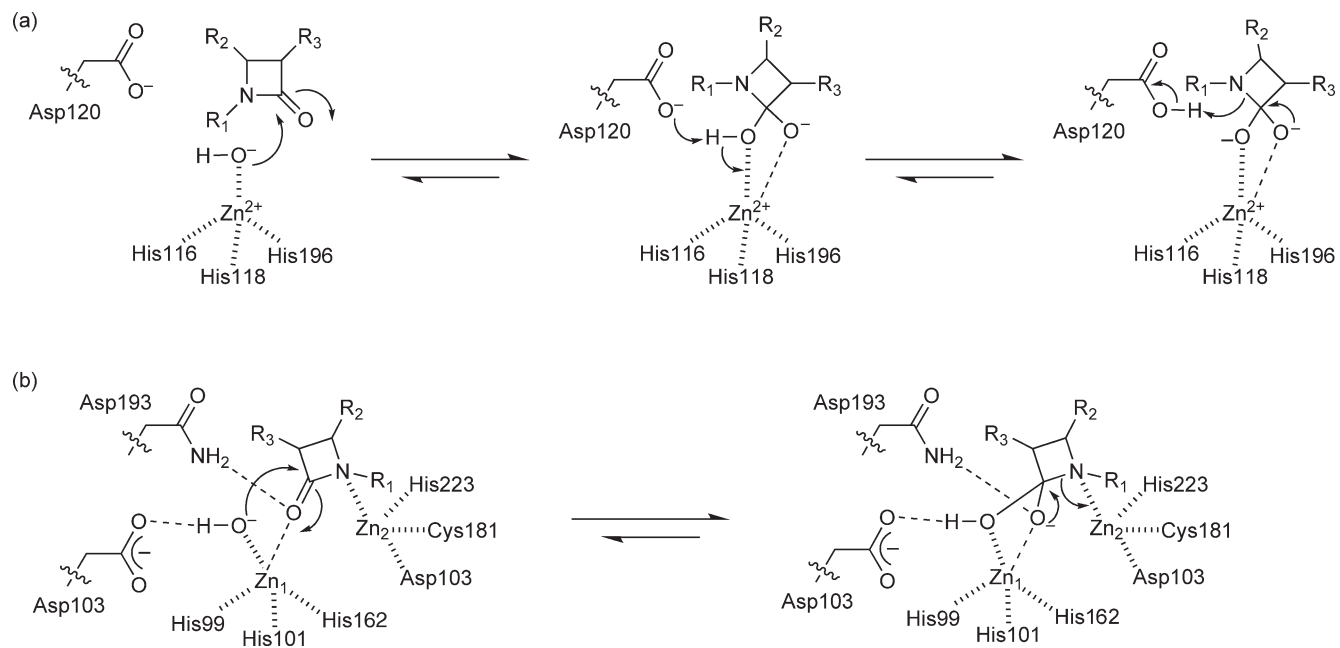
Figure 11 Proposed catalytic mechanisms for (a) mononuclear and (b) binuclear metallo-β-lactamases.

8.15.4 Binuclear Metallohydrolases

Enzymes that contain binuclear metal centers are also well suited to catalyze hydrolysis reactions, including a number of the reactions described above for the mononuclear metallohydrolases. Additionally, several of the examples that are discussed here belong to the enzyme superfamilies described above, specifically the amidohydrolase, zinc α,β-hydrolase, and metallo-βl superfamilies. The substrates for the binuclear metallohydrolases are also biologically diverse, including proteins, peptides, nucleotides, polyamines, and xenobiotics. The binuclear nature of these metal centers produces an active site with altered properties compared to the mononuclear counterparts, and therefore catalysis by these enzymes occurs with alternative reaction mechanisms. Readers are referred to the preceding sections for background information pertaining to enzymes that have already been discussed.

8.15.4.1 Protein and Peptide Metabolism

8.15.4.1.1 Aminopeptidases

Numerous binuclear metalloenzymes are involved in digestion and metabolism of proteins and peptides (**Figure 6(a)**).[44,45,107] The aminopeptidase from *Aeromonas proteolytica* (AAP) is a prototypical example of a binuclear metallopeptidase.[108] The crystal structure of AAP indicates that it is a member of the zinc α,β-hydrolases superfamily with an overall fold described as an eight stranded β-sheet surrounded by eight α-helices.[109] The active site of this enzyme contains a binuclear metal center that has been well characterized using spectroscopic techniques. The two zinc ions of the metal center are separated by 3.5 Å, with each Zn^{2+} best described as a symmetrical five coordinate species with distorted tetrahedral geometry. The ligands for Zn_1 are Asp117, Glu152 (bidentate), His256, and a water molecule, whereas the ligands for Zn_2 are Asp117, Asp179 (bidentate), His97, and a water molecule. In this arrangement, the bound water molecule serves as one bridging ligand for the two zinc ions, and also shares a hydrogen bond with the side chain of Glu151, whereas Asp117 bridges the two zinc ions through a bidentate interaction. Treatment with metal-chelating agents eliminates aminopeptidase activity, indicating that at least one of the bound metal ions is essential for catalytic activity.[110] Metal titration experiments on apo-AAP with Zn^{2+} indicate that two metal ions are required for maximal activity, although the mono-Zn^{2+}-AAP ($K_D^{Zn} \sim 150\,pmol\,l^{-1}$) retains ~80% maximal activity.[110–113] These studies also indicate that the two Zn^{2+} ions bind to the enzyme in a sequential fashion. Mono-substituted AAP containing Cu^{2+}, Co^{2+}, or Ni^{2+} resulted in an enzyme that is more active than the corresponding mono-Zn^{2+}-AAP (7- to 25-fold), whereas the subsequent addition of Zn^{2+} to the Ni^{2+}-AAP or Cu^{2+}-AAP activated the enzyme to an even greater extent ($\leq$100-fold) compared to the Zn_2^{2+}-AAP. However, since the isolated enzyme copurifies with a dizinc center, these are presumed to be the natural cofactors for this enzyme.

The activity of AAP is dependent on a single ionization with a pK_a value of 4.8.[114] Although this was originally thought to reflect ionization of the metal-hydroxide, it is now proposed to reflect ionization of a carboxylate side chain (e.g., metal ligand, Glu151).[108,115] The proposed mechanism for this enzyme is presented in **Figure 12(a)**.[115] First, the carbonyl group on the substrate coordinates with Zn_1 of the binuclear metal center in a step that is coupled with breaking of the Zn_2-water bond, leading to polarization of the substrate and the formation of a terminal Zn_1-water. The side chain of Glu151, which shares a hydrogen bond to the metal–water in the AAP crystal structure, is positioned to act as a GBC to activate the metal–water for attack of the carbonyl group on the substrate. The resulting oxyanion intermediate and flanking transition states are then stabilized by the Zn_1 and Zn_2 ions. The finding that the mono-substituted Zn^{2+}-AAP retains 80% of the activity suggests that in the absence of Zn_2 other active-site side chain(s) can fulfill the function of this metal ion. Protonation of the tetrahedral intermediate by Glu151 facilitates breakdown of the tetrahedral intermediate and formation of the products.

Bovine leucine aminopeptidase (BLAP) is another classical binuclear metalloprotease that belongs to the zinc α,β-hydrolase superfamily of enzymes and is topologically similar to AAP.[49] However, the overall location of the binuclear metal center and the coordination of the individual zinc ions differ from those observed for AAP. Crystal structures of BLAP indicate that the two zinc ions in the active-site binuclear metal center are closer together (2.3–3.0 Å separation) than in the AAP.[116–119] The two zinc sites of the BLAP binuclear metal center are each a five-coordinated unsymmetrical species with what is best described as an octahedral
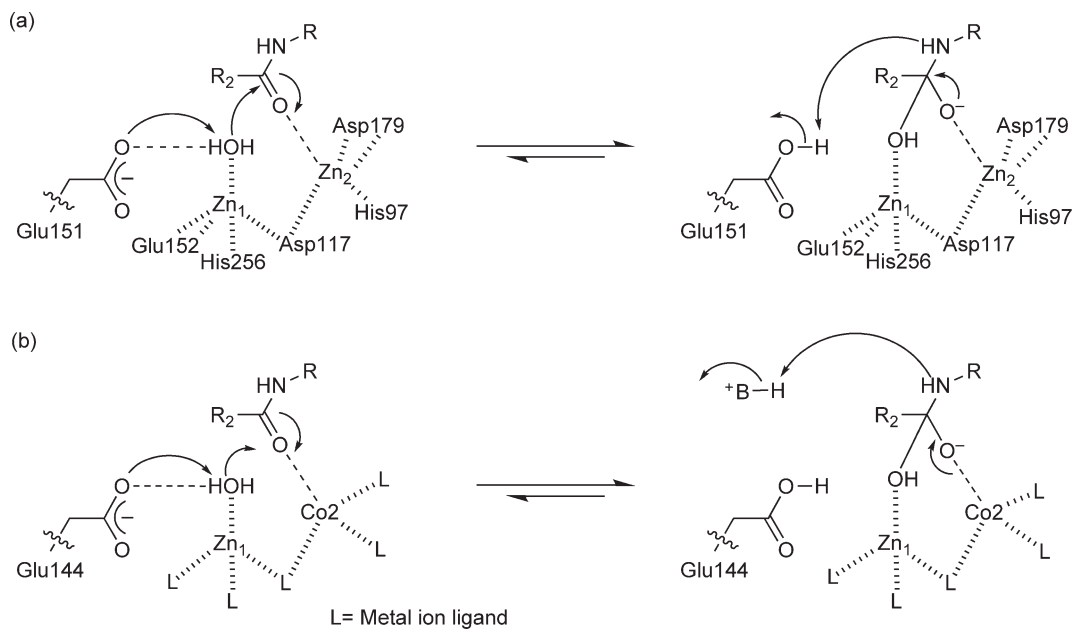
Figure 12 Proposed catalytic mechanisms for the binuclear metallohydrolases using a (a) single bifunctional GABC (AAP) and (b) GABC pair (ArgE).

geometry with one missing vertex. The Zn_1 coordinates with Asp255, Asp332 (bidentate), Glu334, and a water molecule, whereas Zn_2 coordinates with Lys250, Asp255, Asp273, Glu334, and a water molecule. In this arrangement, the two zinc ions of this binuclear metal center are bridged by Glu334 (bidentate), Asp255 (monodentate), and a water molecule. The apparent affinities of BLAP for zinc ions are weaker than observed for other zinc enzymes ($K_D{}^{Zn_1} \gg K_D{}^{Zn_2} = 0.1\text{--}1\,\text{nmol}\,l^{-1}$, compared to $K_D{}^{Zn} \sim 4\,\text{pmol}\,l^{-1}$ for carbonic anhydrase), which may suggest that zinc is not the physiological cofactor for both sites.[117,120–122] Structural differences between the active sites of AAP and BLAP necessitate some dissimilarities in the proposed reaction mechanisms (AAP shown in **Figure 12(a)**).[119,123] In the first step of the BLAP mechanism, the carbonyl group of the substrate coordinates with Zn_1 and the N-terminal amine coordinates with Zn_2. Next the zinc-hydroxide attacks the carbonyl group of the substrate to generate the tetrahedral intermediate. This is unlike the nucleophilic attack proposed for AAP and CPA that uses a GBC to activate the water, as there is no carboxylate side chain positioned to fulfill this role. The positively charged Zn_1 and Zn_2 then function to stabilize the resulting tetrahedral intermediate. Mutagenesis studies indicate that the conserved Lys262 also participates in this transition state stabilization. Proton transfer to assist breakdown of tetrahedral intermediate may be accomplished by water molecules and/or a bicarbonate molecule that is observed in the crystal structure.[119,123]

The metalloproteases are a family of enzymes with a similar overall fold that catalyze very similar reactions; however, the metal centers used by these enzymes are remarkably diverse. The native cofactors for many aminopeptidases have not been identified; however, increasing evidence suggests that cofactor preferences for the metal centers in these enzymes are complex. In addition to the mononuclear metal centers discussed in previous sections, the metal centers of binuclear enzymes also differ with respect to the location (e.g., AAP and LAP)[49] and identity of metal ion cofactors. In contrast to mononuclear enzymes, which typically prefer Zn^{2+} as the cofactor, binuclear metalloproteases can use Zn^{2+} (AAP, CPG),[124,125] Mn^{2+} (MetAP),[33] Co^{2+} (MetAP),[126,127] or Fe^{2+} (MetAP).[128,129] MetAPs are another example of different cofactor preferences for different isozymes and species.

8.15.4.1.2 Aminoacylases

The aminoacylases catalyze the hydrolysis of *N*-acetyl-amino acids, a reaction that is very similar to the hydrolysis of peptide bonds found in protein/peptides (**Figure 3(f)**). Although the mononuclear aminoacylase

discussed above (D-aminoacylase) belongs to the amidohydrolase superfamily, the binuclear L-aminoacylase (ACY1) discussed here belongs to the zinc α,β-hydrolase superfamily.[49] A crystal structure of human (hACY1) ACY1[67] has been solved showing an eight-stranded β-sheet that is surrounded by eight α-helices, similar to the structures described above for CPA and aminopeptidase. The active site of hACY1 contains a binuclear metal center that is occupied by two zinc ions with a Zn–Zn distance of ~3.4 Å, similar to that observed for AAP. Both zinc ions in this metal center coordinate with five ligands, and are bridged by Asp113 (bidentate) and a bound small molecule that is presumed to be water in the resting enzyme. Zn_1 coordinates with the side chains His373, Asp113, Glu148 (bidentate), and a small molecule; Zn_2 coordinates with His80, Asp113, Glu175 (bidentate), and a small molecule. These structures also show that the conserved Glu147 is hydrogen bonded to the small molecule metal ligand. Mutagenesis studies indicate that, Glu147 is essential for catalysis, and consequently this enzyme is proposed to follow a mechanism similar to AAP (**Figure 12(a)**).[67] A homology model of pACY1 suggests that the active site of this enzyme contains only one bound zinc ion, consistent with biochemical studies indicating that this enzyme requires only one metal ion for catalysis.[130]

8.15.4.2 Amine Metabolism

8.15.4.2.1 Arginase

Arginase catalyzes the hydrolysis of arginine to form ornithine and urea (**Figure 6(g)**), an important step in the urea cycle and regulator of several important pathways, including nitric oxide biosynthesis, proline biosynthesis, and polyamine biosynthesis.[131–133] Thus this enzyme has been well studied using a variety of biochemical and biophysical approaches. The crystal structure of arginase has been solved illuminating an overall α,β-fold similar to the HDACs discussed above, and a binuclear metal center.[134] This binuclear center is occupied by two Mn^{2+} ions that are separated by 3.4–3.6 Å. The Mn_A^{2+} (more buried) ion coordinates with the side chains of His101, Asp124, Asp128, Asp232, and a hydroxide molecule in a square pyramidal geometry. The Mn_B^{2+} (more accessible) ion coordinates with the side chains of His126, Asp124, Asp232, Asp234, and a hydroxide molecule in an octahedral geometry. The bridging ligands for the binuclear metal center are Asp124 (bidentate interaction), Asp232 (monodentate interaction), and water/hydroxide, whereas the side chain of Asp128 shares a hydrogen bond with the bridging hydroxide/water molecule. Both bound Mn^{2+} ions are necessary for maximal catalytic activity; monosubstituted Mn^{2+}-arginase retains half of the maximal activity.[135,136] The crystal structure of monosubstituted Mn^{2+}-arginase indicates that only the Mn_B^{2+} site is occupied, with Asp128 hydrogen bonded to the bound metal–water/hydroxide.[132,133]

Arginase is proposed to catalyze the hydrolysis of arginine through the proposed mechanism shown in **Figure 13**.[134] This mechanism is supported by numerous structural and functional studies that are the subject of recent reviews.[132,133] Substrate binding to the enzyme is facilitated through an ionic interaction between the side chain of Glu277 and the guanidinium group, which also serves to position the guanidinium group in close proximity to the metal center. A direct metal–substrate interaction is not proposed since the K_M value does not change upon mutation of the metal ligands or with the number of bound metal ions.[135,136] The two Mn^{2+} ions serve to lower pK_a of the metal–water to provide metal-hydroxide at neutral pH. Consequently, the single ionization (pK_a of 7.9) that is observed in the pH profile (favorable deprotonation) is proposed to reflect the formation of metal-hydroxide.[137,138] Nucleophilic attack by the metal-hydroxide on the guanidinium group in the substrate yields a neutral tetrahedral intermediate. The absence of a negatively charged oxyanion intermediate (as observed in the mechanisms discussed previously) is because the guanidinium group is positively charged (protonated) at neutral pH. Next, the side chain of Asp128 catalyzes the transfer of a proton from the hydroxide to the amine to yield an activated leaving group and oxyanion intermediate that is stabilized by interactions with the Mn^{2+} ions. The breakdown of this tetrahedral intermediate yields the products urea and ornithine. Product release is facilitated by water binding to the metal center and product release. The side chain of His141 may aid in the transfer of a proton from the metal–water cluster to the bulk solvent or ornithine.[132,133]

8.15.4.2.2 N-acetyl ornithine deacetylase

Arginine is an essential molecule that is synthesized through different pathways and intermediates in eukaryotes and prokaryotes. Two key differences in the prokaryotic biosynthetic pathway (eight steps overall) are: (1) the acetylation
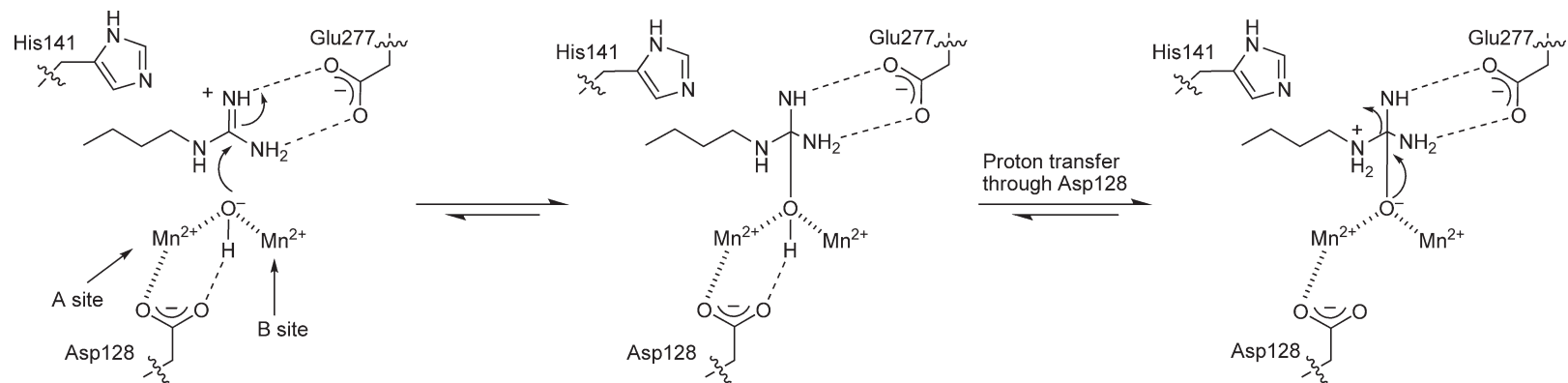
Figure 13 Proposed catalytic mechanism for the dimanganese enzyme arginase.

of glutamate with acetyl-CoA, catalyzed by the enzyme *N*-acetylglutamate synthase and (2) the deacetylation of *N*-acetyl ornithine to form ornithine and acetate (**Figure 6(h)**), catalyzed by the enzyme *N*-acetyl ornithine deacetylase (ArgE).[139] Ornithine is an essential intermediate in bacteria used for the biosynthesis of several critical biological molecules, including arginine, polyamines, proline, siderophores, and antibiotics. Consequently, ArgE is essential for bacterial viability, and therefore is a potential target for the development of new antibiotics.

There are no crystal structures available for ArgE; consequently, the metal center for this protein has not been identified. Sequence homology suggests that ArgE belongs to the zinc α,β-hydrolase superfamily of enzymes (e.g., CPG and ACY1), and alignment of ArgE with members of this superfamily suggests that the ligands for the binuclear metal center are conserved.[140] ArgE copurifies with one bound Zn^{2+}, and the removal of the bound metal ion with chelating agents decreases activity, indicating that the bound metal ion is important for catalysis.[140] However, the physiological cofactor(s) for ArgE have not been identified. The isolated enzyme contains one bound Zn^{2+} per enzyme, and the activity of the purified enzyme is stimulated upon the addition of Zn^{2+} (twofold) or Co^{2+} (eightfold), consistent with ArgE functioning as a binuclear enzyme. The affinity of Zn^{2+}, Co^{2+}, and Mn^{2+} for the M_1 and M_2 sites has been determined as follows: $Zn_1 = 2.7\,\mu mol\,l^{-1}$, $Zn_2 = 51\,\mu mol\,l^{-1}$; $Co_1 = 0.4\,\mu mol\,l^{-1}$, $Co_2 = 153\,\mu mol\,l^{-1}$; $Mn_1 = 0.3\,\mu mol\,l^{-1}$, $Mn_2 = 5.3\,\mu mol\,l^{-1}$.[141,142] These affinity measurements suggest that Mn^{2+} may be the physiologically relevant metal ion(s). Metal titration experiments with Mn^{2+} suggest that only one Mn^{2+} serves a catalytic role, whereas binding of the second Mn^{2+} ion is either structural or inhibitory.[142] The relatively weak affinity of the metal ions for the second binding site may suggest that only one metal site is occupied under physiological conditions.

The mechanism of ArgE has been probed kinetically (in the presence of excess Co^{2+}).[140] The activity of ArgE has a bell-shaped dependence on pH (k_{cat}, k_{cat}/K_M), suggesting that at least two ionizations are important for maximal activity. For the parameter k_{cat}, the apparent pK_a values for the favorable deprotonation (pK_{a1}) and protonation (pK_{a2}) are 5.6 and 7.7, respectively. Both pK_a values are shifted in k_{cat}/K_M to 7.1 and 7.2, respectively, perhaps suggesting that these pK_a values are perturbed upon substrate binding. A deuterium solvent isotope effect of 2.1 is observed for the parameter k_{cat}, consistent with a partially rate-limiting proton transfer; a smaller solvent isotope effect is observed for k_{cat}/K_M (1.3). Proton inventory experiments suggest that the observed isotope effect describes the transfer of single proton. In light of the sequence alignments and structural data for the metalloproteases, the following mechanism has been proposed (**Figure 12(b)**),[140] where $M_1 = Zn^{2+}$ and $M_2 = Co^{2+}$. The carbonyl group on the substrate coordinates with the M_2 site metal ion. Next, the side chain of Glu144 (conserved) serves as a GBC to activate the M_1-water for attack on the carbonyl group to afford a tetrahedral intermediate, which is stabilized by coordination to M_2. A second active-site side chain then functions as a GAC to facilitate breakdown of the tetrahedral intermediate and product formation. This mechanism is interesting because it proposes the use of a GABC pair, and is in contrast to the mechanism of the metalloproteases. Additional experiments are needed to confirm that Glu144 functions as the GBC in this reaction and to identify a second side chain that functions as a GAC.

8.15.4.2.3 *N-succinyl-L,L-diaminopimelic acid desuccinylase (DapE)*

Bacteria synthesize lysine through one of three different pathways that are not present in mammals.[143] An intermediate in these pathways, *meso*-diaminopimelic acid (mDAP), is also an essential component of the bacterial cell wall.[144] Consequently, the enzymes in these pathways are targets for the development of antibiotics for treatment of Gram-negative and Gram-positive bacterial infections. The enzyme DapE catalyzes the hydrolysis of *N*-succinyl-L,L-diaminopimelic acid to form L,L-diaminopimelic acid and succinate in the succinylase pathway of lysine biosynthesis (**Figure 6(i)**).

DapE has significant sequence homology with ArgE, CPG, and ACY1, all binuclear enzymes that are members of the zinc α,β-hydrolase superfamily, but this enzyme contains only one bound zinc ion as isolated.[145] Treatment with a metal chelator (EDTA) decreases the observed activity, whereas the addition of Zn^{2+} or Co^{2+} to the purified enzyme enhances catalytic activity.[145,146] The activity (k_{cat}/K_M) follows the order: ZnCo > ZnZn > CoCo > CoZn, Zn, Co,[146] suggesting that DapE functions as a binuclear metalloenzyme. The mechanism of the DapE reaction has been probed using kinetic and biophysical studies.[145–148] Mutagenesis experiments indicate that Glu134 (conserved) is essential for catalysis. Additionally, the activity of DapE under k_{cat} conditions is dependent on a single ionization (favorable deprotonation) with apparent pK_a values of 6.6 and 5.9 for the Zn^{2+} and Co^{2+} enzymes, respectively, suggesting that this pK_a may reflect ionization of the

metal–water. In contrast, under k_{cat}/K_M conditions, DapE activity exhibits a bell-shaped dependence on pH, with apparent pK_a values of ∼6.5 and 8.3 for both metal ions. The similarity of the value of pK_{a1} (k_{cat}/K_M) to the pK_a observed under k_{cat} conditions suggests that pK_{a1} also describes the ionization of the metal–water under k_{cat}/K_M conditions, whereas pK_{a2} is proposed to reflect an ionization that is important for substrate binding. A significant inverse deuterium solvent isotope effect is observed for the DapE-catalyzed reaction ($k_{cat}^H/k_{cat}^D = 0.62$; $(k_{cat}/K_M)^H/(k_{cat}/K_M)^D = 0.78$), consistent with a metal-hydroxide mechanism. These data led to a proposed mechanism similar to that described for AAP (**Figure 12(a)**), with the side chain of Glu134 serving as the GABC analogous to Glu151 in the AAP mechanism.[145]

8.15.4.3   Nucleotide Metabolism

8.15.4.3.1   *Dihydroorotase*

Dihydroorotase (DHO) catalyzes the reversible hydrolytic conversion of dihydroorotate and carbamoyl aspartate (**Figure 6(e)**), an important step in pyrimidine nucleotide biosynthesis. The crystal structure of DHO from *E. coli*[149] has been solved and illuminates an overall fold described as a $(\beta/\alpha)_8$-barrel indicating that this enzyme is a member of the amidohydrolase superfamily. The active site of DHO contains a binuclear metal center that is occupied with two Zn^{2+} ions separated by 3.6 Å. The zinc ion bound to the M_α site coordinates with the side chains of His16, His18, Lys102, Asp250, and a solvent molecule in a trigonal bipyramidal geometry. The zinc ion bound at the M_β site coordinates with Lys102, His139, His177, and a water molecule in tetrahedral geometry. The bridging ligands for the two zinc ions are Lys102 (carbamoylated) and the solvent molecule. Other members of the amidohydrolase superfamily share a similar binuclear metal ligation and proposed reaction mechanisms, including PTE, iso-aspartyl dipeptidase (IAD) and possibly URE. The binuclear metal centers for these enzymes are occupied with either Zn^{2+} (PTE, IAD) or Ni^{2+} (URE).[26]

Structural information from DHO•substrate and DHO•product complexes have provided valuable insights into the catalytic mechanism of EcDHO (**Figure 14**)[26,149] that have been probed using mutagenesis and kinetic studies.[150] The activity of DHO is sensitive to the identity of the metal ions according to the order $Zn^{2+} > Co^{2+} > Cd^{2+}$. Additionally, DHO activity is dependent on a single ionization for the dihydroorotate (favorable deprotonation) and carbamoyl aspartate (favorable protonation) substrates with apparent pK_a values of ∼6 and 8.2, respectively. The pK_a values change upon substitution with Co^{2+}, suggesting that this reflects a metal-dependent ionization. There is a significant deuterium solvent isotope effect observed in k_{cat} with the dihydroorotate ($k_{cat}^H/k_{cat}^D = 2.5$; $(k_{cat}/K_M)^H/(k_{cat}/K_M)^D = 1.1$) and thio-dihydroorotate ($k_{cat}^H/k_{cat}^D = 2.3$; $(k_{cat}/K_M)^H/(k_{cat}/K_M)^D = 1.6$) substrates consistent with a rate-limiting proton transfer. Mutation of Asp250 (>10^4-fold) and His254 (>10^2-fold) to Ala dramatically decreases DHO activity indicating that these side chains play important roles in catalysis. The proposed mechanism for DHO is shown in **Figure 14**.[149,150] Coordination of the carbonyl group on the substrate to M_β serves to polarize the amide bond, thereby enhancing the electrophilicity of substrate. A direct interaction between the carbonyl group on the substrate and a metal ion is supported by the finding that k_{cat}/K_M for a thio-dihydroorotate substrate is five-fold higher with Cd-DHO compared to Zn-DHO, whereas for the oxygen containing dihydroorotate substrate the reaction is ∼10-fold faster with Zn-DHO compared to Cd-DHO. The activated carbonyl group is attacked by the bridging hydroxide to form a tetrahedral intermediate, which is stabilized by interaction with M_β and M_α.
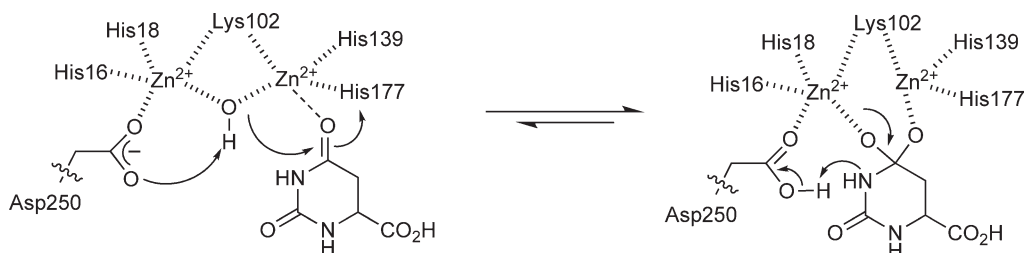
Figure 14   Proposed catalytic mechanism for DHO.

Finally, the side chain of Asp250 is positioned to shuttle the proton from the hydroxyl group to the amine to facilitate breakdown of the tetrahedral intermediate and product generation.

8.15.4.4 β-Lactam Metabolism – *Bacteroides fragilis* β-lactamase (Bfβl)

To date, numerous mononuclear (described above) and binuclear metallo-βl have been identified.[99,100] The βl from *B. fragilis* (Bfβl) is used here as an example of the binuclear enzymes. For additional information on β-lactamases, also see Chapter 8.13. The crystal structure of Bfβl reveals an overall topology similar to other members of the metallo-βl superfamily.[151] The active site of Bfβl contains a binuclear metal center that is occupied by two zinc ions (Zn–Zn separation 3.5 Å). The Zn_1 coordinates the side chains of His99, His101, His162, and a water molecule with tetrahedral geometry; Zn_2 coordinates with the side chains of Asp103, Cys181, His223, and two water molecules with trigonal bipyramidal geometry. A lone water molecule serves as the bridging ligand for this binuclear metal center. Although the ligands for Zn_2 are conserved in the βl from *B. cereus*, it is thought that this site remains unoccupied in Bcβl due to the low affinity for Zn^{2+}.[102,103] Two unusual features of this binuclear metal center are the use of Cys as a ligand and the absence of a bridging protein ligand. Treatment with a metal chelator (EDTA) inhibits activity, and this activity can be restored following the addition of Zn^{2+} or Co^{2+}.[152] Metal titration experiments suggest that the mono-Zn^{2+}-Bfβl retains $\sim$80% activity, whereas the binding of the second zinc modestly enhances catalytic activity.[152,153] These findings also suggest that the two metal binding sites have different affinities for Zn^{2+}. It has been suggested that the importance of the second metal ion may be underestimated, as the rates that are compared may reflect different steps. If the rates of reaction for the chemical step (C–N bond cleavage) are compared, the addition of the second Zn^{2+} appears to enhance the rate of catalysis by $\sim$80-fold.[154]

The mechanism of Bfβl has been probed using kinetic studies.[153–155] Mutagenesis experiments indicate that the conserved Asp103 side chain is important for activity. The activity of Bfβl is pH-dependent, wherein the observed favorable deprotonation (p$K_a \sim$ 5.25), is proposed to reflect formation of the Zn_1-hydroxide nucleophile. Furthermore, deuterium solvent isotope effect experiments (k_{cat}^H / k_{cat}^D = 2.4–2.8) are also consistent with a proton transfer in the rate-limiting step of the reaction. In contrast to other enzymes described above, there is no second ionization observed in the pH rate profile that can be attributed to a GAC (pH 5–10 range). The addition of anions (i.e., fluoride, azide, acetate, cyanate) accelerates the Bfβl reaction. Since these anions do not strongly interact/coordinate with the metal center, they are proposed to facilitate breakdown of the reaction intermediate. Spectroscopic evidence supports the accumulation of a negatively charged enzyme bound intermediate during the reaction.[156] In light of these data, the following mechanism has been proposed for Bfβl and may be applicable to other binuclear βls (**Figure 11(b)**).[154] The substrate carbonyl group coordinates with Zn_1 and the lactam nitrogen coordinates with Zn_2, thus replacing the water ligand and orienting the substrate for nucleophilic attack. Substrate binding is also accompanied by the loss of the bridging water from Zn_2 to form terminally a bound water at Zn_1. This mode of binding serves to polarize the carbonyl group (aided by donation of a hydrogen bond from conserved Asn193) and to activate the amine leaving group on the substrate. The Zn_1-hydroxide attacks the carbonyl carbon to generate a tetrahedral intermediate that is stabilized by coordination with Zn_1 and Zn_2. Breakdown of this intermediate and cleavage of the C–N bond generates negatively charged N leaving group that is stabilized by coordination with Zn_2. Protonation of the amine nitrogen and ligand exchange of the acyl group at Zn_1 with water leads to product formation and dissociation. The proposed role of Asp103 is to orient Zn_2 and hydroxide during the reaction, and to help balance the charge of the active site.

8.15.5 Conclusions

Metal ion cofactors have varied roles to enhance the catalytic efficiency of enzymes in hydrolytic reactions, including facilitate substrate binding (water and organic substrate), gathering/template effects, function as an electrostatic catalyst (carbonyl polarization and transition state stabilization), function as a Lewis acid to lower the pK_a of metal–water and stabilize the formation of the leaving group. Although their properties make several

transition metal ions capable of serving as cofactors for these reactions, we see that Zn^{2+} emerges as the most widely used metal ion for these purposes. Several of the enzymes discussed belong to larger superfamilies, such as the amidohydrolase and zinc-α,β-hydrolase superfamilies, indicating that these enzymes may be evolutionarily related. However, it is not clear how the mononuclear and binuclear enzymes evolved from one another. Although members of enzyme superfamilies share a similar overall fold, these enzymes can have different metal ion requirements for catalysis. For example, there are both mononuclear (e.g., NagA) and binuclear (e.g., DHO) members of the amidohydrolase superfamily, and there are enzymes in this superfamily that utilize Zn^{2+} (e.g., DHO), Fe^{2+} (e.g., CDA) and Ni^{2+} (e.g., URE) as cofactors for hydrolysis. These differences in metal ion preferences can also be extended to include differences in catalytic mechanisms. Interestingly, there is increasing evidence to suggest that individual enzymes may have species-specific metal ion requirements, as observed for PDF, NagA, MetAP, and possibly LpxC and HDACs. The factors that dictate cofactor preferences for these enzymes are not known, but may in part be influenced by differences in metal ion availability. Finally, in terms of participation by enzyme side chains to assist the metal ion cofactor(s) in hydrolytic reactions, there are two main types of catalysts that emerge: those that use a single bifunctional GABC and those that use a GABC pair to facilitate catalysis.

Abbreviations

Aa	*A. aeolicus*
AAP	aminopeptidase from *Aeromonas proteolytica*
ACE	angiotensin converting enzyme
ACY1	aminoacylase-1
ADA	adenosine deaminase
APH	acetylpolyamine amidohydrolase
ArgE	*N*-acetyl ornithine deacetylase
Bc	*Bacillus cereus*
Bf	*Bacteroides fragilis*
BLAP	bovine leucine aminopeptidase
Bs	*Bacillus subtilis*
CDA	cytosine deaminase
CN	coordination number
CPA	carboxypeptidase A
DapE	*N*-succinyl-L,L-diaminopimelic acid desuccinylase
DFT	density functional theory
DHO	dihydroorotase
DPA	dipicolinic acid
Ec	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EXAFS	extended X-ray absorption fine structure
GABC	general acid–base catalysts
GAC	general acid catalyst
GBC	general base catalyst
HDAC	histone deacetylase
HDLP	histone deacetylase-like proteins
HDPR	6-hydroxyl-1,6-dihydropurine riboside
Hs	*Homo sapiens*
HSAB	hard–soft acid–base
IAD	iso-aspartyl dipeptidase
I_n	ionic potential
LPS	lipopolysaccharide
LpxC	UDP-3-*O*-(*R*-3-hydroxymyristoyl)-*N*-acetylglucosamine deacetylase
mDAP	*meso*-diaminopimelic acid

MDD	metal-dependent deacetylase
MetAP	methionine aminopeptidase
MMP	matrix metalloproteases
MSH	mycothiol
MshB	*N*-acetyl-1-D-myo-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase
NagA	*N*-acetylglucosamine-6-phosphate deacetylase
PDF	peptide deformylase
PTE	phosphotriesterase
r	ionic radius
Tm	*Thermatoga maritima*
UDP	uridine diphosphate
URE	urease
z	charge
ZBP	zinc-binding protein
βI	β-lactamase

References

1. J. J. R. Frausto da Silva; R. J. P. Williams, *The Biological Chemistry of the Elements. The Inorganic Chemistry of Life*, 2nd ed.; Oxford University Press: New York, 2001.
2. S. J. Lippard; J. M. Berg, *Principles of Bioinorganic Chemistry*; University Science Books: Mill Valley, California, 1994.
3. W. Kaim; B. Schwederski, *Bioinorganic Chemistry: Inorganic Elements in the Chemistry of Life*, 1st ed.; John Wiley & Sons Ltd: Chichester, 1994.
4. F. E. Jacobsen; J. A. Lewis; S. M. Cohen, *ChemMedChem* **2007**, *2*, 152–171.
5. R. Schiffmann; A. Heine; G. Klebe; C. D. P. Klein, *Angewandte Chem. Int. Ed.* **2005**, *44*, 3620–3623.
6. R. J. White; P. S. Margolis; J. Trias; Z. Y. Yuan, *Curr. Opin. Pharmacol.* **2003**, *3*, 502–507.
7. Z. Y. Yuan; J. Trias; R. J. White, *Drug Discov. Today* **2001**, *6*, 954–961.
8. A. Agrawal; D. Romero-Perez; J. A. Jacobsen; F. J. Villarreal; S. M. Cohen, *ChemMedChem* **2008**, *3*, 812–820.
9. C. T. Supuran; A. Casini; A. Scozzafava, *Med. Res. Rev.* **2003**, *23*, 535–558.
10. J. L. Hu; P. E. Van den Steen; Q. X. A. Sang; G. Opdenakker, *Nat. Rev. Drug Discov.* **2007**, *6*, 480–498.
11. C. R. H. Raetz; C. Whitfield, *Annu. Rev. Biochem.* **2002**, *71*, 635–700.
12. M. Anderson; H. Bull; S. Galloway; T. Kelly; S. Mohan; K. Radika; C. Raetz, *J. Biol. Chem.* **1993**, *268*, 19858–19865.
13. D. A. Whittington; K. M. Rusche; H. Shin; C. A. Fierke; D. W. Christianson, *Proc. Natl. Acad. Sci.* **2003**, *100*, 8146–8150.
14. B. E. Coggins; X. C. Li; A. L. McClerren; O. Hindsgaul; C. R. H. Raetz; P. Zhou, *Nat. Struct. Biol.* **2003**, *10*, 645–651.
15. J. E. Jackman; C. R. H. Raetz; C. A. Fierke, *Biochemistry* **1999**, *38*, 1902–1911.
16. J. E. Jackman; C. R. H. Raetz; C. A. Fierke, *Biochemistry* **2001**, *40*, 514–523.
17. C. P. McClure; K. M. Rusche; K. Peariso; J. E. Jackman; C. A. Fierke; J. E. Penner-Hahn, *J. Inorg. Biochem.* **2003**, *94*, 78–85.
18. M. Hernick; C. A. Fierke, *Biochemistry* **2006**, *45*, 14573–14581.
19. M. Hernick; C. A. Fierke, *Arch. Biochem. Biophys.* **2005**, *433*, 71–84.
20. M. Hernick; H. A. Gennadios; D. A. Whittington; K. M. Rusche; D. W. Christianson; C. A. Fierke, *J. Biol. Chem.* **2005**, *280*, 16969–16978.
21. A. L. McClerren; P. Zhou; Z. Guan; C. R. H. Raetz; J. Rudolph, *Biochemistry* **2005**, *44*, 1106–1113.
22. M. Hernick; C. A. Fierke, *Biochemistry* **2006**, *45*, 15240–15248.
23. B. E. Coggins; A. L. McClerren; L. Jiang; X. Li; J. Rudolph; O. Hindsgaul; C. R. H. Raetz; P. Zhou, *Biochemistry* **2005**, *44*, 1114–1126.
24. J. J. Robinet; J. W. Gauld, *J. Phys. Chem. B* **2008**, *112*, 3462–3469.
25. T. Uehara; K. Suefuji; N. Valbuena; B. Meehan; M. Donegan; J. T. Park, *J. Bacteriol.* **2005**, *187*, 3643–3649.
26. C. M. Seibert; F. M. Raushel, *Biochemistry* **2005**, *44*, 6383–6391.
27. L. Holm; C. Sander, *Proteins* **1997**, *28*, 72–82.
28. F. M. Ferreira; G. Mendoza-Hernandez; M. Castaneda-Bueno; R. Aparicio; H. Fischer; M. L. Calcagno; G. Oliva, *J. Mol. Bio.* **2006**, *359*, 308–321.
29. R. S. Hall; S. Brown; A. A. Fedorov; E. V. Fedorov; C. F. Xu; P. C. Babbitt; S. C. Almo; F. M. Raushel, *Biochemistry* **2007**, *46*, 7953–7962.
30. F. Vincent; D. Yates; E. Garman; G. J. Davies; J. A. Brannigan, *J. Biol. Chem.* **2004**, *279*, 2809–2816.
31. R. S. Hall; D. F. Xiang; C. F. Xu; F. M. Raushel, *Biochemistry* **2007**, *46*, 7942–7952.
32. K. T. Nguyen; J. C. Wu; J. A. Boylan; F. C. Gherardini; D. Pei, *Arch. Biochem. Biophys.* **2007**, *468*, 217–225.
33. J. Wang; G. S. Sheppard; P. Lou; M. Kawai; C. Park; D. A. Egan; A. Schneider; J. Bouska; R. Lesniewski; J. Henkin, *Biochemistry* **2003**, *42*, 5035–5042.
34. S. L. Clugston; J. F. J. Barnard; R. Kinach; D. Miedema; R. Ruman; E. Daub; J. F. Honek, *Biochemistry* **1998**, *37*, 8754–8763.

35. S. L. Gantt; S. G. Gattis; C. A. Fierke, *Biochemistry* **2006**, *45*, 6170–6178.
36. G. L. Newton; K. Arnold; M. S. Price; C. Sherrill; S. B. Delcardayre; Y. Aharonowitz; G. Cohen; J. Davies; R. C. Fahey; C. Davis, *J. Bacteriol.* **1996**, *178*, 1990–1995.
37. G. L. Newton; R. C. Fahey, *Arch. Microbiol.* **2002**, *178*, 388–394.
38. D. Sareen; G. L. Newton; R. C. Fahey; N. A. Buchmeier, *J. Bacteriol.* **2003**, *185*, 6736–6740.
39. M. Rawat; G. L. Newton; M. Ko; G. J. Martinez; R. C. Fahey; Y. Av-Gay, *Antimicrob. Agents Chemother.* **2002**, *46*, 3348–3355.
40. G. L. Newton; P. Ta; K. P. Bzymek; R. C. Fahey, *J. Biol. Chem.* **2006**, *281*, 33910–33920.
41. J. T. Maynes; C. Garen; M. M. Cherney; G. Newton; D. Arad; Y. Av-Gay; R. C. Fahey; M. N. G. James, *J. Biol. Chem.* **2003**, *278*, 47166–47170.
42. G. L. Newton; M. Ko; P. Ta; Y. Av-Gay; R. C. Fahey, *Protein Expr. Purif.* **2006**, *47*, 542–550.
43. V. E. Fadouloglou; A. Deli; N. M. Glykos; E. Psylinakis; V. Bouriotis; M. Kokkinidis, *FEBS J.* **2007**, *274*, 3044–3054.
44. W. T. Lowther; B. W. Matthews, *Chem. Rev.* **2002**, *102*, 4581–4607.
45. W. N. Lipscomb; N. Strater, *Chem. Rev.* **1996**, *96*, 2375–2433.
46. D. S. Auld, *Biometals* **2001**, *14*, 271–313.
47. S. Aoki; E. Kimura, *Compr. Coord. Chem. II* **2004**, *8*, 601–640.
48. K. S. Makarova; N. V. Grishin, *J. Mol. Biol.* **1999**, *292*, 11–17.
49. M. A. Wouters; A. Husain, *J. Mol. Biol.* **2001**, *314*, 1191–1207.
50. D. W. Christianson; W. N. Lipscomb, *Acc. Chem. Res.* **1989**, *22*, 62–69.
51. J. E. Coleman; B. L. Vallee, *J. Biol. Chem.* **1960**, *235*, 390–395.
52. J. E. Coleman; B. L. Vallee, *J. Biol. Chem.* **1961**, *236*, 2244–2249.
53. K. S. Larsen; D. S. Auld, *Biochemistry* **1991**, *30*, 2613–2618.
54. M. GomezOrtiz; F. X. GomisRuth; R. Huber; F. X. Aviles, *FEBS Lett.* **1997**, *400*, 336–340.
55. D. C. Rees; M. Lewis; W. N. Lipscomb, *J. Mol. Biol.* **1983**, *168*, 367–387.
56. K. Zhang; B. Chance; D. S. Auld; K. S. Larsen; B. L. Vallee, *Biochemistry* **1992**, *32*, 1159–1168.
57. T. Meinnel; Y. Mechulam; S. Blanquet, *Biochimie* **1993**, *75*, 1061–1075.
58. K. T. Nguyen; X. Hu; C. Colton; R. Chakrabarti; M. X. Zhu; D. Pei, *Biochemistry* **2003**, *42*, 9952–9958.
59. D. Groche; A. Becker; I. Schlichting; W. Kabsch; S. Schultz; A. F. V. Wagner, *Biochem. Biophys. Res. Commun.* **1998**, *246*, 342–346.
60. P. T. R. Rajagopalan; X. C. Yu; D. H. Pei, *J. Am. Chem. Soc.* **1997**, *119*, 12418–12419.
61. M. K. Chan; W. Gong; P. T. R. Rajagopalan; B. Hao; C. M. Tsai; D. Pei, *Biochemistry* **1997**, *36*, 13904–13909.
62. P. T. R. Rajagopalan; S. Grimme; D. H. Pei, *Biochemistry* **2000**, *39*, 779–790.
63. H. Deng; R. Callender; J. Zhu; K. T. Nguyen; D. Pei, *Biochemistry* **2002**, *41*, 10563–10569.
64. R. K. Jain; B. Hao; R. P. Liu; M. K. Chan, *J. Am. Chem. Soc.* **2005**, *127*, 4558–4559.
65. M. W. Anders; W. Dekant, *Adv. Pharmacol.* **1994**, *27*, 431–448.
66. S. H. Liaw; S. J. Chen; T. P. Ko; C. S. Hsu; C. J. Chen; A. H. J. Wang; Y. C. Tsai, *J. Biol. Chem.* **2003**, *278*, 4957–4962.
67. H. A. Lindner; V. V. Lunin; A. Alary; R. Hecker; M. Cygler; R. Menard, *J. Biol. Chem.* **2003**, *278*, 44496–44504.
68. W. L. Lai; L. Y. Chou; C. Y. Ting; R. Kirby; Y. C. Tsai; A. H. J. Wang; S. H. Liaw, *J. Biol. Chem.* **2004**, *279*, 13962–13967.
69. C. S. Hsu; W. L. Lai; W. W. Chang; S. H. Liaw; Y. C. Tsai, *Protein Sci.* **2002**, *11*, 2545–2550.
70. M. Wakayama; H. Yada; S. Kanda; S. Hayashi; Y. Yatsuda; K. Sakai; M. Moriguchi, *Biosci. Biotech. Biochem.* **2000**, *64*, 1–8.
71. D. D. Leipe; D. Landsman, *Nucleic Acids Res.* **1997**, *25*, 3693–3697.
72. C. A. Hassig; S. L. Schreiber, *Curr. Opin. Chem. Biol.* **1997**, *1*, 300–308.
73. C. Hildmann; D. Riester; A. Schwienhorst, *Appl. Microbiol. Biotechnol.* **2007**, *75*, 487–497.
74. I. Gregoretti; Y.-M. Lee; H. V. Goodson, *J. Mol. Biol.* **2004**, *338*, 17–31.
75. J. R. Somoza; R. J. Skene; B. A. Katz; C. Mol; J. D. Ho; A. J. Jennings; C. Luong; A. Arvai; J. J. Buggy; E. Chi, *Structure* **2004**, *12*, 1325–1334.
76. M. S. Finnin; J. R. Donigian; A. Cohen; V. M. Richon; R. A. Rifkind; P. A. Marks; R. Breslow; N. P. Pavletich, *Nature* **1999**, *401*, 188–193.
77. T. K. Nielsen; C. Hildmann; A. Dickmanns; A. Schwienhorst; R. Ficner, *J. Mol. Biol.* **2005**, *354*, 107–120.
78. A. Schuetz; J. Min; A. Allali-Hassani; M. Schapira; M. Shuen; P. Loppnau; R. Mazitschek; N. P. Kwiatkowski; T. A. Lewis; R. L. Maglathin; T. H. McLean; A. Bochkarev; A. N. Plotnikov; M. Vedadi; C. H. Arrowsmith, *J. Biol. Chem.* **2008**, *283*, 11355–11363.
79. C. A. Hassig; J. K. Tong; T. C. Fleischer; T. Owa; P. G. Grable; D. E. Ayer; S. L. Schreiber, *Proc. Nat. Acad. Sci. U.S.A.* **1998**, *95*, 3519–3524.
80. B. E. Schultz; S. Misialek; J. Wu; J. Tang; M. T. Conn; R. Tahilramani; L. Wong, *Biochemistry* **2004**, *43*, 11083–11091.
81. D. Kadosh; K. Struhl, *Genes Dev.* **1998**, *12*, 797–805.
82. K. Moreth; D. Riester; C. Hildmann; R. Hempel; D. Wegenert; A. Schobert; A. Schwienhorst, *Biochem. J.* **2007**, *401*, 659–665.
83. C. Corminboeuf; P. Hu; M. E. Tuckerman; Y. Zhang, *J. Am. Chem. Soc.* **2006**, *128*, 4530–4531.
84. L. J. Marton; A. E. Pegg, *Annu. Rev. Pharmacol. Toxicol.* **1995**, *35*, 55–91.
85. C. Moinard; L. Cynober; J. P. De Bant, *Clin. Nutr.* **2005**, *24*, 184–197.
86. K. Sakurada; T. Ohta; K. Fujishiro; M. Hasegawa; K. Aisaka, *J. Bacteriol.* **1996**, *178*, 5781–5786.
87. A. P. Gerber; W. Keller, *Trends Biochem. Sci.* **2001**, *26*, 376–384.
88. M. Ohman, *Biochimie* **2007**, *89*, 1171–1176.
89. D. K. Wilson; F. B. Rudolph; F. A. Quiocho, *Science* **1991**, *252*, 1278–1284.
90. G. C. Ireton; G. McDermott; M. E. Black; B. L. Stoddard, *J. Mol. Biol.* **2002**, *315*, 687–697.
91. D. J. T. Porter; E. A. Austin, *J. Biol. Chem.* **1993**, *268*, 24005–24011.
92. B. F. Cooper; V. Sideraki; D. K. Wilson; D. Y. Dominguez; S. W. Clark; F. A. Quiocho; F. B. Rudolph, *Protein Sci.* **1997**, *6*, 1031–1037.
93. Z. Wang; F. A. Quiocho, *Biochemistry* **1998**, *37*, 8314–8324.
94. D. K. Wilson; F. A. Quiocho, *Biochemistry* **1993**, *32*, 1689–1694.

95. V. Sideraki; D. K. Wilson; L. C. Kurz; F. A. Quiocho; F. B. Rudolph, *Biochemistry* **1996**, *35*, 15019–15028.
96. V. Sideraki; K. A. Mohamedali; D. K. Wilson; Z. Chang; R. E. Kellems; F. A. Quiocho; F. B. Rudolph, *Biochemistry* **1996**, *35*, 7862–7872.
97. K. A. Mohamedali; L. C. Kurz; F. B. Rudolph, *Biochemistry* **1996**, *35*, 1672–1680.
98. M. Babic; A. M. Hujer; R. A. Bonomo, *Drug Resist. Updat.* **2006**, *9*, 142–156.
99. J. A. Cricco; E. G. Orellano; R. M. Rasia; E. A. Ceccarelli; A. J. Vila, *Coord. Chem. Rev.* **1999**, *192*, 519–535.
100. C. Bebrone, *Biochem. Pharmacol.* **2007**, *74*, 1686–1701.
101. A. Carfi; S. Pares; E. Duee; M. Galleni; C. Duez; J. M. Frere; O. Dideberg, *EMBO J.* **1995**, *14*, 4914–4921.
102. D. de Seny; U. Heinz; S. Wommer; M. Kiefer; W. Meyer-Klaucke; M. Galleni; J.-M. Frere; R. Bauer; H.-W. Adolph, *J. Biol. Chem.* **2001**, *276*, 45065–45078.
103. S. Wommer; S. Rival; U. Heinz; M. Galleni; J.-M. Frere; N. Franceschini; G. Amicosante; B. Rasmussen; R. Bauer; H.-W. Adolph, *J. Biol. Chem.* **2002**, *277*, 24142–24147.
104. L. I. Llarrull; S. M. Fabiane; J. M. Kowalski; B. Bennett; B. J. Sutton; A. J. Vila, *J. Biol. Chem.* **2007**, *282*, 18276–18285.
105. S. Bounaga; A. P. Laws; M. Galleni; M. I. Page, *Biochem. J.* **1998**, *331*, 703–711.
106. A. Badarau; M. I. Page, *Biochemistry* **2006**, *45*, 10654–10666.
107. D. E. Wilcox, *Chem. Rev.* **1996**, *96*, 2435–2458.
108. R. C. Holz, *Coord. Chem. Rev.* **2002**, *232*, 5–26.
109. B. Chevrier; C. Schalk; H. D'Orchymont; J. M. Rondeau; D. Moras; C. Tarnus, *Structure* **1994**, *2*, 283–291.
110. J. M. Prescott; S. H. Wilkes; F. W. Wagner; K. J. Wilson, *J. Biol. Chem.* **1971**, *246*, 1756–1764.
111. J. M. Prescott; F. W. Wagner; B. Holmquist; B. L. Vallee, *Biochemistry* **1985**, *24*, 5350–5356.
112. M. E. Bayliss; J. M. Prescott, *Biochemistry* **1986**, *25*, 8113–8117.
113. J. M. Prescott; F. W. Wagner; B. Holmquist; B. L. Vallee, *Biochem. Biophys. Res. Commun.* **1983**, *114*, 646–652.
114. J. O. Baker; J. M. Prescott, *Biochemistry* **1983**, *22*, 5322–5331.
115. B. Chevrier; H. D'Orchymont; C. Schalk; C. Tarnus; D. Moras, *FEBS. J.* **1996**, *237*, 393–398.
116. S. K. Burley; P. R. David; R. M. Sweet; A. Taylor; W. N. Lipscomb, *J. Mol. Biol.* **1992**, *224*, 113–140.
117. H. Kim; W. N. Lipscomb, *Biochemistry* **1993**, *32*, 8465–8478.
118. N. Strater; W. N. Lipscomb, *Biochemistry* **1995**, *34*, 9200–9210.
119. N. Strater; W. N. Lipscomb, *Biochemistry* **1995**, *34*, 14792–14800.
120. F. H. Carpenter; J. M. Vahl, *J. Biol. Chem.* **1973**, *248*, 294–304.
121. G. A. Thompson; F. H. Carpenter, *J. Biol. Chem.* **1976**, *251*, 53–60.
122. G. A. Thompson; F. H. Carpenter, *J. Biol. Chem.* **1976**, *251*, 1618–1624.
123. N. Strater; L. Sun; E. R. Kantrowitz; W. N. Lipscomb, *Proc. Natl. Acad. Sci. U.S.A* **1999**, *96*, 11151–11155.
124. R. C. Holz; K. P. Bzymek; S. I. Swierczek, *Curr. Opin. Chem. Biol.* **2003**, *7*, 197–206.
125. J. L. McCullough; B. A. Chabner; J. R. Bertino, *J. Biol. Chem.* **1971**, *246*, 7207–7213.
126. S. M. Arfin; R. L. Kendall; L. Hall; L. H. Weaver; A. E. Stewart; B. W. Matthews; R. A. Bradshaw, *Proc. Natl. Acad. Sci. U.S.A* **1995**, *92*, 7714–7718.
127. X. V. Hu; X. Chen; K. C. Han; A. S. Mildvan; J. O. Liu, *Biochemistry* **2007**, *46*, 12833–12843.
128. L. Meng; S. Ruebush; V. M. D'souza; A. J. Copik; S. Tsunasawa; R. C. Holz, *Biochemistry* **2002**, *41*, 7199–7208.
129. V. M. D'Souza; R. C. Holz, *Biochemistry* **1999**, *38*, 11079–11085.
130. Z. G. Liu; Z. L. Zhen; Z. Y. Zuo; Y. L. Wu; A. F. Liu; Q. M. Yi; W. X. Li, *J. Biochem.* **2006**, *139*, 421–430.
131. D. E. Ash; J. D. Cox; D. W. Christianson, In *Metal Ions in Biological Systems: Manganese and Its Role in Biological Processes*; A. Sigel, H. Sigel, Eds.; CRC Press: 2000; Vol. 37, pp 407–428.
132. D. E. Ash, *J. Nutr.* **2004**, *134*, 2760S–2764S.
133. D. W. Christianson, *Acc. Chem. Res.* **2005**, *38*, 191–201.
134. Z. F. Kanyo; L. R. Scolnick; D. E. Ash; D. W. Christianson, *Nature* **1996**, *383*, 554–557.
135. L. R. Scolnick; Z. F. Kanyo; R. C. Cavalli; D. E. Ash; D. W. Christianson, *Biochemistry* **1997**, *36*, 10558–10565.
136. E. Cama; F. A. Emig; D. E. Ash; D. W. Christianson, *Biochemistry* **2003**, *42*, 7748–7758.
137. N. J. Kuhn; J. Talbot; S. Ward, *Arch. Biochem. Biophys.* **1991**, *286*, 217–221.
138. N. J. Kuhn; S. Ward; M. Piponski; T. M. Young, *Arch. Biochem. Biophys.* **1995**, *320*, 24–34.
139. R. H. Davis, *Microbiol. Rev.* **1986**, *50*, 280–313.
140. F. Javid-Majd; J. S. Blanchard, *Biochemistry* **2000**, *39*, 1285–1293.
141. W. C. McGregor; S. I. Swierczek; B. Bennett; R. C. Holz, *J. Am. Chem. Soc.* **2005**, *127*, 14100–14107.
142. W. C. McGregor; S. I. Swierczek; B. Bennett; R. C. Holz, *J. Biol. Inorg. Chem.* **2007**, *12*, 603–613.
143. G. Scapin; J. S. Blanchard, *Adv. Enzymol. Rel. Areas of Mol. Biol.* **1998**, *1998*, 279–324.
144. T. L. Born; J. S. Blanchard, *Curr. Opin. Chem. Biol.* **1999**, *3*, 607–613.
145. T. L. Born; R. J. Zheng; J. S. Blanchard, *Biochemistry* **1998**, *37*, 10478–10487.
146. D. L. Bienvenue; D. M. Gilner; R. S. Davis; B. Bennett; R. C. Holz, *Biochemistry* **2003**, *42*, 10756–10763.
147. D. H. Broder; C. G. Miller, *J. Bacteriol.* **2003**, *185*, 4748–4754.
148. R. Davis; D. Bienvenue; S. I. Swierczek; D. M. Gilner; L. Rajagopal; B. Bennett; R. C. Holz, *J. Biol. Inorg. Chem.* **2006**, *11*, 206–216.
149. J. B. Thoden; G. N. Phillips; T. M. Neal; F. M. Raushel; H. M. Holden, *Biochemistry* **2001**, *40*, 6989–6997.
150. T. N. Porter; Y. Li; F. M. Raushel, *Biochemistry* **2004**, *43*, 16285–16292.
151. N. O. Concha; B. A. Rasmussen; K. Bush; O. Herzberg, *Structure* **1996**, *4*, 823–836.
152. Z. Wang; S. J. Benkovic, *J. Biol. Chem.* **1998**, *273*, 22402–22408.
153. M. W. Crowder; Z. Wang; S. L. Franklin; E. P. Zovinka; S. J. Benkovic, *Biochemistry* **1996**, *35*, 12126–12132.
154. Z. Wang; W. Fast; S. J. Benkovic, *Biochemistry* **1999**, *38*, 10013–10023.
155. W. Fast; Z. Wang; S. J. Benkovic, *Biochemistry* **2001**, *40*, 1640–1650.
156. Z. Wang; W. Fast; S. J. Benkovic, *J. Am. Chem. Soc.* **1998**, *120*, 10788–10789.

Biographical Sketches

Marcy Hernick was born in Troy, NY in 1975. After obtaining a B.S. in Pharmacy at Albany College of Pharmacy in 1998, she earned a Ph.D., in Medicinal Chemistry at Purdue University in 2002 working in the laboratory of Richard Borch on the development of phosphoramidate prodrugs as anticancer agents. As a postdoc in Carol Fierke's laboratory at the University of Michigan, she studied the mechanism of the metal-dependent deacetylase LpxC. In 2007, she joined the faculty at Virginia Tech in the Department of Biochemistry and is currently working on examining the mechanisms of metal-dependent hydrolases and the development of inhibitors against these enzymes.

Carol Fierke was born in Stillwater, MN in 1955. After obtaining a B.A. in Chemistry at Carleton College in 1978, she earned a Ph.D., in Biochemistry at Brandeis University in 1984 working in the laboratory of William Jencks. After completion of a postdoctoral fellowship in Stephen Benkovic's laboratory at Pennsylvania State University, she began her academic career in the Biochemistry Department at the Duke University Medical Center in 1987, where she received multiple awards, including the American Cancer Society Junior Faculty Research Award, the Packard Foundation Fellowship, and the American Heart Association Established Investigator, and was promoted to associate professor with tenure. She moved to the University of Michigan in 1999 as Professor of Chemistry and Biological Chemistry. At Michigan, she has been awarded a Faculty Achievement Award (2001), the Jerome and Isabella Collegiate Professorship (2003), a Distinguished Faculty Achievement Award (2005), the Sarah Power Goddard Award (2005), and elected a Fellow of the AAAS (2006). She currently serves as the Chair of the Chemistry Department and the Chair for the Biological Chemistry Division of the American Chemical Society. She also serves on the Editorial Boards of *Chemical Biology & Drug Design*, *Biochemistry* and *Biopolymers*. She has published nearly two hundred research articles and reviews. Her research integrates biochemistry, cell biology, structural biology, and drug design. Her current research interests include the

catalytic mechanism, biological function, and inhibition of metalloenzymes that catalyze post-translational lipidation and deacetylation that are potential antibiotic and anticancer targets; structure and function of protein–RNA complexes; homeostasis of intracellular metal ions; and the development of protein-based sensors for the measurement of intracellular analytes, including metal ions.

8.16 Dioxygenase Enzymes and Oxidative Cleavage Pathways

Timothy D. H. Bugg, University of Warwick, Coventry, UK

8.16.1 Bacterial Aromatic Degradation Pathways

8.16.1.1 Oxidative Pathways for Aromatic Degradation

Soil is extremely rich in bacteria (10^6–10^8 cells per g soil), which survive by utilizing nutrients and carbon sources present there. Naturally occurring aromatic compounds are present in soil from the breakdown of lignin from woody plants and phenylpropanoids found in plants; the aromatic fraction of leached oil and coal; as well as the aromatic amino acids L-phenylalanine, L-tyrosine, and L-tryptophan from protein breakdown (**Figure 1**). Man-made aromatic compounds include pesticides, detergents, oils, solvents, paints, and explosives.

Many simple aromatic compounds such as benzene, toluene, xylenes, benzoic acid, phenylacetic acid, and phenylpropionic acid are degraded by aerobic soil bacteria such as *Pseudomonas*, *Acinetobacter*, and *Rhodococcus*.[1] Phenylacetic acid and phenylpropionic acid are also degraded by *Escherichia coli*, an enteric bacterium.[2] Many man-made compounds in the environment can be degraded via the same pathways used by microorganisms for degradation of naturally occurring aromatic compounds, for example, the insecticide carbaryl can be degraded via the bacterial naphthalene degradation pathway.[3] However, some man-made chemicals are highly persistent in the environment, especially chlorinated aromatics, whose degradation is discussed below. Examples of highly persistent chlorinated aromatics include the insecticide dichloro-diphenyl-trichloroethane (DDT), the fungicide pentachlorophenol, and the industrial chemicals trichloroethylene (TCE) and tetrachloroethylene (also known as perchloroethylene (PCE)), used for the manufacture of polyvinyl chloride (PVC) plastics.
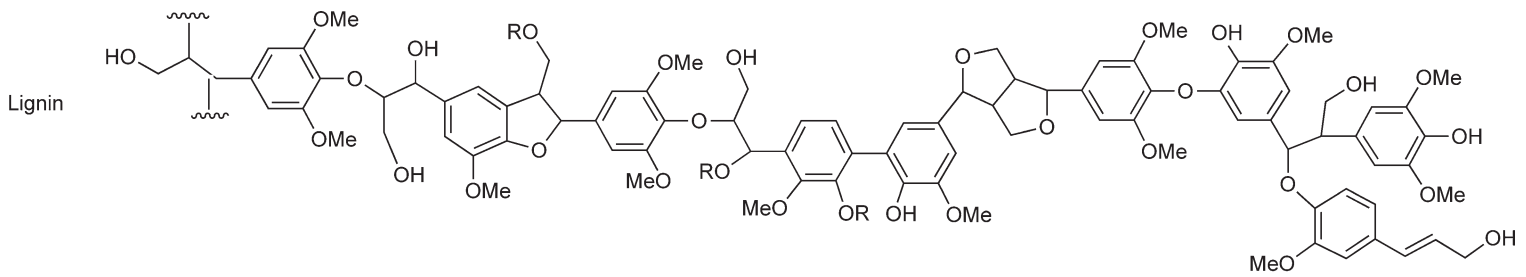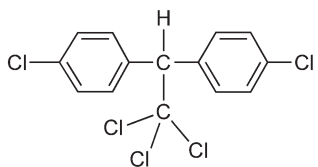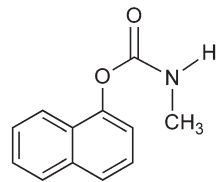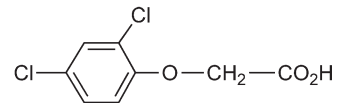
Figure 1 Aromatic compounds present in soil.
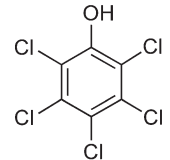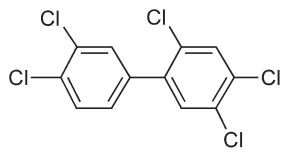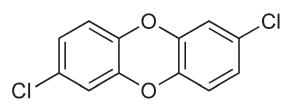
Two classes of highly persistent aromatic compounds are the polychlorinated biphenyls (PCBs) and the polychlorinated benzodioxins (PCDDs). PCBs have been widely used in transformer oils, heat transfer fluids, dielectric fluids, and plasticizers. Degradation of PCBs is very slow, due to their very low solubility and low reactivity, but is further complicated by their presence as a complex mixture of isomers.[4] Their industrial synthesis involves chlorination of biphenyl under high temperature and pressure, resulting in a mixture of all of the 208 possible isomers, which hinders their enzymatic degradation, since enzymes are inherently selective in their action. Certain bacterial strains, such as *Pseudomonas* sp. LB400, are able to degrade lightly chlorinated PCBs, but heavily chlorinated PCBs are extremely persistent in the environment.[4]

Small quantities of PCDDs are produced as by-products of the industrial manufacture of chlorinated phenolic chemicals and pesticides.[5] They are especially hazardous environmental pollutants, since they have carcinogenic and teratogenic properties. Their release in 1976 in an industrial explosion at Seveso, Italy was implicated in the widespread poisoning of livestock and contamination of the local population.

Many of the pathways used by soil bacteria to degrade aromatic compounds were discovered through the work of Stanley Dagley (1916–87), and are summarized in a well-written article that he published in 1975.[1] More recently, the University of Minnesota Biocatalysis and Biodegradation Database (http://umbbd.msi.umn. edu) has compiled nearly 200 pathways for microbial degradation of aromatic and aliphatic compounds, in written and graphic formats.[6] The following section will present a selection of the more common strategies for bacterial aromatic degradation.

Most of the pathways for bacterial degradation are inducible by the presence of the aromatic compound, and expression of a gene cluster for the degrading enzymes is induced through a regulatory gene. The genes encoding many of the degradation pathways are clustered on bacterial plasmids, indeed, the TOL pWWO plasmid of *Pseudomonas putida mt-2* was one of the first bacterial plasmids to be studied in detail.[7]

A general feature of bacterial degradation pathways is that they are convergent; very often, several aromatic metabolites can be converted into a common intermediate, that is then processed via a single pathway. The common bacterial aromatic degradation pathways are oxidative, and involve monooxygenase hydroxylation reactions and dioxygenase reactions.[1] A key intermediate in the breakdown of benzenoid compounds is catechol, which is formed from benzene via dioxygenase-catalyzed oxidation to the *cis*-dihydo-diol, followed by dehydrogenation of one alcohol group, and aromatization to catechol, as shown in **Figure 2**. Catechol can be formed from phenol, via monooxygenase-catalyzed hydroxylation, and can also be formed by other oxidative transformations, such as from salicylic acid, via the corresponding *cis*-diol.

Catechol is then a substrate for two types of oxidative ring cleavage. The ortho-cleavage pathway proceeds via intradiol dioxygenase cleavage of catechol to give *cis,cis*-muconic acid, which is then cyclized to form muconolactone, followed by isomerization and ring opening to form β-keto-adipate. This pathway, sometimes known as the β-keto-adipate pathway, is shown in **Figure 3**.
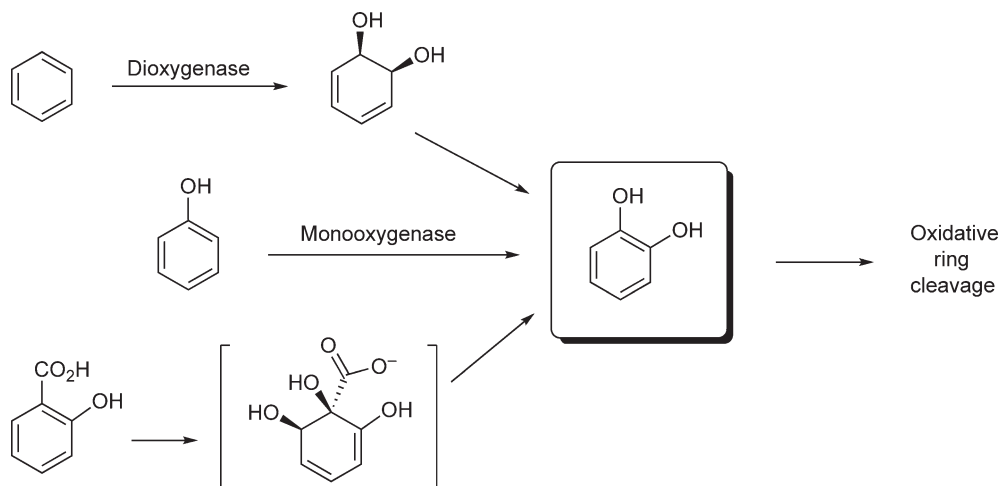

**Figure 2** Convergence of aromatic degradation upon catechol.
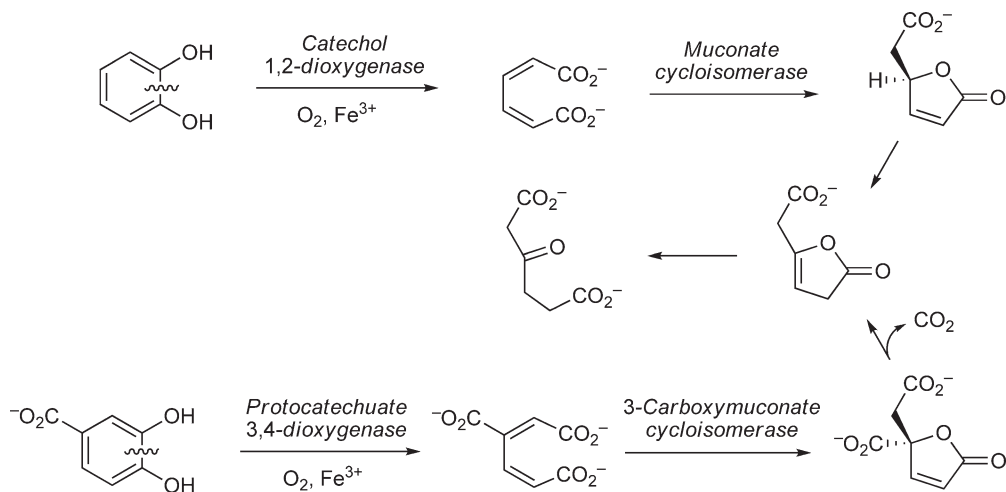
Figure 3   Ortho-cleavage pathways for catechol and protocatechuic acid.

The catechol meta-cleavage pathway involves extradiol dioxygenase cleavage of catechol to 2-hydroxymuconate semialdehyde, as shown in **Figure 4**. In *P. putida*, this intermediate is oxidized to 2-hydroxymuconic acid. Isomerization of the dienol to the corresponding enone, followed by decarboxylation, gives 2-hydroxypentadienoic acid, which is further degraded by hydratase-catalyzed addition of water to the 4,5-double bond, to give 2-keto-4-hydroxypentanoic acid, which is cleaved by an aldolase enzyme to give pyruvate and acetaldehyde. In some bacteria, 2-hydroxymuconate semialdehyde can be converted directly to 2-hydroxypentadienoic acid and formaldehyde by a C–C hydrolase enzyme (to be discussed in more detail is Section 8.16.1.8).
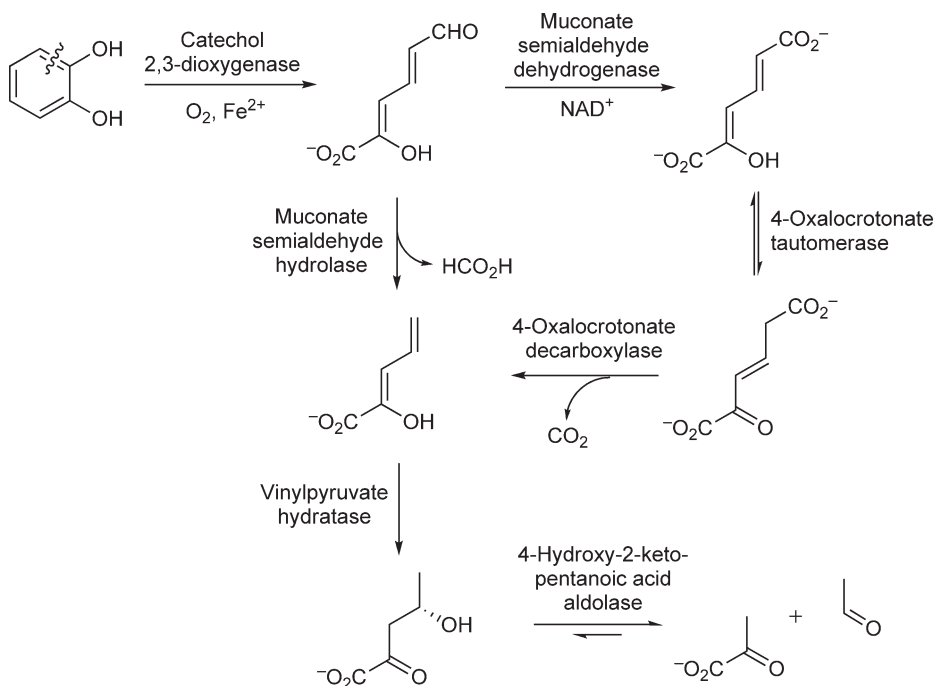
Figure 4   Meta-cleavage pathway.
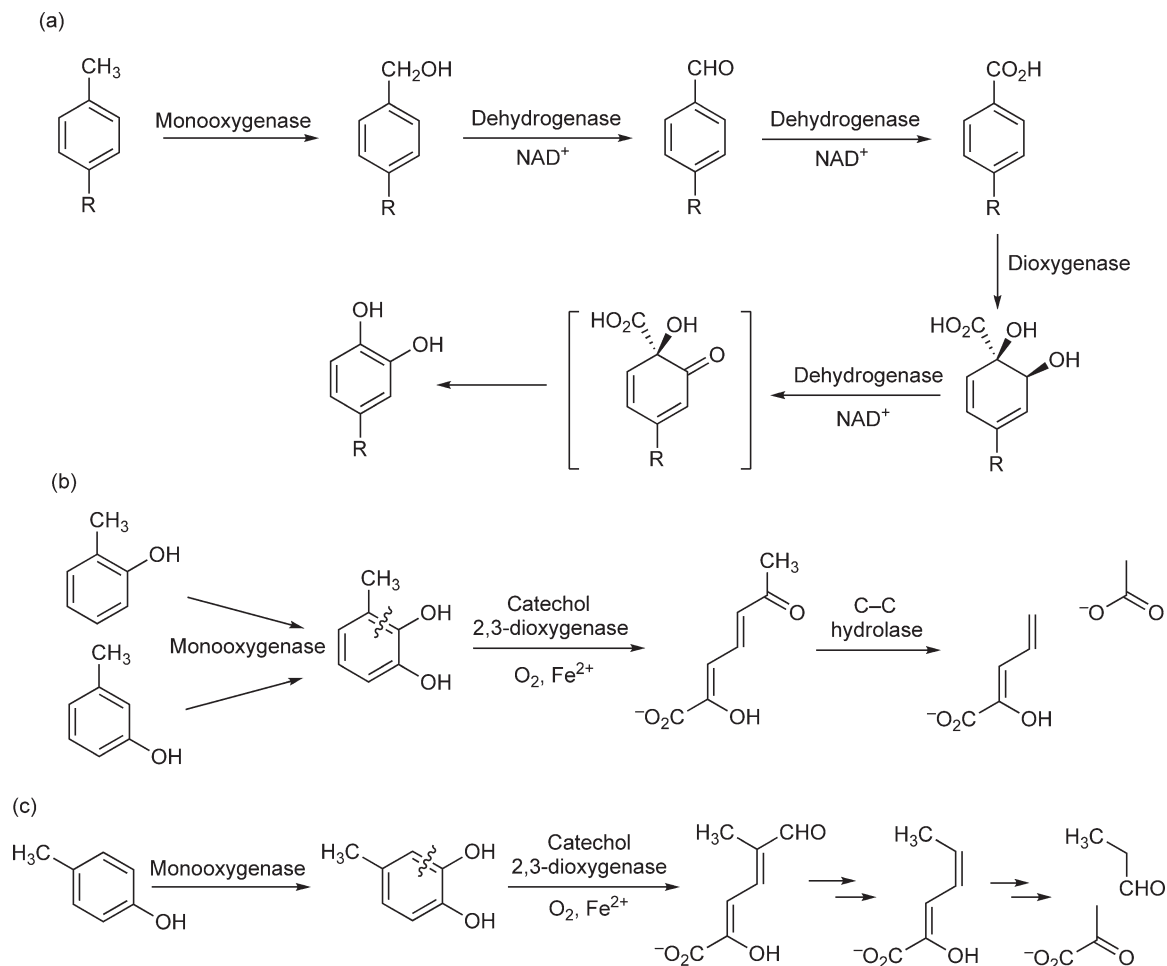
Figure 5 Degradation of alkyl-substituted aromatics. (a) Oxidation of side chain: toluene (R = H) and *p*-xylene (R = CH$_3$). (b) Removal after ring cleavage: *o*- and *m*-cresol. (c) Different end products: *p*-cresol.

Aromatic compounds containing alkyl side chains are metabolized using three general strategies, illustrated in **Figure 5**. The first involves oxidation of the alkyl side chain: toluene is oxidized to benzyl alcohol, and then further oxidized to benzaldehyde and benzoic acid. Benzoate dioxygenase then converts benzoate directly to catechol, via the corresponding *cis*-diol.

Chlorinated aromatic compounds pose several problems for degrading organisms, leading to the persistence in soil of man-made chlorinated aromatics such as DDT and PCBs.[8] The C–Cl bond is much harder to break than the C–H bond, and the reactions to remove chlorine often depend upon elimination of a nonaromatic intermediate. The chlorinated compounds may not induce the relevant catabolic enzymes, and the chlorinated intermediates are often less reactive. In some cases, reactive intermediates can be generated, for example, 3-chlorocatechol is an inhibitor for most catechol 2,3-dioxygenase ring cleavage enzymes. In a few cases, dehalogenase enzymes exist for hydrolytic removal of Cl$^-$ from certain compounds, as shown in **Figure 6**. Chlorobenzene is metabolized to 3-chlorocatechol, which is cleaved via ortho-cleavage to give chloro-muconic acid. After cyclization to the muconolactone, elimination of HCl is possible, as shown in **Figure 6**. 2-Chlorobenzoate can be oxidized directly to catechol by benzoate dioxygenase, involving an elimination of chloride from the *cis*-diol intermediate.

Polycyclic aromatic compounds are degraded via oxidation of the outer ring structure, followed by oxidative metabolism to give a single ring structure. As shown in **Figure 7**, biphenyl can be degraded via meta-cleavage of 2,3-dihydroxybiphenyl, followed by C–C cleavage to release benzoic acid. *Pseudomonas* sp. LB400, which operates this pathway, is able to metabolize lightly chlorinated biphenyls via this pathway.[4] Naphthalene is
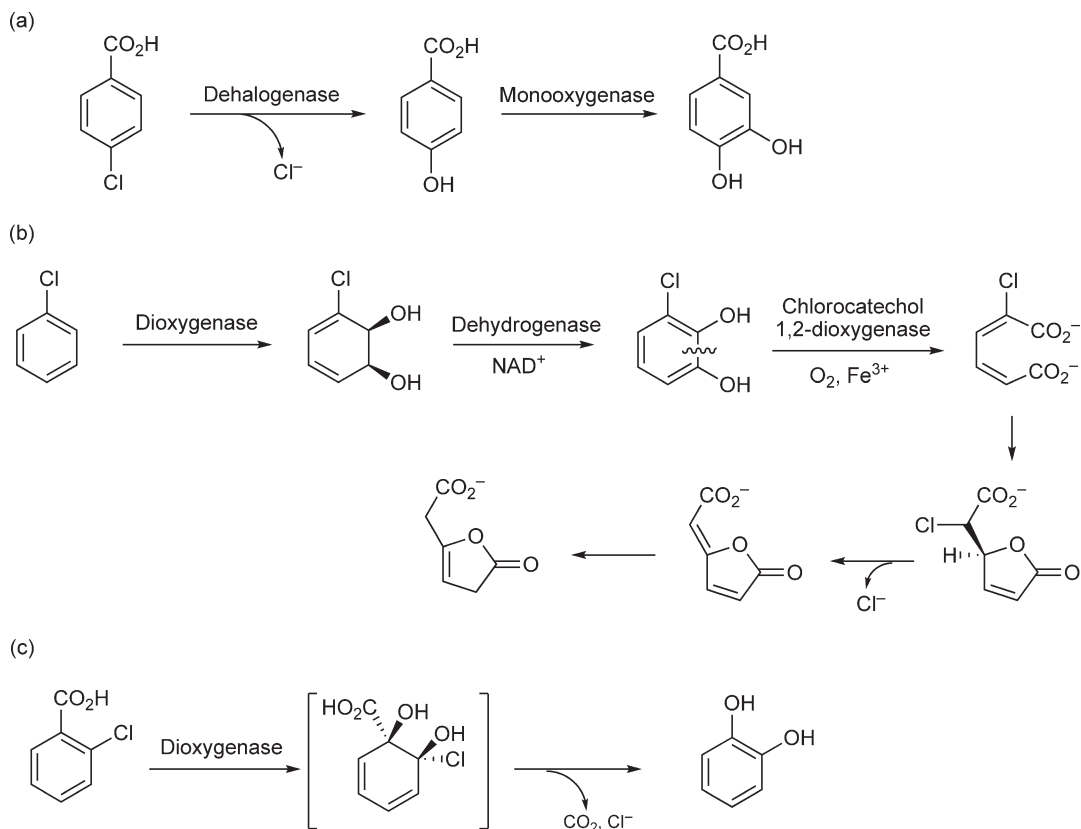
Figure 6   Degradation of chlorinated aromatics. (a) Hydrolytic cleavage: 4-chlorobenzoic acid. (b) Elimination of chloride after ring cleavage: chlorobenzene. (c) Elimination of chloride before ring cleavage: 2-chlorobenzoic acid.

degraded via meta-cleavage of naphthalene 1,2-diol, followed by hydration/aldolase cleavage to yield 2-hydroxy-benzaldehyde.

8.16.1.2   Reductive Pathways for Aromatic Degradation

Aromatic compounds can also be degraded anaerobically, via reductive pathways, which have been reviewed.[9] The best-characterized example is the degradation of benzoic acid in *Rhodopseudomonas palustris* and *Thauera aromatica* via the pathway shown in **Figure 8**. Benzoyl-CoA is formed, and then a reductase enzyme is able to reduce the aromatic ring to a cyclohexadiene. Following two consecutive additions of water, and oxidation to a β-keto ester, hydrolytic cleavage gives a linear 7-carbon CoA thioester, which can then be broken down via fatty acid β-oxidation.

Other reductive transformations are sometimes involved in nitroaromatic degradation pathways. *Pseudomonase pseudoalcalignes* JS45 is able to grow on nitrobenzene as sole carbon source.[10] Nitrobenzene is reduced, first to nitrosobenzene, and then to the hydroxylamine, which is isomerized via a mutase enzyme to 2-aminophenol, as shown in **Figure 8**. The remaining pathway then follows an oxidative meta-cleavage route.

8.16.1.3   Microbial Lignin Degradation

Much of the aromatic material present in soil is derived from lignin, the aromatic polymer that comprises 15–30% of lignocellulose in woody plants. Lignin is a heterogeneous polymer that is extremely resistant to biodegradation, but it is slowly degraded by white- and brown-rot fungi, and by some soil bacteria. The best-studied lignin degrader is the white-rot fungus *Phanerochaete chrysosporium*, which produces an extracellular
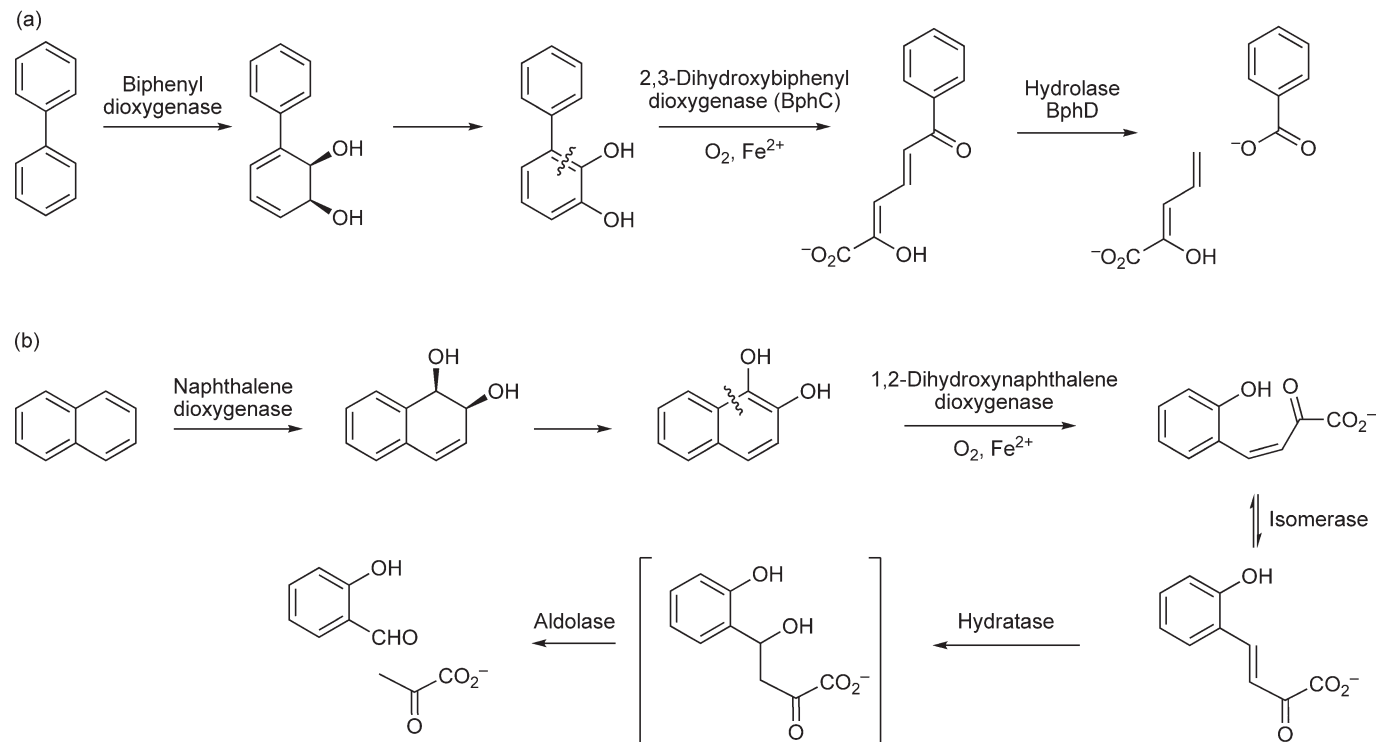
Figure 7 Degradation of biphenyl and naphthalene. (a) Biphenyl degradation pathway. (b) Naphthalene degradation pathway.
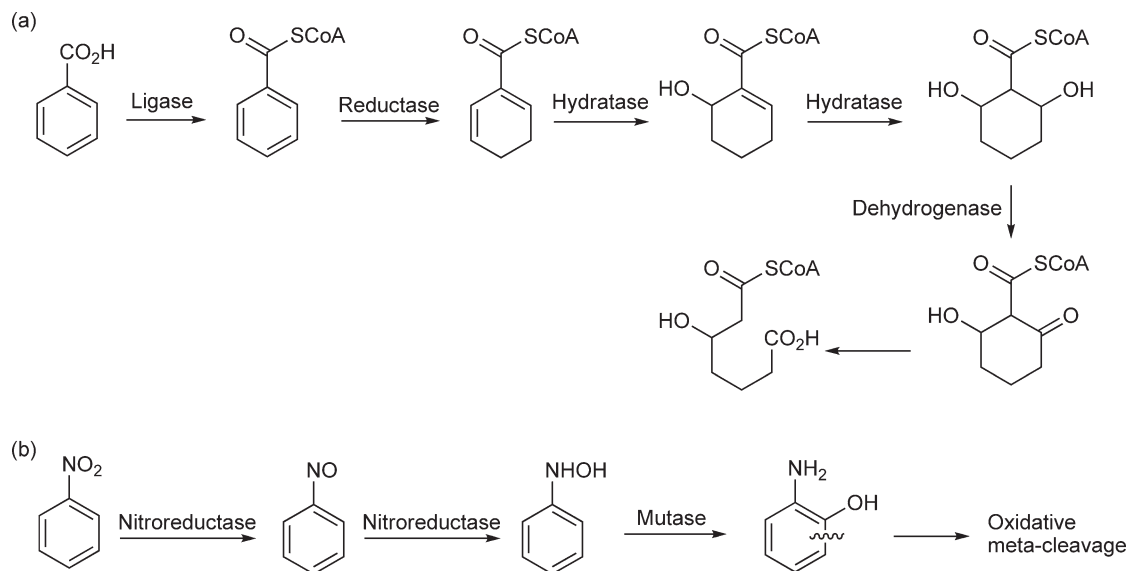
Figure 8 Reductive degradation pathways. (a) Anaerobic degradation of benzoate. (b) Reductive degradation of nitrobenzene by *Pseudomonas pseudoalcaligenes* JS45.

peroxidase enzyme that breaks down lignin into smaller dimeric fragments, via a radical depolymerization mechanism.[11] The major breakdown products from lignin are shown in **Figure 9**. Actinomycetes such as *Streptomyces viridosporus* are also able to break down lignin via extracellular lignin peroxidases.[12]

Several soil bacteria are known to break down the lignin fragments, via oxidative cleavage pathways. The most abundant fragment is a β-aryl ether, which can be broken down in *Sphingomonas* via a pathway involving a glutathione-dependent β-etherase activity, as shown in **Figure 10**.[13] The aromatic product vanillin is then converted to protocatechuic acid (3,4-dihydroxybenzoic acid), which is a substrate for oxidative meta-cleavage by protocatechuate 4,5-dioxygenase (4,5-PCD).[13]

The diarylpropane lignin fragment is converted in *Pseudomonas* to lignostilbene via an unusual fragmentation reaction, generating formaldehyde.[14] Lignostilbene is then cleaved oxidatively by a nonheme iron-dependent lignostilbene dioxygenase,[15] a member of the carotenoid cleavage dioxygenase (CCD) family.

8.16.1.4 Arene (Rieske) Dioxygenases

The initial step of most aromatic degradation pathways is the oxidation of an arene to the corresponding *cis*-dihydro diol. This remarkable reaction is catalyzed by a family of nonheme iron-dependent multicomponent dioxygenases, which have been reviewed.[16,17] The best-characterized example of this family is naphthalene dioxygenase (**Figure 11**).

Naphthalene dioxygenase consists of three components, which form an electron transfer chain: an NADH-dependent flavoprotein reductase,[18] a ferredoxin containing two [2Fe2S] Rieske iron–sulfur clusters,[19] and a Rieske oxygenase containing both a [2Fe2S] Rieske iron–sulfur cluster and a mononuclear iron(II) center in the enzyme active site.[20,21]

The crystal structure of the terminal dioxygenase component of naphthalene dioxygenase was solved by Kauppi *et al.* in 1998, revealing a mononuclear iron(II) center in the active site, coordinated by His-208, His-213, and a bidentate Asp-362 (**Figure 12**).[22] The structure revealed that the mononuclear iron(II) center was positioned within 12 Å of the [2Fe2S] cluster of another subunit in the $\alpha_3\beta_3$ oxygenase domain.[22] Refinement of the structure revealed electron density for an indole hydroperoxide, ligated to the iron(II) center, with the indole ring positioned at about 4 Å from the iron(II) center.[23] The presence of indole was believed to arise from the presence of L-tryptophan in the growth media, and the observation of a hydroperoxide species suggested that an alkyl hydroperoxide intermediate might be formed in the catalytic mechanism.[23] Further recent crystallographic studies have yielded the structures of ternary complexes with
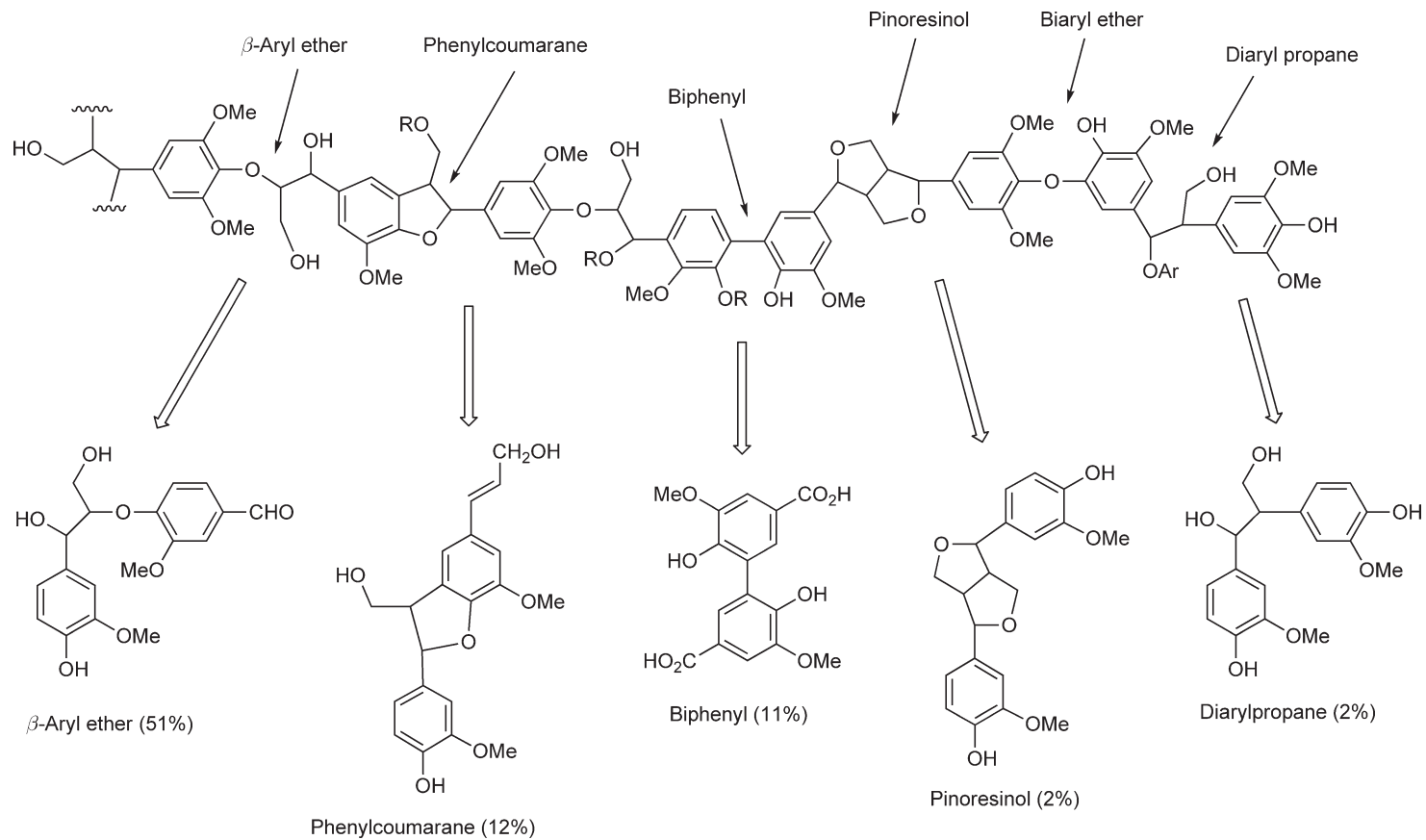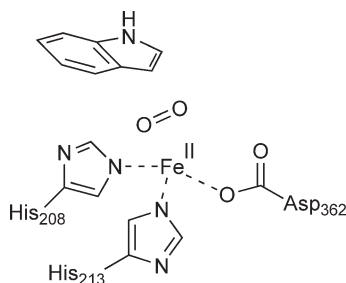
Figure 9 Lignin breakdown products.
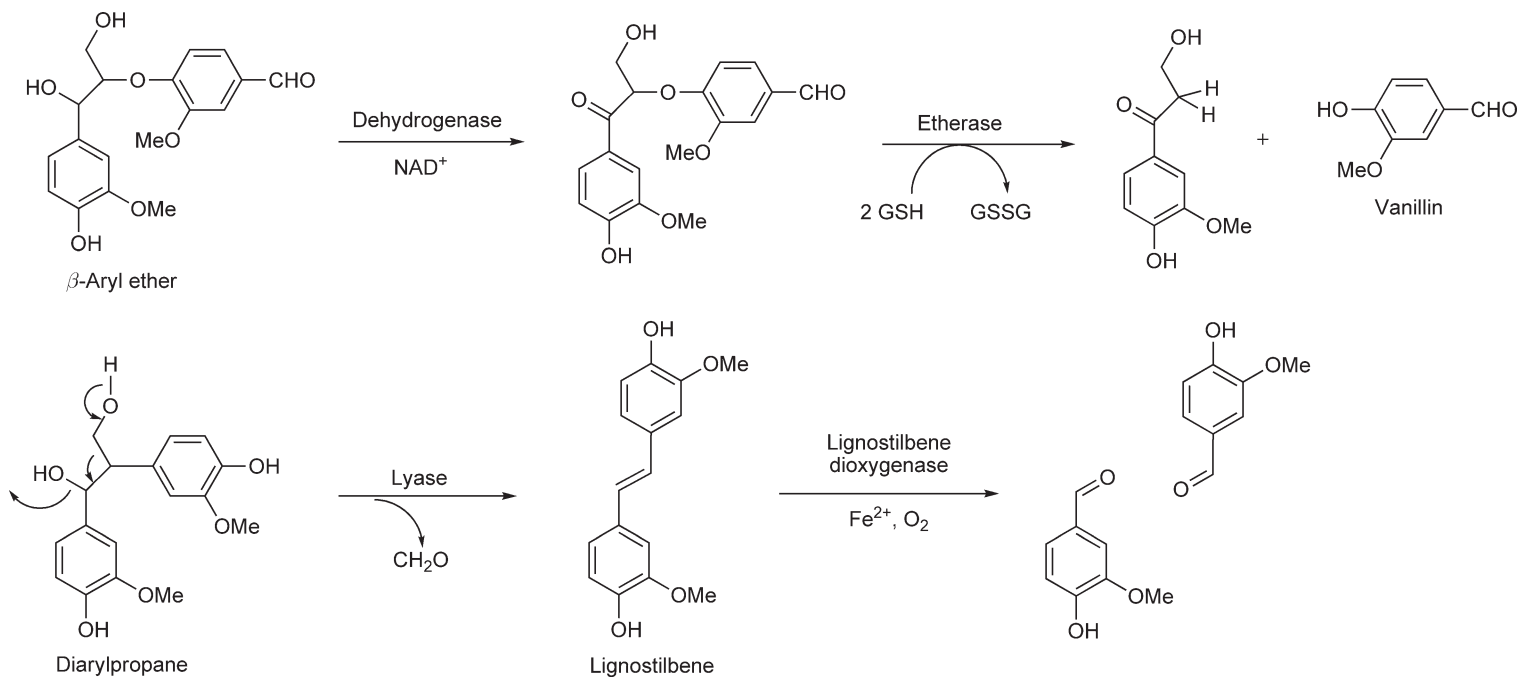
Figure 10 Breakdown of lignin fragments.
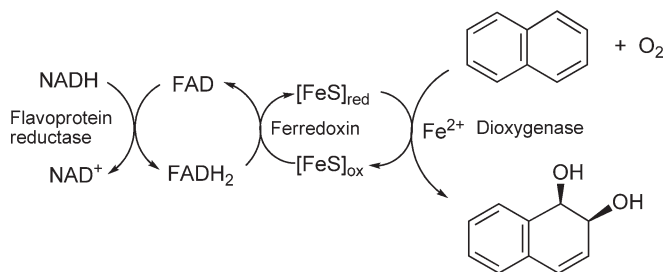
**Figure 11** Naphthalene dioxygenase reaction.

Figure 12 Active site of naphthalene 1,2-dioxygenase with bound dioxygen.

substrates (naphthalene or indole) and dioxygen.[24] Remarkably, dioxygen was found to be bound side-on to the iron(II) center, with Fe–O distances of 1.8 and 2.0 Å, and an O–O distance of 1.4 Å. The aryl substrate was positioned slightly further from the iron(II) center, but close to the bound oxygen (see **Figure 29**).[24] The electron transfer pathway from the [2Fe2S] cluster to the active site iron(II) center passes through Asp-205, whose replacement results in loss of enzyme activity.[25]

Despite the detailed structural data for this enzyme, there are relatively few insights into the catalytic mechanism for this remarkable transformation. The original mechanistic proposal for dihydroxylation, involving a dioxetane intermediate,[26] now seems unlikely, given the presence of iron–sulfur clusters in the enzyme, suggesting one-electron transfers in the catalytic mechanism. Toluene dioxygenase and naphthalene dioxygenase both possess monooxygenase activity using alternate substrates, suggesting that dihydroxylation is stepwise, and not concerted.[27,28] The existence of monooxygenase activity suggests the possible involvement of iron–oxo intermediates in catalysis. It has also been observed that processing of benzene by naphthalene dioxygenase leads to the production of hydrogen peroxide, via uncoupling of oxygen activation from substrate hydroxylation.[29] Furthermore, fully reduced benzoate 1,2-dioxygenase is able to utilize hydrogen peroxide to form *cis*-diol product.[30] These observations suggest that dioxygen is activated via superoxide, which can be further reduced to an iron(III) hydroperoxy intermediate, which might be the active oxidant or which might undergo O–O bond cleavage to form a $Fe^V(=O)$–OH species.

Single turnover studies of the naphthalene dioxygenase reaction have shown that the catalytic cycle commences with active site nonheme cofactor as iron(III), which is reduced to iron(II) by electron transfer from a Rieske [2Fe2S] cluster, and a further one-electron transfer occurs during the catalytic cycle.[31] Wolfe *et al.* have suggested that O–O bond cleavage could occur first, to give an $O=Fe(V)$–OH intermediate, which could effect dihydroxylation in a similar fashion to the dihydroxylation of alkenes by $NaIO_4$ or OsO_4.[31] This proposal is consistent with model studies using iron(III) complexes, which are able to oxidize alkene substrates using hydrogen peroxide as oxidant, to form a mixture of *cis*-diol and epoxide products.[32–34] However, computational studies appear to disfavor the formation of an iron(V)–oxo species.[35] Radical trap studies using naphthalene dioxygenase have shown that ring opening of norcarene occurs, to give 60–70% of a ring-opened product, consistent with a substrate radical intermediate.[36] Therefore, the weight of evidence favors a radical mechanism proceeding via an iron(III) hydroperoxy intermediate, which could either react directly with the aryl substrate via a radical mechanism (mechanism A, **Figure 13**), or undergo O–O bond homolysis to form a reactive $Fe^V(=O)$–OH species (mechanism B, **Figure 13**).
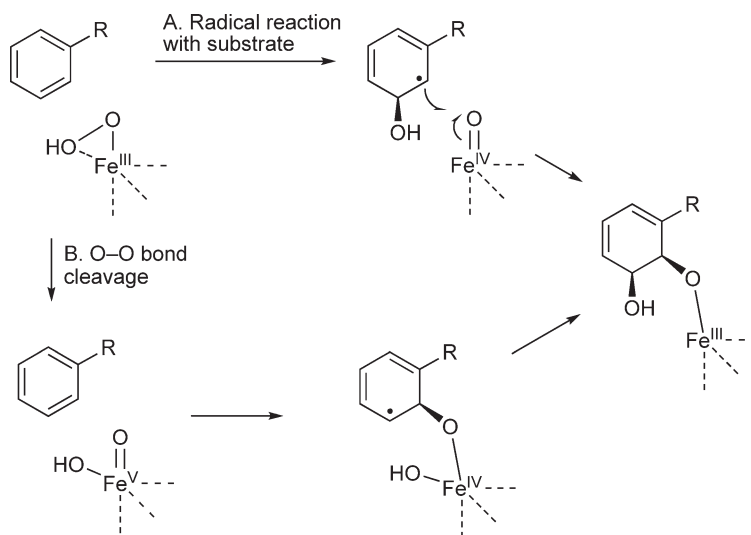
Figure 13 Possible catalytic mechanisms for *cis*-dihydroxylation.

## 8.16.1.5 Intradiol Catechol Dioxygenases

The catechol dioxygenases were discovered through the work of Osamu Hayaishi, who established that the dioxygenases responsible for intradiol catechol cleavage required iron(III), whereas the enzymes responsible for extradiol cleavage utilized iron(II).[37] Hayaishi was able to demonstrate in 1955, using $^{18}O_2$ labeling experiments, that catechol 1,2-dioxygenase (1,2-CTD) from *Pseudomonas* incorporated two atoms of oxygen from dioxygen into the reaction products,[38] consistent with a mechanism involving a four-membered dioxetane intermediate (see **Figure 14**).

The first X-ray structure of a catechol dioxygenase, the intradiol-cleaving protocatechuate 3,4-dioxygenase (3,4-PCD) from *P. putida*, was solved by Ohlendorf *et al.* in 1988.[39] The enzyme consists of an oligomeric $(\alpha\beta Fe)_{12}$ structure. The nonheme iron(III) cofactor is ligated by four amino acid side chains: the imidazole side chains of His-460 and His-462, and the phenolic side chains of Tyr-408 and Tyr-447 (see **Figure 15(a)**). A fifth water ligand completes a trigonal bipyramidal structure. The two tyrosinate ligands give the enzyme its characteristic deep red color due to ligand-to-metal charge transfer interactions, which give rise to characteristic resonance Raman vibrations at 1254 and 1266 cm^{-1}.[40]

The structure of 1,2-CTD from *Acinetobacter* sp. ADP1, which is sequence related to 3,4-PCD, consists of an α_2 homodimer with one iron(III) cofactor per subunit.[41] The tertiary structure of the 1,2-CTD enzyme is similar to that found in 3,4-PCD, although 1,2-CTD contains a novel helical zipper motif at the interface of the two subunits, with two molecules of bound phospholipid. The active site of 1,2-CTD contains a very similar
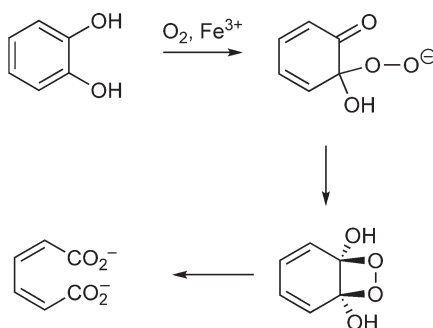
Figure 14 Dioxetane mechanism.
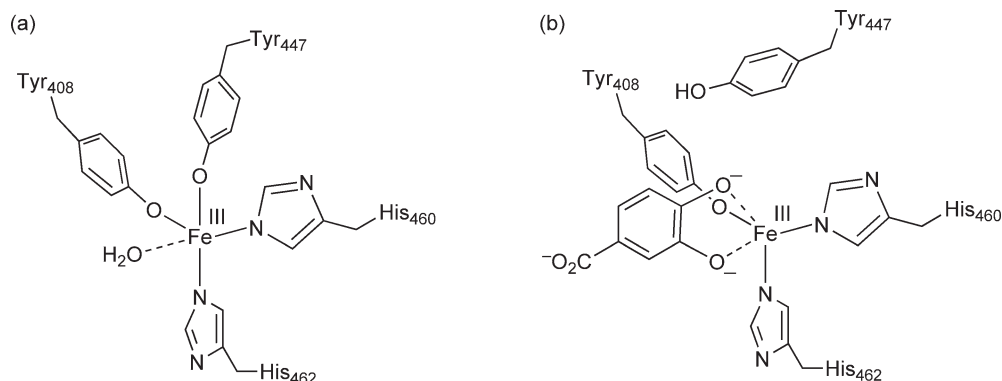
Figure 15   Active site structures of 3,4-PCD, without substrate (a) and with substrate (b).

arrangement of iron(III) ligands: Tyr-200 and His-226 are the axial ligands, and Tyr-164, His-224 and a water molecules are the equatorial ligands.[41]

Structures of 3,4-PCD with bound catechol substrates reveal that, upon substrate binding, the axial tyrosine ligand Tyr-447 and equatorial water ligand are both displaced, to form a monodentate substrate complex.[42,43] Tyr-447 swings away from the iron(III) cofactor to leave an iron(III) center with approximately octahedral geometry, but containing a vacant coordination site opposite His-460. Subsequent structures of 3,4-PCD with substrates and inhibitors in the presence of NO and CN^- have revealed that either NO or CN^- can bind in the vacant coordination site, to form a ternary complex.[44,45] The structures of 1,2-CTD with bound catechol have shown, similarly, that the axial Tyr-200 ligand swings away from the iron(III) center upon substrate binding.

Reaction of the iron(III)–catechol complex with dioxygen has been proposed to occur via a substrate activation mechanism, in which electron transfer occurs from the iron(III) cofactor to substrate to form an iron(II)–semiquinone, which is able to react directly with dioxygen to form a hydroperoxide intermediate. Evidence in support of this mechanism comes from the study of model complexes such as iron(III) TPA, in which Que and coworkers found that the most reactive model complexes showed the most iron(II)–semiquinone character.[46,47] The literature on catechol dioxygenase model chemistry has been reviewed.[16,48] Studies of the 3,4-PCD reaction using MCD spectroscopy and electronic structure calculations by Solomon and coworkers have indicated that a highly covalent iron(III)–semiquinone complex interacts with dioxygen through a strong π interaction, with simultaneous interaction with the iron center, to form a five-coordinate dioxygen adduct.[49] By replacement of equatorial ligand Tyr-408 with NH_3, it was found that Tyr-408 is essential for stabilizing the iron(III) center for reaction with dioxygen.[49]

Subsequent reaction of the cyclohexadienyl hydroperoxide intermediate is believed to occur via a Criegee rearrangement, with migration of the adjacent acyl group (acyl migration) to yield muconic anhydride as an intermediate, which then undergoes hydrolysis to give the product muconic acid. $^{18}O_2$ labeling studies on (1,2-CTD) from *Pseudomonas arvilla* have revealed that the intradiol cleavage products contain 99% incorporation of a single atom of ^{18}O, and 74% incorporation of a second atom of ^{18}O, with 24% incorporation of only one atom of ^{18}O.[50] These data are not consistent with a dioxetane intermediate, but are consistent with a Criegee rearrangement to give an anhydride intermediate, followed by the partial exchange of the iron(III) ^{18}O-hydroxide with solvent water.[50]

The role of the axial Tyr-447 ligand that swings away from the iron(III) center during catalysis has been probed by construction of a Y447H mutant 3,4-PCD enzyme.[45] The mutant enzyme has a 600-fold lower k_{cat} than the native enzyme, but kinetic studies have shown that reaction with dioxygen occurs at similar rates to the native enzyme, therefore the lower k_{cat} is due to slower substrate binding and product release.[45] The role of Tyr-447 is thought to be as a base, deprotonating the second phenolic hydroxyl group of the substrate to form a catechol dianion. The proposed mechanism is shown in **Figure 16**.
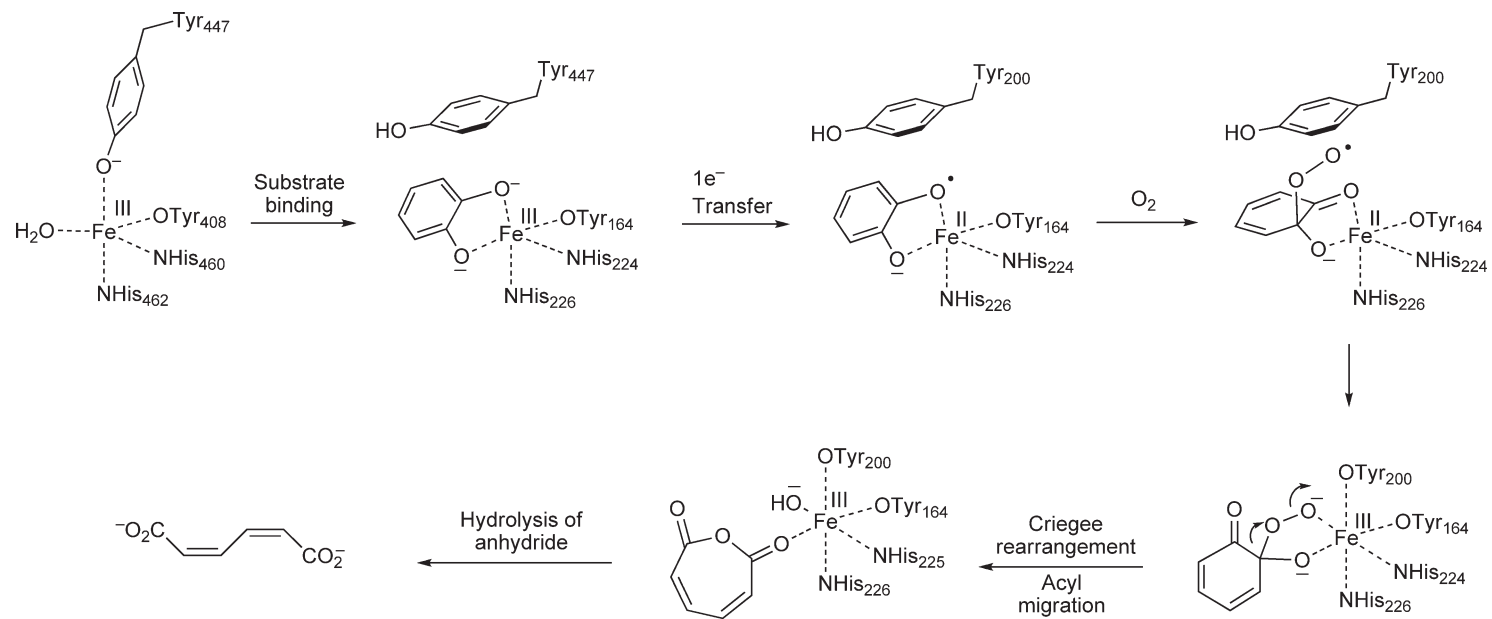
Figure 16 Intradiol mechanism.

8.16.1.6   Enzymology of Ortho-Cleavage Pathways

The ortho-cleavage pathways for catechol and protocatechuic acid are illustrated in **Figure 3**. The intradiol oxidative cleavage product, *cis,cis*-muconic acid, is cyclized by muconate cycloisomerase to form muconolactone (**Figure 17**). The stereochemical course of *P. putida* muconate cycloisomerase has been shown to be a *syn*-addition.[51] The corresponding enzyme on the protocatechuate pathway, 3-carboxymuconate cycloisomerase, has been shown to proceed with antistereochemistry in *P. putida*,[52] whereas the enzyme in *Neurospora crassa* catalyzes cyclization of the same substrate onto the 4-position, with *syn*-stereochemistry.[53] Decarboxylation of the 3-carboxymuconolactone in *P. putida* gives an unsaturated lactone, which is isomerized by a Δ-isomerase enzyme, which has been shown to proceed via a 1,3-suprafacial shift.[54]

The X-ray crystal structure of *P. putida* muconate lactonizing enzyme (cycloisomerase) was determined in 1987, and was found to contain an α/β barrel fold, also found in triosephosphate isomerase and enolase.[55] Remarkably, the structure of *P. putida* mandelate racemase, which catalyzes a mechanistically distinct reaction earlier in the same pathway, was found in 1990 to have a homologous structure, indicating that the structural fold of the enolase superfamily is able to support a range of enzyme-catalyzed reactions.[56] The *P. putida* 3-carboxy-*cis,cis*-muconate lactonizing enzyme, in contrast, shares sequence similarity with a class II fumarase enzyme,[57] and determination of its structure in 2004 has shown that it shares the same fold as the class II fumarase superfamily, hence these two catalysts of similar reactions have evolved from different ancestors.[58]

8.16.1.7   Extradiol Catechol Dioxygenases

The extradiol catechol dioxygenases catalyze the oxidative cleavage of catechol substrates, through cleavage of the bond adjacent to the two hydroxyl groups, to give a 2-hydroxymuconaldehyde product, using iron(II) as a cofactor (see **Figure 4**). It was shown by Hayaishi that catechol 2,3-dioxygenase from *Pseudomonas arvilla* incorporated two atoms of ^{18}O from $^{18}O_2$ into the product,[59] and thus a mechanism involving a dioxetane intermediate was also possible for this family of enzymes.

The structure of 2,3-dihydroxybiphenyl 1,2-dioxygenase (BphC) from *Pseudomonas* LB400, a strain capable of degrading chlorinated biphenyls, was solved by Han *et al.* in 1996.[60] The tertiary structure of the enzyme consists of two similar $\beta\alpha\beta\beta$ domains, only one of which contains an iron(II) cofactor. A funnel-shaped cavity leads to the active site, where the iron(II) center is ligated by three amino acid side chains: His-146, His-210, and Glu-260 (see **Figure 18**).[60] This His_2Glu/Asp motif is found in a number of other nonheme iron(II)-dependent oxygenases, including the α-ketoglutarate-dependent dioxygenases and isopenicillin N synthase.[61] The crystal structure of BphC from *Pseudomonas* KKS102 with bound substrate shows a similar coordination geometry, however the crystals contain iron(III) rather than iron(II).[62]

In 1998, Kita *et al.* reported the structure of catechol 2,3-dioxygenase from *P. putida mt-2*, an α_4 tetramer.[63] The subunit structure is very similar to that of BphC, and the iron(II) cofactor is bound by His-153, His-214, and Glu-265.[63] In 1999, Sugimoto *et al.* reported the structure of 4,5-PCD from *Sphingomonas paucimobilis* SYK-6,
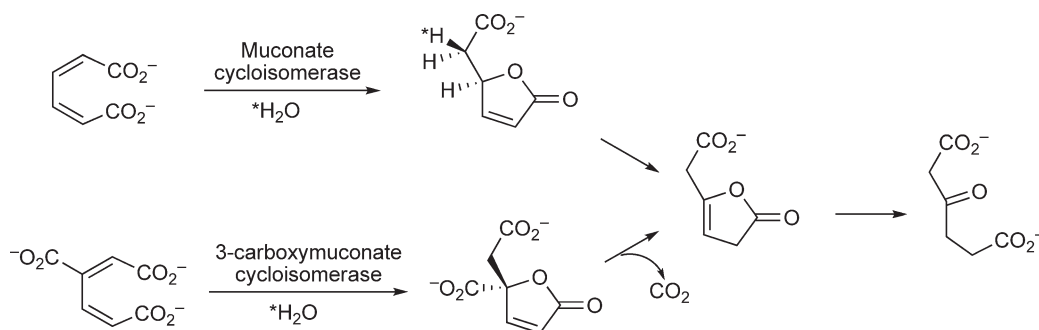
Figure 17   Stereochemical course of ortho-cleavage pathway enzymes.
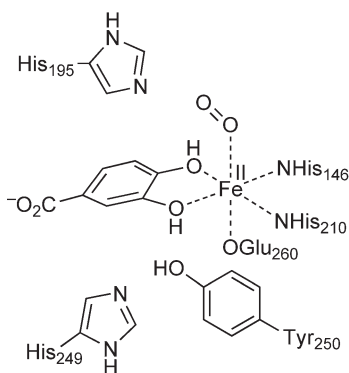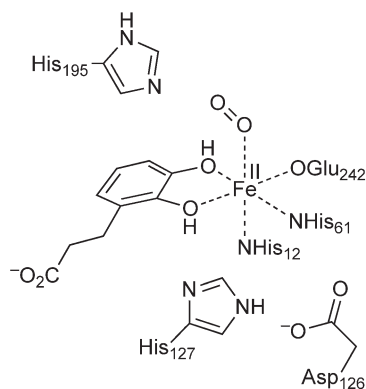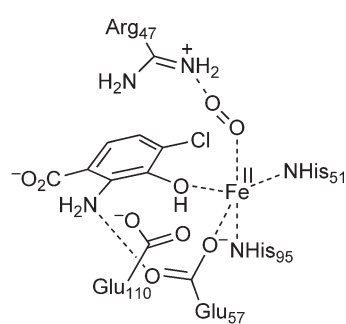
Figure 18   Active sites of extradiol dioxygenases: (a) BphC, (b) LigB, and (c) HAD.

which is composed of an $\alpha_2\beta_2$ tetramer.[64] This enzyme has no sequence similarity to BphC, yet the active site iron(II) ligands are very similar: the metal center is coordinated by His-12, His-61, and Glu-242 (see **Figure 18**).[64] The structure of an aminophenol cleavage dioxygenase, 3-hydroxyanthranilate 3,4-dioxygenase from *Ralstonia metallidurans*, was reported in 2005 by Zhang *et al.*[65] The structure of a complex with a 4-chloro-substituted inhibitor was solved, in the presence of bound NO, providing insight into enzyme–substrate binding interactions. This enzyme also contains a His,His,Glu motif for iron(II) cofactor binding, however the identity of outer sphere active site residues is different to the catechol dioxygenases. Arg-47 is positioned close to bound NO, and the only active site acid–base residue is Glu-110.[65]

There is clearly more than one family of extradiol catechol dioxygenases, since some enzymes closely resemble BphC in sequence and tertiary structure, whereas 4,5-PCD is clearly unrelated. Spence *et al.* proposed in 1996 a classification of three classes of extradiol dioxygenase, based upon sequence alignments: class I enzymes containing a single domain, of Mr 21–24 kDa; class II enzymes including BphC containing two domains, of which the C-terminal domain binds Fe^{2+}; and class III enzymes including 4,5-PCD and 2,3-dihydroxyphenylpropionate 1,2-dioxygenase (MhpB), with active site residues in the N-terminal half of the protein sequence.[66] More recently, a new classification based upon structural fold has been proposed by Vaillancourt *et al.*[67]

The catalytic mechanism of the extradiol catechol dioxygenases proceeds through several enzyme-bound intermediates. EPR spectroscopic studies by Lipscomb and coworkers of the NO complex of the extradiol dioxygenase 4,5-PCD demonstrated that the iron(II) cofactor binds both catecholic hydroxyl groups and NO.[68] The catechol substrate has been shown in the case of dioxygenase BphC, using UV/visible and resonance Raman spectroscopy, to be bound as the catecholate monoanion.[69] The iron(II) cofactor is thought to activate dioxygen as superoxide, and activate the catechol substrate as its semiquinone radical, as has been observed in model transition metal complexes.[70] Evidence for a semiquinone radical intermediate has been obtained in the reaction catalyzed by MhpB from *E. coli*, using substrate analogues containing cyclopropyl radical traps.[71] Processing of *trans*- and *cis*-substituted cyclopropyl analogues by MhpB was found to proceed with isomerization of the cyclopropyl ring substituents. The most plausible explanation of these results is that a radical-mediated reversible opening of the cyclopropyl ring is taking place, upon formation of a transient iron(II)–semiquinone–superoxide intermediate.[71]

Thus, it is believed that the iron(II) cofactor mediates one-electron transfer with dioxygen and with the bound catechol substrate, leading to C–O bond formation between the two activated substrates, to give a hydroperoxide intermediate. Evidence in favor of the proximal hydroperoxide has been obtained for MhpB, from enzyme inhibition by 'carba' analogues of the proximal hydroperoxide, in which the –OOH functional group was replaced by –CH$_2$OH, and the cyclohexadienyl ring simplified to a cyclohexanone ring.[72] Enzyme inhibition was only observed in analogues in which the hydroxymethyl substituent is positioned in an axial orientation with respect to the cyclohexanone ring, indicating that the conformation adopted by the

hydroperoxide was of importance. The proof of the existence of a proximal hydroperoxide intermediate has come from its direct observation in the crystal structure of homoprotocatechuate 2,3-dioxygenase (2,3-HPCD) from *Brevibacterium fuscum*, in the presence of slow substrate 4-nitrocatechol.[73] Three different intermediates were observed in the tetrameric crystal: the ternary complex of bound substrate and dioxygen, the proximal hydroperoxide intermediate formed upon C–O bond formation, and the cleaved extradiol product. The hydroperoxide group is positioned axially with respect to the cyclohexadiene ring of the substrate, in agreement with earlier stereoelectronic considerations of the catalytic mechanism.[74]

The proximal hydroperoxide is then believed to undergo Criegee rearrangement to give a seven-membered lactone intermediate, which is hydrolyzed to give the extradiol product. $^{18}O_2$ labeling studies carried out on *E. coli* MhpB revealed that, although both the acid and ketone carbonyls could be labeled with ^{18}O from $^{18}O_2$, upon reaction in $H_2^{18}O$ the carboxylate position was labeled to the extent of 30%, consistent with the formation of an α-ketolactone intermediate, and exchange of iron(II) hydroxide with solvent ^{18}O-labeled water.[75] The enzyme was also found to catalyze the hydrolysis of a saturated seven-membered lactone analogue.[75] These studies implicate a lactone intermediate arising from Criegee rearrangement in the extradiol cleavage reaction mechanism, shown in **Figure 19**.

Both intradiol and extradiol catechol dioxygenase mechanisms proceed through the same proximal hydroperoxide intermediate, but then proceed via different Criegee rearrangements: intradiol cleavage occurs via 1,2-acyl migration to give an anhydride, whereas extradiol cleavage occurs via 1,2-alkenyl migration to give a lactone. Recent studies have shown that extradiol specificity can be altered to intradiol cleavage by point mutations.[76,77] In *B. fuscum* 2,3-HPCD, mutant H200F converts unnatural substrate 2,3-dihydroxybenzoic acid to the intradiol product.[76] In *E. coli* MhpB, point mutants generated by directed evolution studies, were found to generate 5–15% intradiol product.[77] These observations support the existence of a common reaction intermediate for extradiol and intradiol cleavage, and imply that iron(II) is able to support extradiol and intradiol cleavage.

It has been proposed that acid–base catalysis is a key factor in extradiol cleavage.[78] Studies of a biomimetic model reaction for extradiol cleavage using $FeCl_2$, 1,4,7-triazanonane, and pyridine, have shown that pyridine has two roles in the reaction: pyridine initially generates the catecholate monoanion and the pyridinium cation then acts as a proton donor to assist the Criegee rearrangement.[78] These observations are consistent with the presence of two acid–base groups in the outer sphere of the extradiol catechol dioxygenases. Site-directed mutagenesis of His-179 and His-115 in *E. coli* MhpB has shown that both residues are essential for catalysis.[79] Replacement of adjacent active site residues indicates that His-179 is the active site base, whereas His-115, possessing an abnormally high pK_a of 8.0 due to adjacent Asp-114, acts as a proton donor for the final lactone hydrolysis step.[79] In the BphC family of extradiol dioxygenases, a crystal structure of *Pseudomonas* KKS102 BphC complexed
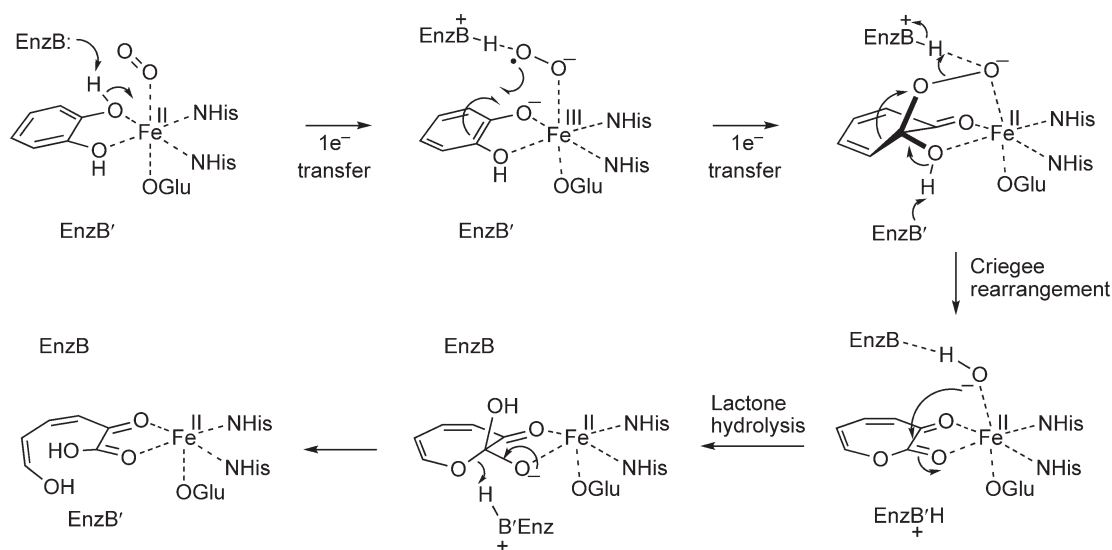
Figure 19 Catalytic mechanism for extradiol cleavage.

with substrate and NO showed that His-194 was positioned close to NO, which occupies a sixth coordination site at the iron(II) center.[80] His-194 was therefore proposed to act as a base to deprotonate the substrate at C-3, to form the catechol monoanion, and then the protonated imidazolium side chain of His-194 could stabilize the negative charge of bound superoxide, hence assisting oxygen activation. However, replacement of His-200 in *B. fuscum* 2,3-HPCD by Ala, Gln, or Asn gives mutant enzymes that retain 30–40% k_{cat}, compared to wild-type enzyme.[81] Thus, different results are obtained upon replacement of active site histidine residues in MhpB and 2,3-HPCD, however these are not sequence related, therefore it appears that the precise roles of active site acid–base residues in the two families of extradiol dioxygenases are somewhat different. Nevertheless, acid–base catalysis via outer sphere active site residues appears to be an important factor in extradiol cleavage catalysis. Model studies also indicate that the facial tridentate ligand geometry for iron(II) is an important factor in extradiol selectivity.[78,81–83]

Although iron(II) is normally utilized as the metal ion cofactor in these enzymes, dioxygenase MndD from *Arthrobacter globiformis* contains manganese(II) at its active site.[84] A similar catalytic mechanism can be envisaged for Mn^{2+}, although the redox potential for the Mn^{3+}/Mn^{2+} redox couple is significantly higher, at +1.6 V. Gentisate dioxygenase, which catalyzes the oxidative cleavage of a *p*-hydroquinone substrate, requires iron(II), which has been shown to ligate the C-2 hydroxyl group of the substrate, and appears to follow a similar mechanism.[85]

8.16.1.8   Enzymology of Meta-Cleavage Pathways

Following extradiol oxidative cleavage, there are two strategies for degradation of the meta-ring fission product, as illustrated in **Figure 4**. In *P. putida*, 2-hydroxymuconalde semialdehyde is oxidized by an NAD^+-dependent dehydrogenase to give 2-hydroxymuconic acid, which is tautomerized by 4-oxalocrotonate tautomerase.[86] The ketonization catalyzed by this tautomerase is stereospecific, as shown in **Figure 20**.[87] The structure of this 62-amino acid enzyme has been solved by X-ray crystallography,[88] and also by ^1H nuclear magnetic resonance (NMR) spectroscopy.[89] The catalytic base has been identified as the N-terminal proline residue.[89] Two active site arginine residues, Arg-11 and Arg-39, have been shown by site-directed mutagenesis to be important in substrate binding and catalysis,[90] and replacement by citrulline has shown that Arg-39 stabilizes the anionic transition state by electrostatic stabilization.[91] The ketonized product is then a substrate for 4-oxalocrotonate decarboxylase, to give 2-hydroxypentadienoic acid.[92]

The other, more common strategy is a hydrolytic cleavage of the meta-ring fission product, catalyzed by a family of C–C hydrolase enzymes (**Figure 21**). There are X-ray crystal structures of *E. coli* MhpC,[93] *Burkholderia xenovorans* BphD,[94] *Pseudomonas fluorescens* CumD,[95] and *Pseudomonas resinovorans* CarC.[96] The structural fold of each enzyme is that of the α/β-hydrolase fold, containing eight parallel β-sheets, with catalytic Ser, His, and Asp residues situated on loops protruding into the active site, and a four-helix 'lid' positioned over the active site.

A series of mechanistic studies have been carried out on hydrolase MhpC, found on the *E. coli* phenylpropionic acid catabolic pathway. The overall stereochemistry of the enzyme-catalyzed reaction proceeds with insertion of the H-5$_E$ hydrogen by the enzyme, with partial exchange of the H-5$_Z$ hydrogen.[97] Stopped-flow kinetic studies at pH 5.0 and 4.0 have identified a fast initial step, corresponding to the ketonization of the substrate dienol (ring fission product (RFP)) into a keto tautomer (ketonised ring fission product (RFPk)), followed by rate-determining C–C cleavage and product release steps.[98] The existence of a covalent acyl enzyme intermediate, implied by the presence of a serine catalytic triad, was investigated using a range of kinetic and trapping experiments; however, very low stoichiometries (<1%) of covalently bound acyl species were found under conditions where acyl enzyme hydrolysis should be slow.[99] Evidence in favor of a mechanism involving base-catalyzed attack of water has been obtained from the incorporation of 4–6% of two atoms of ^{18}O from $H_2^{18}O$ into the product succinate, and from the enzyme-catalyzed exchange of ^{18}O into a noncleavable analogue 4-keto-nona-1,9-dioic acid (KNDA).[99] Further evidence for the existence of a *gem*-diol intermediate has been obtained by ^{13}C NMR spectroscopy, using a 13C-labeled substrate.[100] Using kinetically impaired mutants of hydrolases MhpC and BphD, a signal corresponding to the *gem*-diol intermediate was observed during the reaction at 128 ppm, whereas no acyl enzyme was observed at ~170 ppm.[100] Furthermore, hydrolase BphD has been shown to catalyze the C–C cleavage of a reduced substrate analogue containing an alcohol at C-6 in place of a ketone, consistent with a general base mechanism, but not with a nucleophilic mechanism.[101]
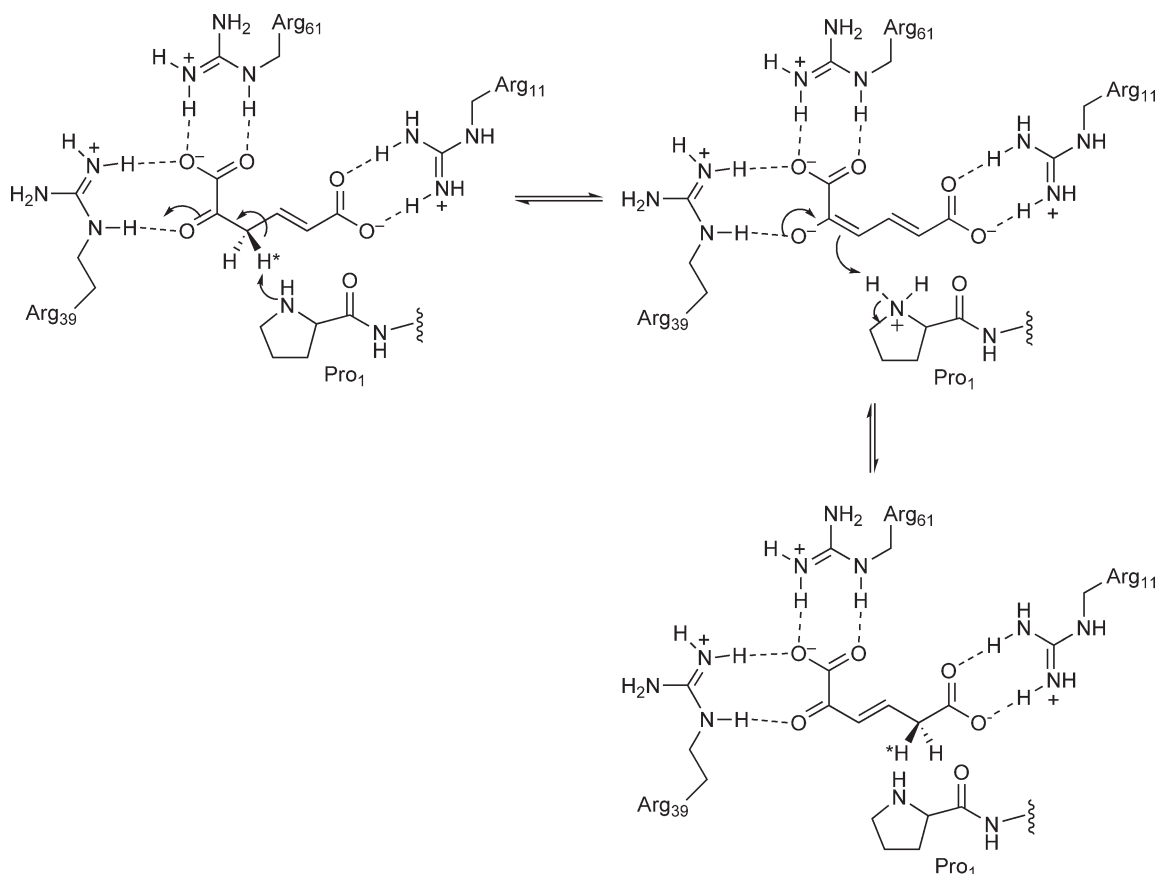
Figure 20   4–Oxalocrotonate tautomerase catalytic mechanism.

Replacement of Ser-110 or His-263 of hydrolase MhpC with Ala caused 10^4-fold drop in k_{cat} in each case, but stopped-flow kinetic analysis of the S110A mutant demonstrated that ketonization proceeded at the same rate as wild-type enzyme, but C–C cleavage was dramatically slowed, whereas mutant H263A was impaired in both ketonization and C–C cleavage steps.[102] Active site residue Arg-188, which binds the C-1 carboxylate of the substrate, was also found to have a catalytic role in substrate ketonization.[103] The proposed catalytic mechanism for MhpC involves protonation at C-5 by His-263, followed by attack of water at the C-6 carbonyl group, deprotonated by His-263. Ser-110 has an important role in C–C cleavage, but appears not to act as a nucleophile, and it has been proposed that Ser-110 could form a strong hydrogen bond with the oxyanion intermediate.[104]

The product of the C–C hydrolase enzymatic cleavage is 2-hydroxypentadienoic acid, a dienol with a half-life of 5–10 min in aqueous solution.[105] This intermediate is the substrate for addition of water, by a metal-dependent hydratase enzyme, which in the case of *E. coli* hydratase MhpD has been shown to be Mn^{2+}.[106] The structure of the analogous hydratase enzyme HpcG, on the homoprotocatechuate pathway of *E. coli*, has been determined in the presence of an oxalate inhibitor.[107]

The 4-hydroxy-2-ketopentanoic acid product is converted by an aldolase enzyme into pyruvate and acetaldehyde. Aldolase MhpE on the phenylpropionate pathway in *E. coli* has been shown to be a class I aldolase,[108] and is coupled to a dehydrogenase MhpF, which converts acetaldehyde into acetyl-CoA.[109] The terminal aldolase on the homoprotocatechuate pathway of *E. coli*, HpcH, has in contrast been shown to be a class II aldolase, requiring a divalent metal ion cofactor, and its structure has been determined.[110] Aldolase HpaI on the *E. coli* hydroxyphenylacetate degradation pathway has also been characterized, and shown to be a class II aldolase.[111]
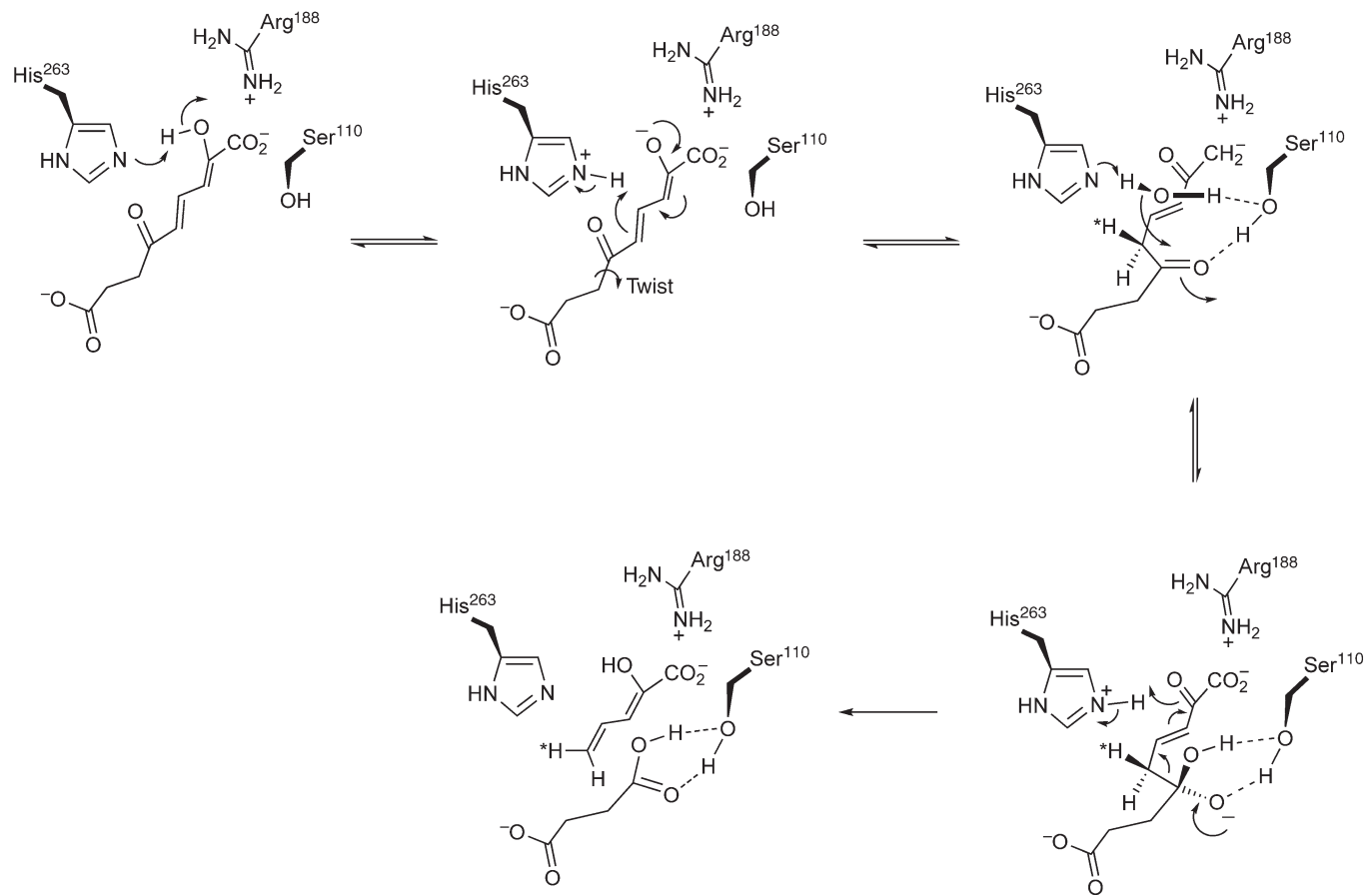
Figure 21 Catalytic mechanism of C–C hydrolase.

8.16.2   Mammalian Aromatic Amino Acid Degradation Pathways

8.16.2.1   L-Tyrosine Degradation Pathway

The amino acids L-phenylalanine and L-tyrosine are broken down via aromatic degradation pathways that are found in mammals and bacteria, to form organic acids that can be utilized for growth. These pathways are the only aromatic degradation pathways found in mammals, and are of some medical significance, since there are several inherited metabolic diseases (phenylketonuria, alkaptonuria, tyrosinemia) that are caused by mutations in enzymes in these pathways.

The mammalian L-tyrosine degradation pathway is shown in **Figure 22**. Transamination of L-phenylalanine gives *p*-hydroxyphenylpyruvate, which is a substrate for *p*-hydroxyphenylpyruvate dioxygenase (HPPD), an iron(II)-dependent enzyme, which converts *p*-hydroxypehylpyruvate to homogentisic acid. Homogentisic acid is a substrate for oxidative cleavage by homogentisic acid dioxygenase, to give maleylacetoacetate, which is isomerized by a glutathione-dependent isomerase to give fumarylacetoacetate, followed by C–C hydrolytic cleavage to give fumarate and oxaloacetate. Patients suffering from type I tyrosinemia show dramatically reduced levels of fumarylacetoacetate hydrolase (FAH), resulting in accumulation of the ring fission intermediate, fumarylacetoacetate, which is toxic, due to reaction of its unsaturated carbonyl functional group with cellular proteins.[112]

The reaction catalyzed by HPPD is related to that of the α-ketoglutarate dioxygenases (see Section 8.16.4), which catalyze the oxidative decarboxylation of α-ketoglutarate to form succinate and an iron(IV)–oxo species. HPPD catalyzes a similar oxidative decarboxylation of the α-keto acid functional group of *p*-hydroxyphenylpyruvate, to generate an iron(IV)–oxo species, which then mediates an oxidative rearrangement to homogentisic acid, as shown in **Figure 23**. Electrophilic reaction of the iron(IV)–oxo species with the aromatic ring generates a carbocation intermediate, which undergoes a 1,2-rearrangement, followed by elimination of H^+, to give homogentisic acid. The crystal structures of HPPD from *P. fluorescens*[113] and *Streptomyces avermitilis*[114] have been determined, in the latter case in complex with a herbicide NTBC that targets this enzyme in plants.[114]

Oxidative cleavage of homogentisic acid to maleylacetoacetate is catalyzed by homogentisate dioxygenase. Mutations in homogentisate dioxygenase cause the metabolic disease alkaptonuria, a benign condition first noted in the seventeenth century (for a colorful description of alkaptonuria, see Stryer[115]) in which the oxidation of homogentisic acid causes a darkly colored pigment in urine. This metabolic disease was the first example of an 'inborn error of metabolism' caused by a genetic mutation, proposed by Garrod in 1902.[116] The crystal structure of human homogentisate dioxygenase was determined in 2000: the active site iron(II) cofactor is ligated by His-335, Glu-341, and His-371,[117] in a similar arrangement to the bacterial extradiol catechol dioxygenases (see Section 8.16.1.6).

The initial ring cleavage product, maleylacetoacetate, undergoes a *cis–trans* isomerization, catalyzed by a glutathione-dependent isomerase enzyme.[118] The hydrolytic cleavage of fumarylacetoacetate is catalyzed by FAH, a member of the β-ketolase family of C–C hydrolase enzymes.[119] Mutations in the FAH gene are the primary cause of hereditary tyrosinemia type I.[120] The crystal structure of human FAH was determined in 1999,[121] and the structure of a complex with a phosphinate transition state analogue has since been determined.[122] The active site contains a divalent metal ion cofactor (Ca^{2+} or Mg^{2+}), which ligates the β-keto acid functional group. His-133 is proposed to act as a base to deprotonate water, in a general base mechanism, and the oxyanion intermediate formed is stabilized by Arg-237 and Gln-240. C–C cleavage is assisted by the β-keto acid functional group, and reprotonation of the acetoacetate product is by Lys-253, as shown in **Figure 24**.

Bacterial degradation of naphthalene (see **Figure 7**) leads to the formation of salicylic acid (2-hydroxybenzoic acid), which is converted via a monooxygenase enzyme in *Pseudomonas* sp. into gentisic acid.[123] Oxidative cleavage of gentisic acid is catalyzed by an iron(II)-dependent gentisate 1,2-dioxygenase, which has been purified from *Pseudomonas testosteroni* and *Pseudomonas acidovorans*.[85] Following ring cleavage, genes encoding a glutathione-dependent isomerase and fumarylpyruvate hydrolase have been identified.[124]

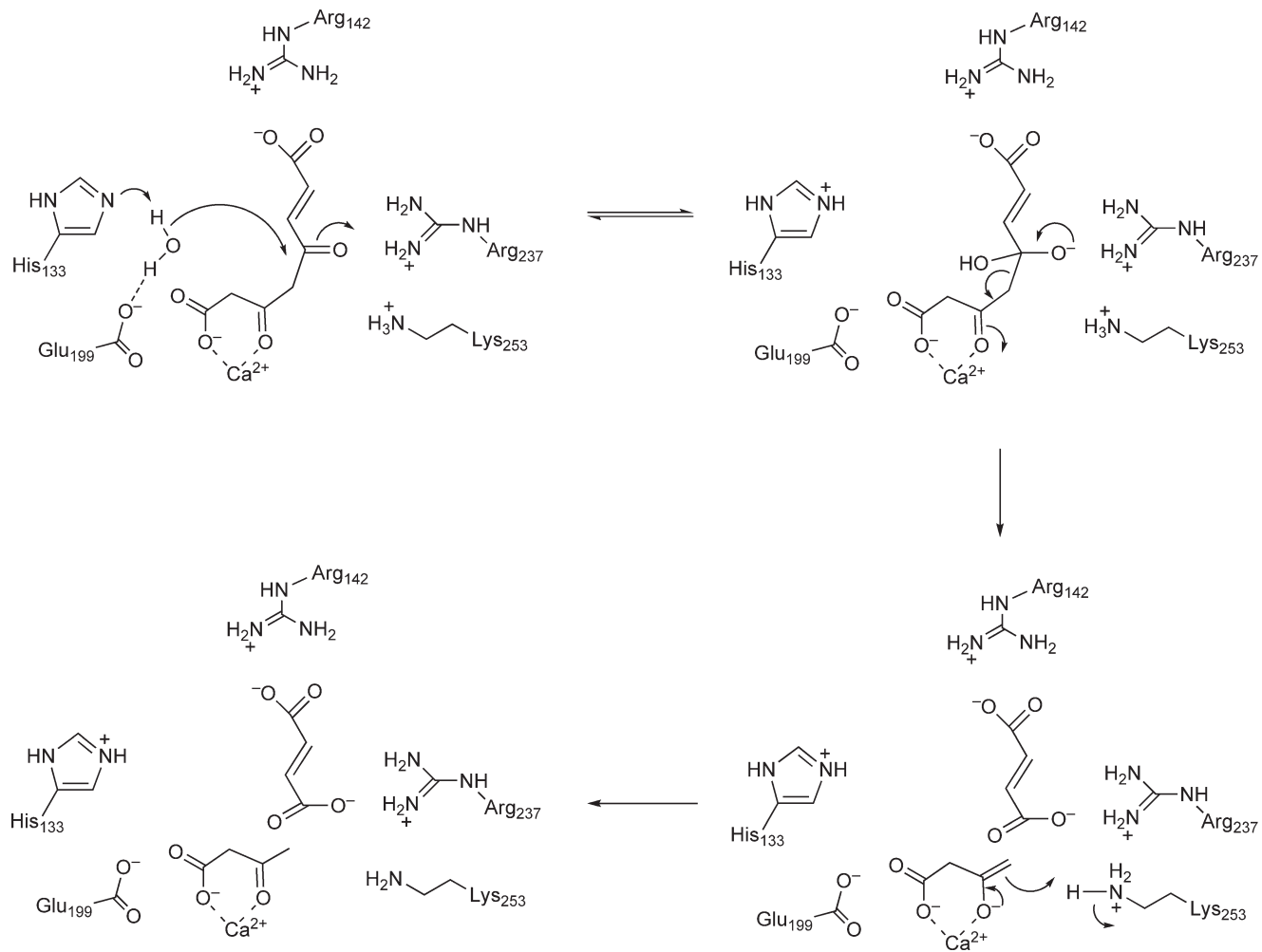### 8.16.2.2   L-Tryptophan Degradation Pathway

L-Tryptophan is degraded via oxidative cleavage of the C-2,C-3 bond, to give *N*-formyl-kynurenine, as shown in **Figure 25**. Tryptophan 2,3-dioxygenase (TDO), found in mammals and bacteria, is selective for cleavage of tryptophan, whereas indoleamine 2,3-dioxygenase (IDO), found only in mammals, is able to cleave other
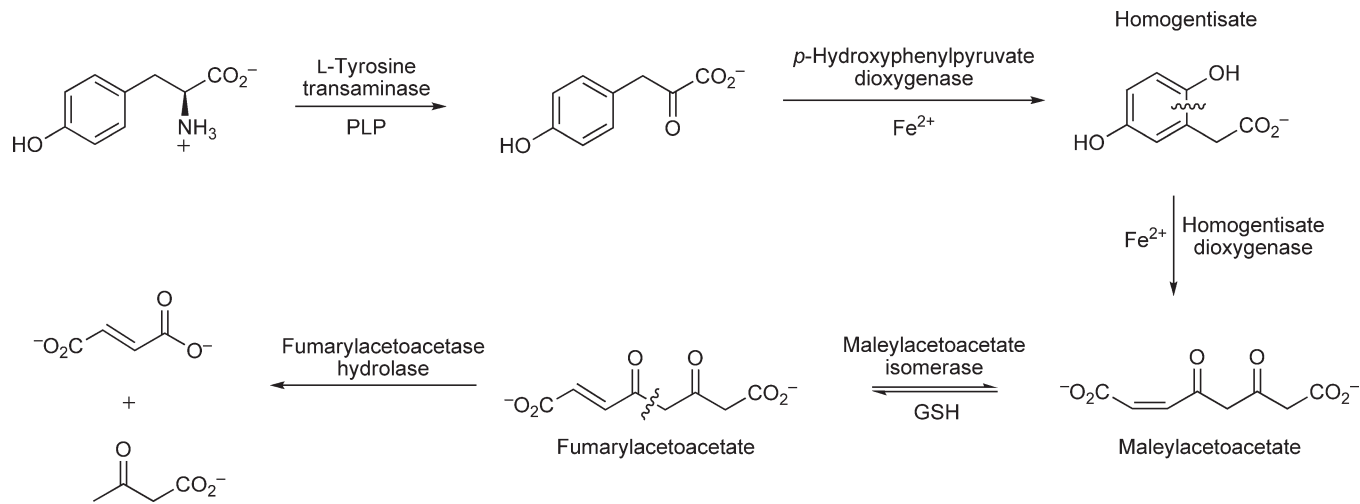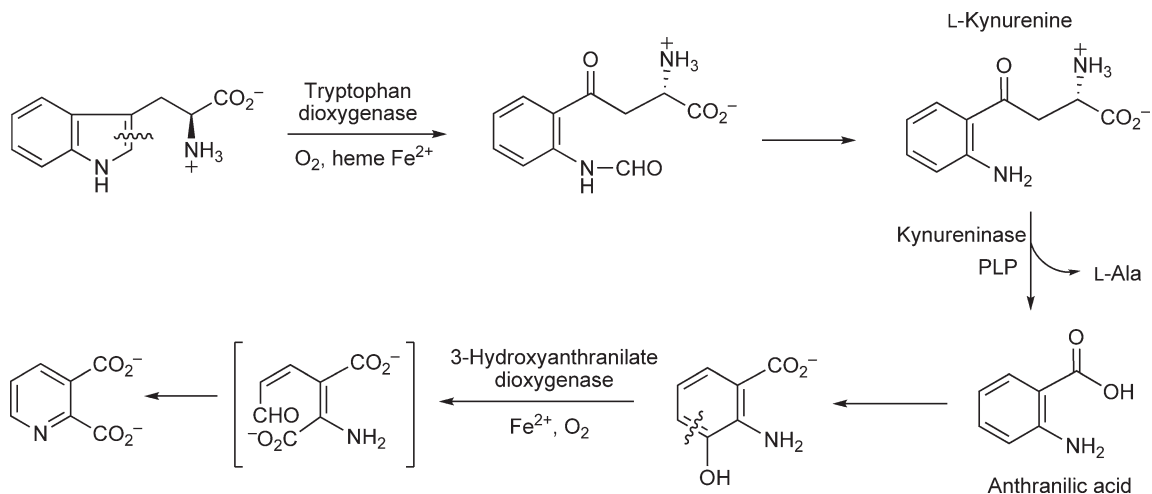
Figure 22 Tyrosine degradation pathway.
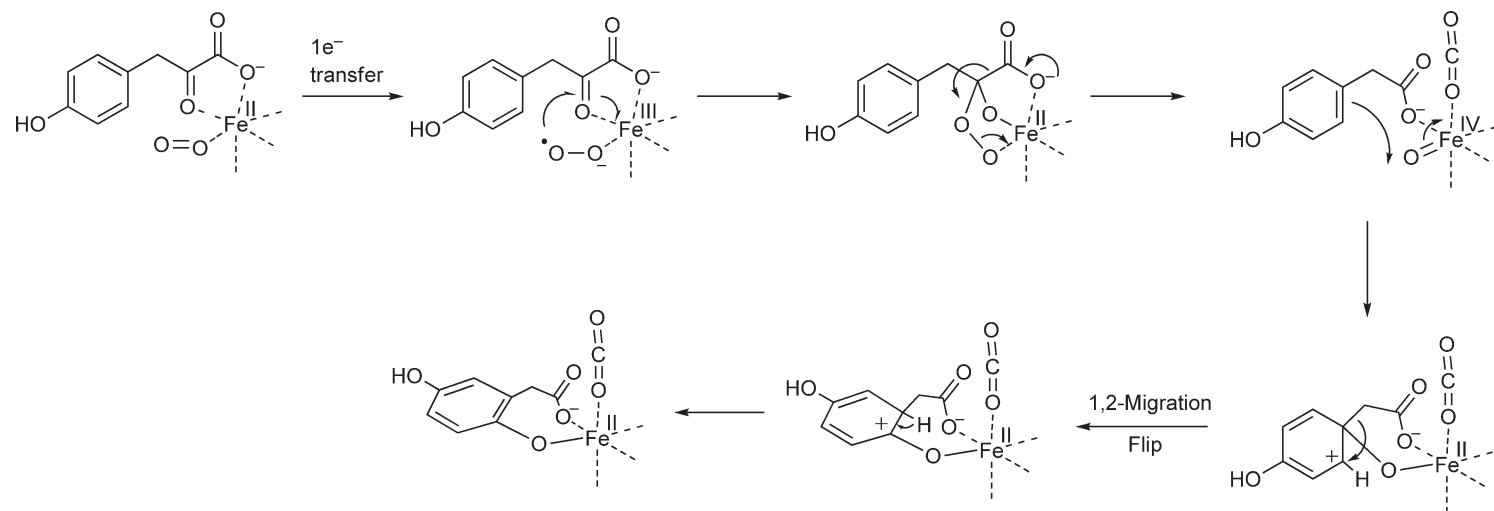
Figure 23 Catalytic mechanism of HPPD.

Figure 24 Catalytic mechanism of fumarylacetoacetate hydrolase.

Figure 25   L-Tryptophan degradation pathway.

indoleamine derivatives such as tryptamine and serotonin.[125] Both enzymes contain a heme iron cofactor, unlike the catechol dioxygenases involved in other aromatic degradation pathways, which utilize nonheme iron.[125]

The catalytic mechanism for IDO and TDO is believed to proceed via the formation of a hydroperoxide at C-3 of the indole ring, followed either by dioxetane formation or Criegee rearrangement,[125] as shown in **Figure 26**. Formation of the hydroperoxide could either take place via nucleophilic attack upon heme-bound dioxygen, or via the formation of an indole radical, followed by recombination with iron(III)–superoxide.[125] The structure of human IDO was published in 2006.[126] Site-directed mutagenesis of active site residues has established that Phe-226, Phe-227, and Arg-231 contribute toward catalysis.[126]

It was discovered in 1998 that expression of IDO activity in the mouse fetus represses the maternal T-cell activity and hence protects the fetus from the maternal immune system.[127] Pregnant mice treated with the IDO inhibitor 1-methyltryptophan rejected the embryos via their immune system, thus either IDO itself or a product of tryptophan catabolism is able to suppress the maternal T-cell activity.[127] IDO is also expressed in response to interferon γ from activating T-cells, inhibiting T-cell proliferation[128] and contributing toward the antiviral activity of interferon γ.[129] The end product of the L-tryptophan degradation pathway, quinolinic acid, has neurological effects,[130] hence the IDO pathway is implicated in several mammalian regulatory pathways.

Deformylation of the cleavage product yields L-kynurenine, which is the substrate for hydrolytic C–C cleavage by kynureninase, a pyridoxal 5′-phosphate-dependent enzyme. The α-amino group of L-kynurenine is attached to the PLP cofactor, and deprotonation of the α-carbon forms a ketimine linkage, which provides an electron sink to assist C–C hydrolytic cleavage, as shown in **Figure 27**.

The crystal structure of kynureninase from *P. fluorescens* was solved in 2004.[131] The enzyme shares the same structural fold as aspartate aminotransferase, but shares low sequence similarity. An active site arginine residue (Arg-375) was identified, which is important in substrate binding.[131] The structure of the human kynureninase, which shows a catalytic preference for 3-hydroxy-kynurenine over L-kynurenine, was solved in 2007.[132] The human enzyme shares the same fold as the *P. fluorescens* enzyme, and also contains an active site arginine residue (Arg-434).[132] The catalytic mechanism requires two acid/base residues, which have not yet been unambiguously assigned. The hydrolytic cleavage step is believed to proceed via a general base mechanism.[133]

Kynurenine is converted to 3-hydroxykynurenine in mammals by a specific hydroxylase enzyme.[134] Oxidative cleavage of 3-hydroxykynurenine is catalyzed by a nonheme iron-dependent dioxygenase enzyme. The gene for the human 3-hydroxyanthranilate dioxygenase was identified in 1994.[135] A 4-chloro-substituted substrate analogue is known to be a potent inhibitor of the human enzyme, and the neurological effects of this inhibitor have been investigated.[136] The bacterial genes encoding enzymes of this pathway have been identified
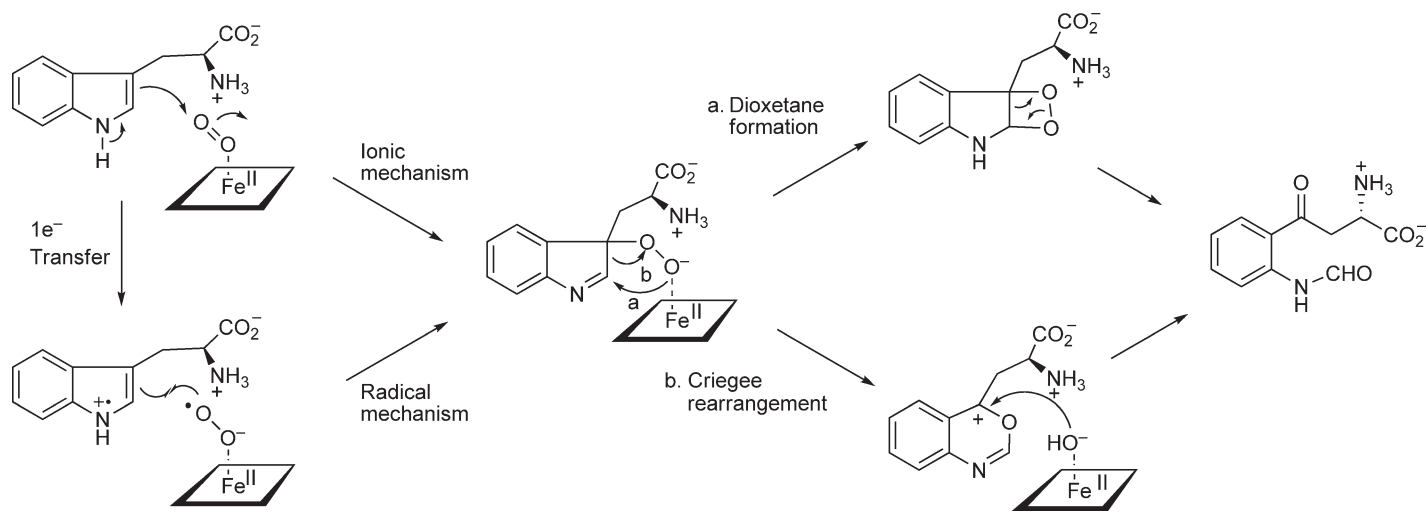
Figure 26 Possible catalytic mechanisms for tryptophan/indolamine dioxygenase.
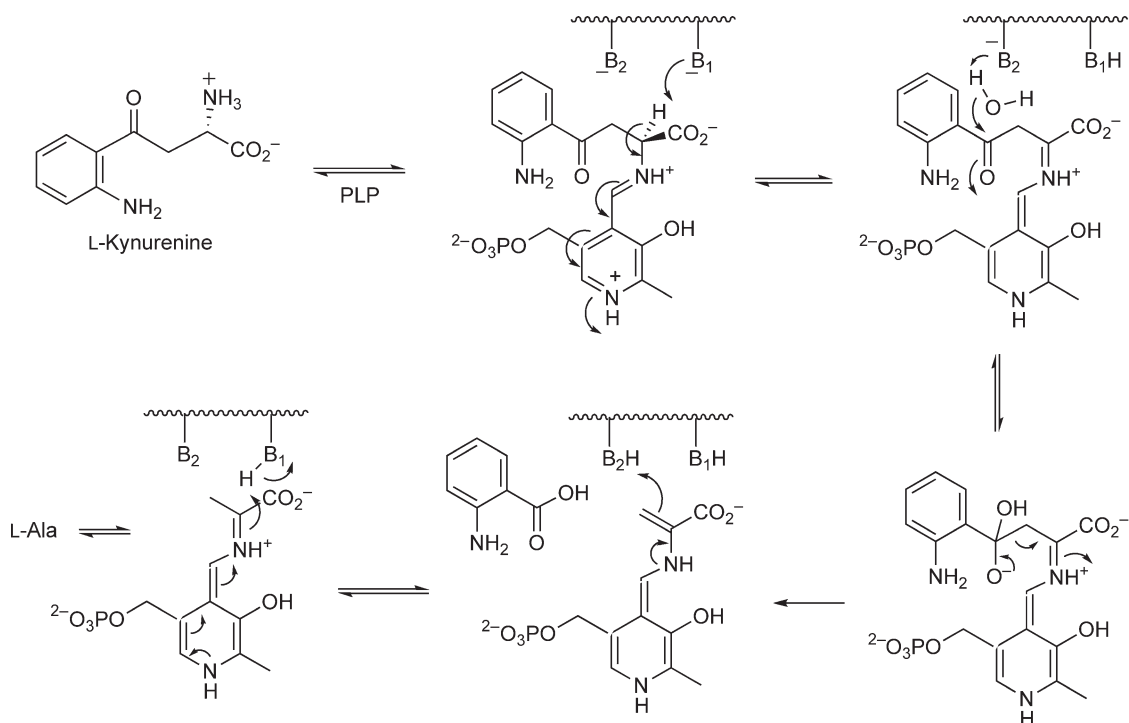
**Figure 27** Catalytic mechanism for kynureninase.

in *Ralstonia metallidurans*, and a 22 kDa 3-hydroxyanthranilate dioxygenase enzyme overexpressed and purified to homogeneity.[137] The crystal structure of 3-hydroxyanthranilate 3,4-dioxygenase from *Ralstonia metallidurans* was reported in 2005 by Zhang *et al.*[138] The structure of a complex with a 4-chloro-substituted inhibitor was solved, in the presence of bound NO, providing insight into enzyme–substrate binding interactions. The enzyme contains a His,His,Glu motif for iron(II) cofactor binding; however the identity of outer sphere active site residues is different to the catechol dioxygenases. Arg-47 is positioned close to bound NO, and the only active site acid–base residue is Glu-110.[138] Using the recombinant bacterial enzyme, the mechanism of inactivation by 4-chloro-3-hydroxyanthranilate was investigated: it was found that the enzyme was able to activate dioxygen in the presence of inhibitor, but then was unable to complete the reaction, and dissociated superoxide, to leave an inactive iron(III) center.[139] The bacterial L-tryptophan degradation pathway is responsible for the biosynthesis of the nicotinamide base of the coenzyme nicotinamide adenine dinucleotide.[137]

8.16.3 Carotenoid Oxidative Cleavage Pathways

8.16.3.1 Biosynthesis of Retinal in Mammals

The oxidative cleavage of vitamin A, β-carotene, to form retinal, used as a cofactor in the light-sensing protein rhodopsin, has been known since 1965 (**Figure 28**),[140] and this important transformation rationalizes why vitamin A is required in the human diet for night vision. However, only in 2000 were the genes encoding this enzyme identified in chicken intestinal mucosa,[141] *Drosophila melanogaster*,[142] and mouse kidney.[143] Overexpression of the enzyme has enabled a series of mechanistic studies to be undertaken on the recombinant enzyme, which requires iron(II).[144] Substrate specificity studies revealed that the enzyme is selective for a rod-like polyene substrate, but that some modifications to the substituents of the polyene are tolerated by the enzyme.[145] Incubations carried out in $^{17}O_2$ and $H_2^{18}O$ using an asymmetric substrate revealed approximately 50
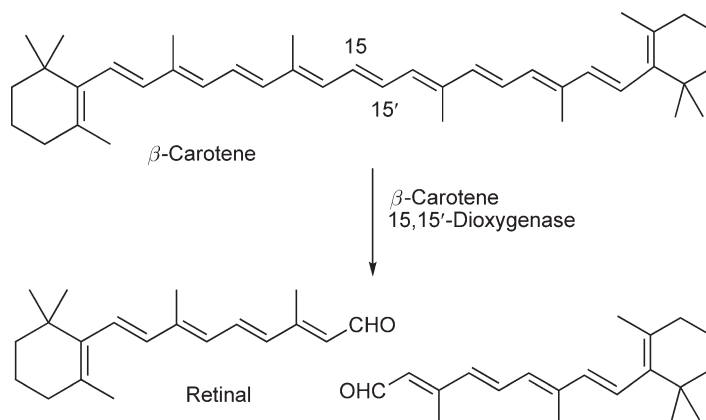
Figure 28   Oxidative cleavage by β-carotene dioxygenase.

atom % incorporation of oxygen into each aldehyde product from each source, suggesting the possibility of an epoxide intermediate.[144]

8.16.3.2   The Carotenoid Cleavage Dioxygenase Family

Since 2000, it has become apparent that there is a much larger family of nonheme iron-dependent dioxygenases that catalyze oxidative cleavage reactions on carotenoid substrates, predominantly in plants. The first member of this family to be identified was 9-*cis*-epoxycarotenoid dioxygenase (NCED), which catalyzes the oxidative cleavage of the epoxy-carotenoid 9-*cis*-neoxanthin, a key step in the biosynthesis of plant growth regulator abscisic acid (**Figure 29**).

The gene encoding NCED was identified in maize as the *vp14* gene in 1997, and the encoded enzyme shown to catalyze oxidative cleavage of 9-*cis*-neoxanthin (**Figure 30**).[146,147] The genome sequence of *Arabidopsis thaliana* contains nine gene sequences with sequence similarity to *vp14*,[148] and the family of encoded enzymes is known as the CCDs.[149,150] Five of the *Arabidopsis* CCDs appear to encode NCED isoenzymes.[148] Of the remaining enzymes, CCD1 cleaves several carotenoid substrates at the 9,10 or 9′,10′ positions, to release apocarotenoids such as β-ionone, a component of flower fragrance,[151] as shown in **Figure 34**. The other cleavage product can be converted biosynthetically into rosafluene, a natural product found in roses. Several other apocarotenoid cleavage products are known, including crocin and crocetin, the pigments found in the saffron crocus (*Crocus sativum*), which arise from 7,8-cleavage.[152]

It is thought that the CCD family catalyzes a variety of oxidative cleavage reactions at different points along the polyene chain of carotenoid substrates, to produce apocarotenoid natural products, at least some of which have functions as signaling molecules.[153] The *max3* (CCD7) and *max4* (CCD8) genes have been implicated in the shoot branching response in plants, and are believed to catlyze two steps in the biosynthesis of an unknown plant signaling molecule, which controls shoot branching.[154,155]

There is a crystal structure of one member of the CCD family, a *Synechocystis* 15,15′-dioxygenase enzyme.[156] The nonheme iron(II) cofactor is ligated by four histidine residues, a different coordination state compared with other nonheme iron-dependent oxygenases. The iron(II) center is positioned 5–6 Å from the bound carotenoid substrate, which is held in a hydrophobic channel lined with a number of aromatic amino acid side chains.[156] The distance from the iron(II) cofactor to the polyene chain of the substrate suggests that dioxygen binds between the metal cofactor and the substrate, which then reacts at the distal end of dioxygen. Another possible clue to the catalytic mechanism is the existence of a tertiary amine inhibitor, Abamine, for NCED (**Figure 31**).[157] At neutral pH this tertiary amine will be protonated, and may therefore mimic a carbocation intermediate, which could arise as shown in **Figure 35**. In the later stages of the mechanism, either dioxetane formation or Criegee rearrangement steps are possible.
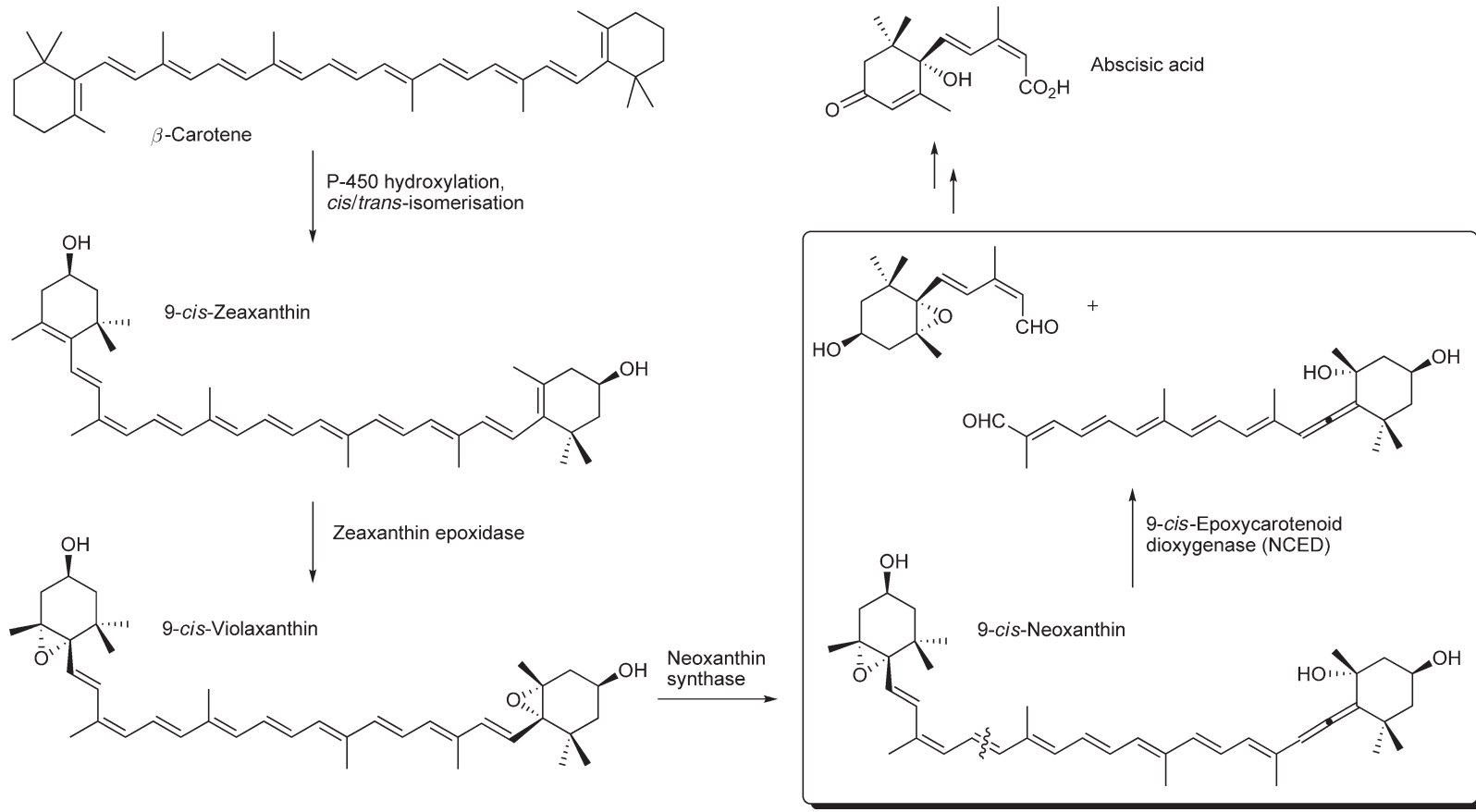
Figure 29 Biosynthesis of abscisic acid via NCED.
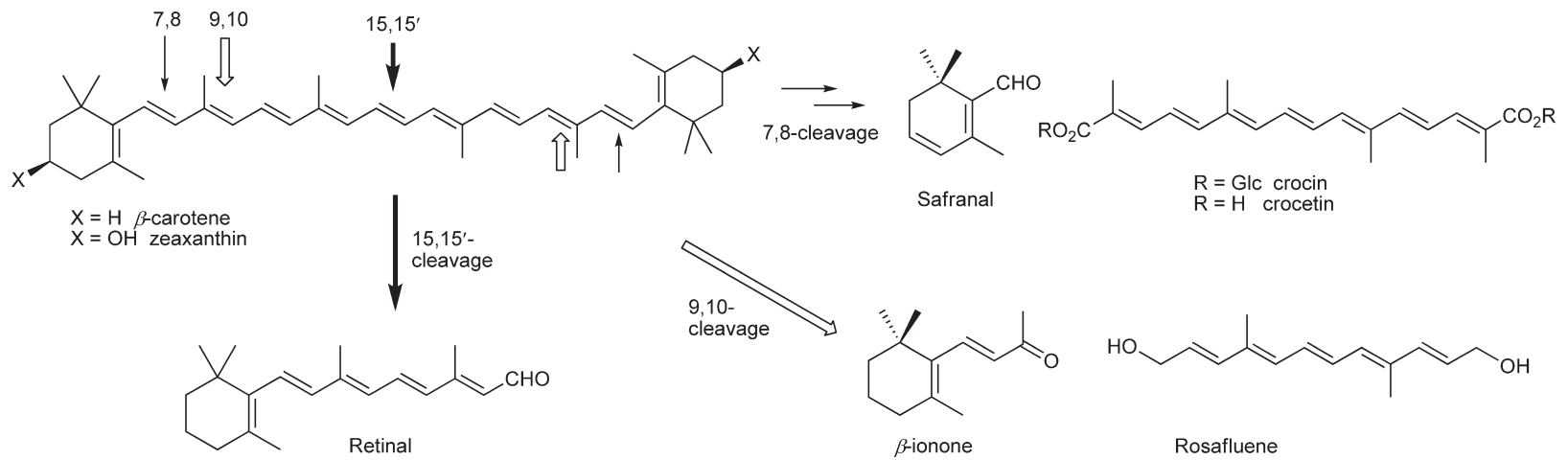
Figure 30 Biosynthesis of apocarotenoids via 7,8- and 9,10-oxidative cleavage.
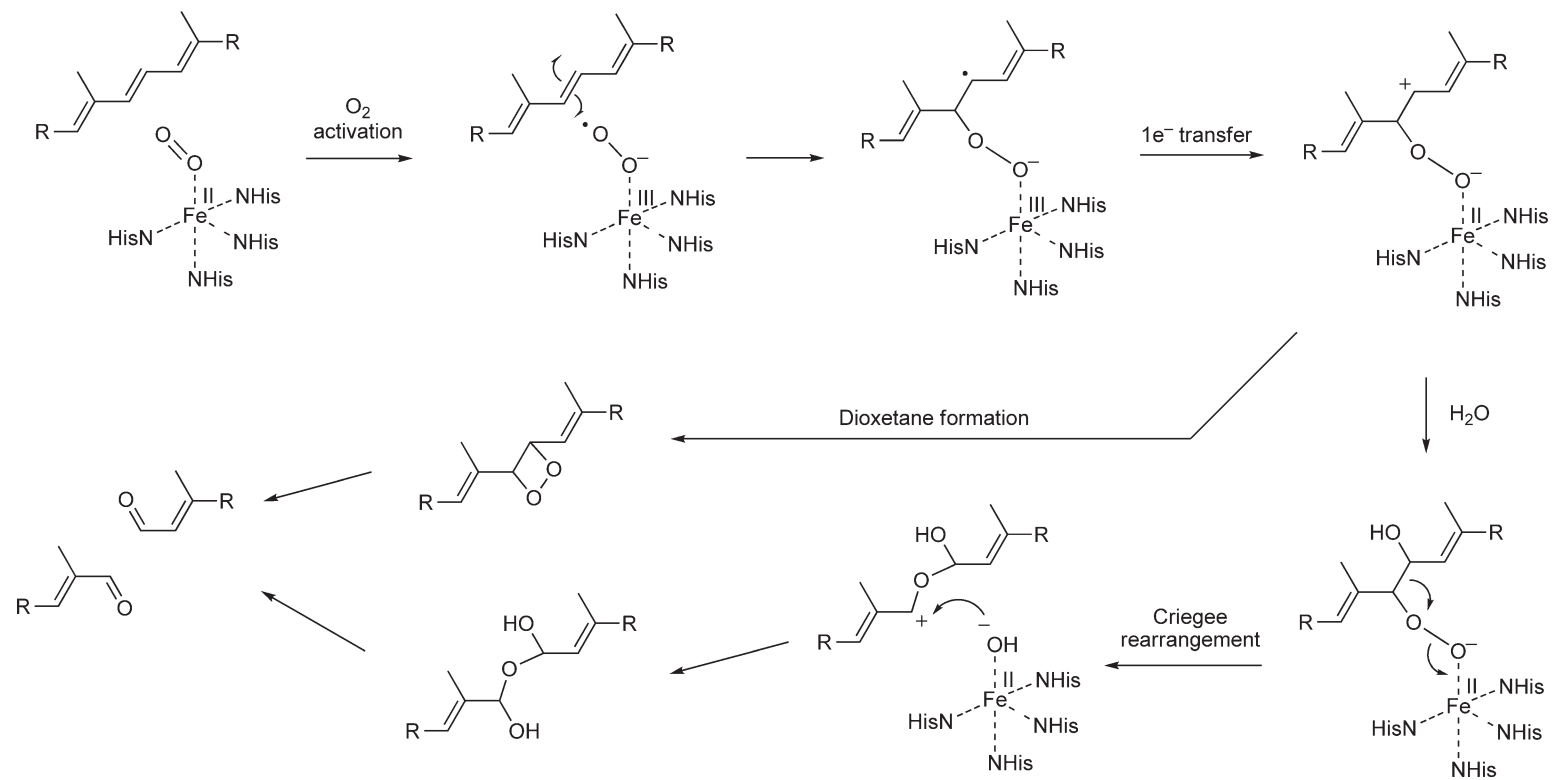
Figure 31 Possible catalytic mechanisms for NCED-catalyzed reaction.

8.16.4   Other Dioxygenase Enzymes Involved in Catabolic and Biosynthetic Pathways

8.16.4.1 α-Ketoglutarate-Dependent Dioxygenases

The other major class of nonheme iron-dependent dioxygenases are the α-ketoglutarate-dependent dioxygenases, which catalyze the oxidative decarboxylation of cosubstrate α-ketoglutarate to form succinate and an iron(IV)–oxo intermediate, which is then used to carry out a range of hydroxylation, desaturation, and other oxidative reactions. While the majority of reactions catalyzed by this family of enzymes are involved in biosynthetic pathways, enzymes such as HPPD (see Section 8.16.2.1) are involved in degradation pathways, therefore it is appropriate to discuss this family of enzymes, and contrast them with the nonheme iron-dependent dioxygenases described in Section 8.16.1.

Prolyl hydroxylase was the first α-ketoglutarate-dependent dioxygenase to be identified, in 1967, by Udenfriend.[158] This enzyme catalyzes the hydroxylation of prolyl residues in collagen to 4-hydroxy-prolyl residues (Figure 32). One oxygen atom is incorporated from dioxygen into the hydroxylated product, and one into succinate.[159]

A large family of these enzymes is now known, and their enzymology and structures have been reviewed.[160–162] A number of crystal structures have been obtained for enzymes in this family, and in each case the mononuclear iron(II) center is coordinated by a His,His,Glu motif, also observed in the extradiol catechol dioxygenases, and in other nonheme iron-dependent enzymes.[161,162] Structural studies on clavaminic acid synthase have indicated the structural basis for the separate hydroxylation and oxidative cyclization/desaturation reactions catalyzed by this enzyme.[163]

By analogy with cytochrome P-450-catalyzed monooxygenase reactions, it has been proposed that the catalytic mechanism of these enzymes involves a high-valent iron–oxo intermediate.[164,165] The iron–oxo intermediate has been shown to undergo partial exchange with the oxygen atom of solvent water in deacetoxy/deacetylcephalosporin C synthase,[166] p-hydrophenylpyruvate hydroxylase,[167] α-ketoisocaproate oxygenase,[168] and lysyl hydroxylase,[169] but not in prolyl hydroxylase.[170] Evidence for a radical intermediate has been provided by the mechanism-based inactivation of prolyl hydroxylase by a substrate analogue containing a labile N–O bond adjacent to the site of hydroxylation.[171]

Therefore, the catalytic mechanism is believed to proceed via formation of an iron(III)–superoxide complex, followed by attack of superoxide upon the ketone carbonyl group of α-ketoglutarate (Figure 33). Decarboxylation of the resulting hydroperoxide intermediate, with cleavage of the O–O bond, then generates succinate and an iron(IV)–oxo intermediate. The iron–oxo species then effects hydroxylation of the substrate, probably via hydrogen atom abstraction to form a substrate radical intermediate (see Figure 29). Direct spectroscopic evidence has more recently been obtained for the iron(IV)–oxo intermediates: in oxygenase TauD using Raman spectroscopy, where a band at $859 \, \text{cm}^{-1}$ corresponding to Fe(IV)=O was observed, which was shifted in the presence of ^{18}O;[172] and in prolyl hydroxylase, where an Fe(IV)=O intermediate was characterized kinetically and spectroscopically, using Mössbauer spectroscopy.[173]

In the reaction catalyzed by HPPD, the α-keto acid used for oxidative decarboxylation is in the substrate molecule. Interestingly, another dioxygenase enzyme also uses the same substrate to catalyze a different
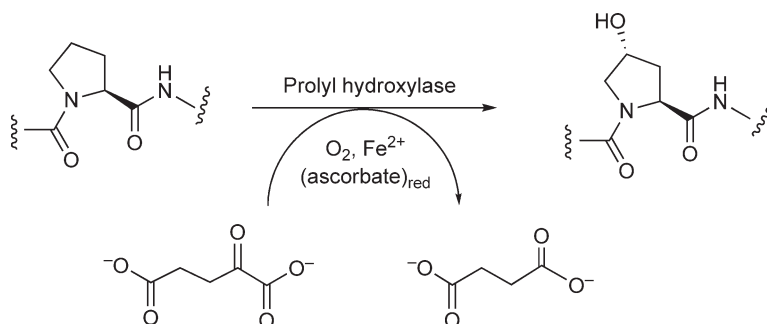
Figure 32   Reaction catalysed by prolyl hydroxylase.
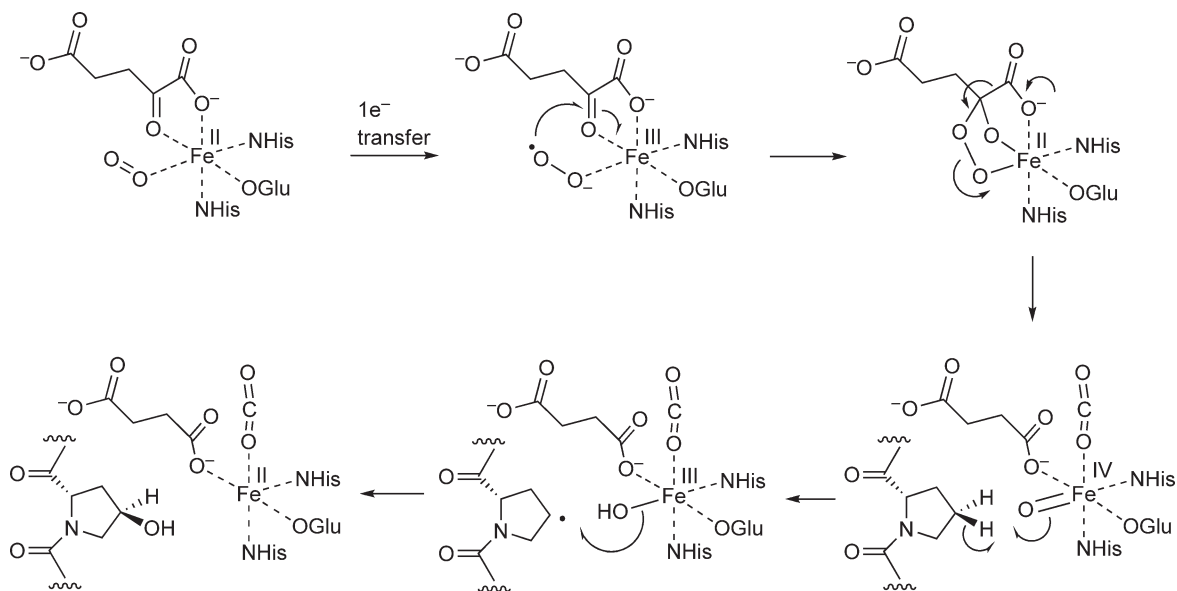
Figure 33   Catalytic mechanism for prolyl hydroxylase.

oxidative conversion. *p*-Hydroxymandelic acid synthase (HMAS) catalyzes the conversion of *p*-hydroxyphenylpyruvic acid into *p*-hydroxymandelic acid, as part of the biosynthetic pathway to the glycopeptide antibiotic vancomycin.[174,175] This enzyme shares 34% amino acid sequence identity with HPPD, which converts the same substrate into homogentisic acid, as part of the tyrosine degradative pathway (see Section 8.16.2.1). Both enzymes effect the oxidative decarboxylation of *p*-hydrophenylpyruvate to *p*-hydroxyphenylacetate, generating the iron(IV)–oxo intermediate, which then carries out either hydroxylation in the benzylic position or electrophilic hydroxylation at C-1 of the aromatic ring, followed by a 1,2-alkyl shift, as shown in **Figure 34**.

Members of the α-ketoglutarate-dependent dioxygenase family are involved in a number of biosynthetic pathways,[160–162] and have recently been implicated in mammalian oxygen sensing by hypoxia inducible factor (HIF), via hydroxylation of Pro-402, Pro-564, and Asn-803 of HIF-1α.[176,177]

8.16.4.2   Flavonoid Oxidative Cleavage by Quercetin 2,3-Dioxygenase

Flavonoid natural products occur widely in plants, and are degraded by *Aspergillus*. The initial step of oxidative ring cleavage is catalyzed by quercetin 2,3-dioxygenase, which is the only known copper-dependent dioxygenase enzyme (**Figure 35**).[178,179] Oxidative cleavage of two carbon–carbon bonds takes place, liberating carbon monoxide as a by-product. The structure of *Aspergillus japonicus* quercetin 2,3-dioxygenase was solved in 2002 by Steiner *et al.*, revealing that the mononuclear copper(II) center is ligated by three active site histidine residues and one glutamic acid residue, and is positioned close to the C-3 hydroxyl group, allowing radical formation at this center.[180] Reaction with dioxygen is assisted by the binding of the substrate in a bent conformation, thereby lowering the activation energy by relief of strain.[180] The catalytic mechanism then proceeds via intramolecular nucleophilic attack of a hydroperoxide anion, followed by cheletropic ring opening, releasing CO.

8.16.4.3   Bacterial Degradation of Quinolines via Cofactor-Independent 2,4-Dioxygenases

The bacterial degradation of quinoline heterocycles proceeds via oxidation to 3-hydroxy-4-oxoquinolines, which are substrates for a novel family of dioxygenases.[181] The reaction is similar to the copper-dependent quercetin dioxygenase (see Section 8.16.4.2), involving oxidative cleavage of two C–C bonds, and liberation of carbon monoxide; however, remarkably, these enzymes have no cofactor requirement. Two dioxygenases QDO
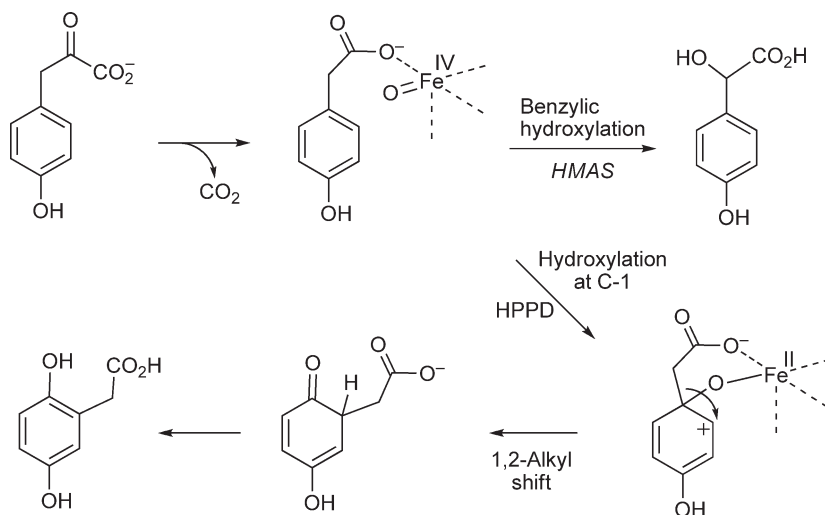
Figure 34   Reactions catalysed by HPPD and HMAS.
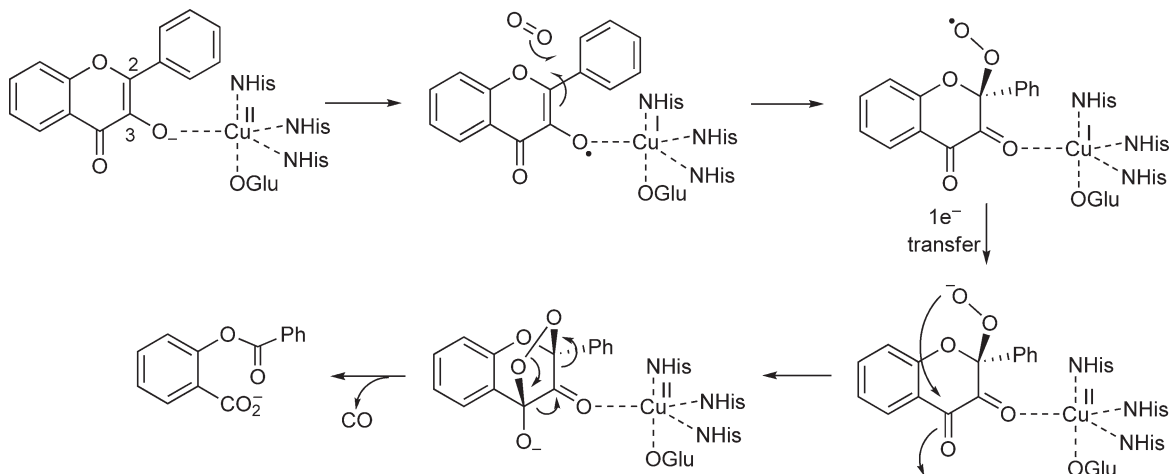
Figure 35   Reaction mechanism of quercetin 2,3-dioxygenase.

(1*H*-3-hydroxy-4-oxoquinoline 2,4-dioxygenase) and HOD (1*H*-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase) have been purified, and neither contains a metal ion nor an organic cofactor.[182]

Amino acid sequence alignments have revealed that QDO and HOD are members of the $\alpha\beta$-hydrolase superfamily, containing a Ser–His–Asp triad, the normal function of which is to participate as a nucleophile in amide and ester hydrolysis reactions.[183] Replacement of Ser-95 in QDO and Ser-101 in HOD by Ala by site-directed mutagenesis gave mutant enzymes with approximately 10% wild-type activity, indicating that the putative active site serine is not essential for activity.[184] The probable catalytic mechanism (**Figure 36**) involves the reaction of a substrate carbanion with dioxygen, via single electron transfer to give superoxide and a stable substrate radical. However, the role of a serine triad in this reaction mechanism is not clear, and thus it is conceivable that a protein radical might be formed.[181]

8.16.4.4   Biosynthesis of Betalain Pigments via Oxidative Cleavage of Dopa

There are several oxidative natural product biosynthetic pathways known to proceed via oxidative cleavage of an aromatic precursor.[185] One of the best-characterized examples is the biosynthesis of the betalain family of pigments, found in plants of the Centrospermae order, and mushrooms of the *Amanita* and *Hygocybe* genera,
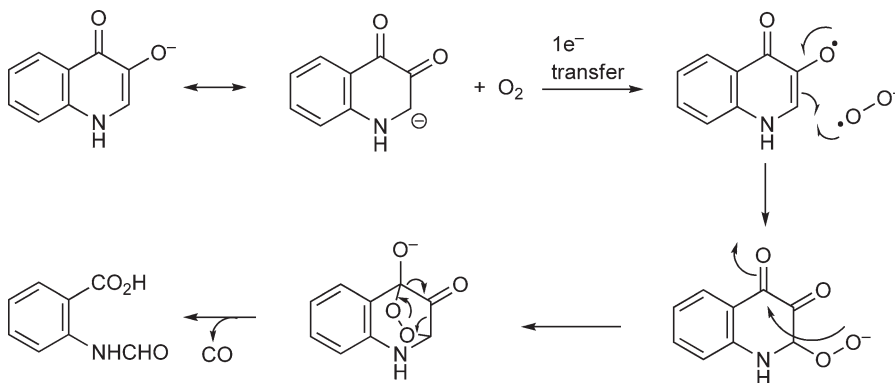
Figure 36   Mechanism for 2,4-dioxygenases.

which occurs via oxidative cleavage of 3,4-dihydroxyphenylalanine (dopa),[186] as shown in **Figure 37**. An extradiol dioxygenase enzyme has been purified from *Amanita muscaria* with catalyzes 4,5-oxidative cleavage of dopa, leading to the formation of betalamic acid and indicaxanthin.[187] The recombinant dioxygenase enzyme was also found to catalyze 2,3-oxidative cleavage, forming muscaflavine, which suggests that there may be a single enzyme responsible for both types of cleavage reactions.[188] A plant gene encoding a dopa 4,5-dioxygenase has also been identified in *Portulaca grandiflora*, which shows sequence similarity with bacterial 4,5-PCD, but no similarity with the *Amanita* dopa 4,5-dioxygenase.[189]

8.16.4.5   Methionine Salvage Pathway Dioxygenases in *Klebsiella pneumoniae*

The biological cofactor S-adenosyl methionine can be converted metabolically into 5′-thiomethyladenosine, which is recycled via opening of the ribose ring to the amino acid L-methionine, as shown in **Figure 38**. An unusual transformation in this pathway is the oxidative cleavage of aci-reductone to 2-keto-4-
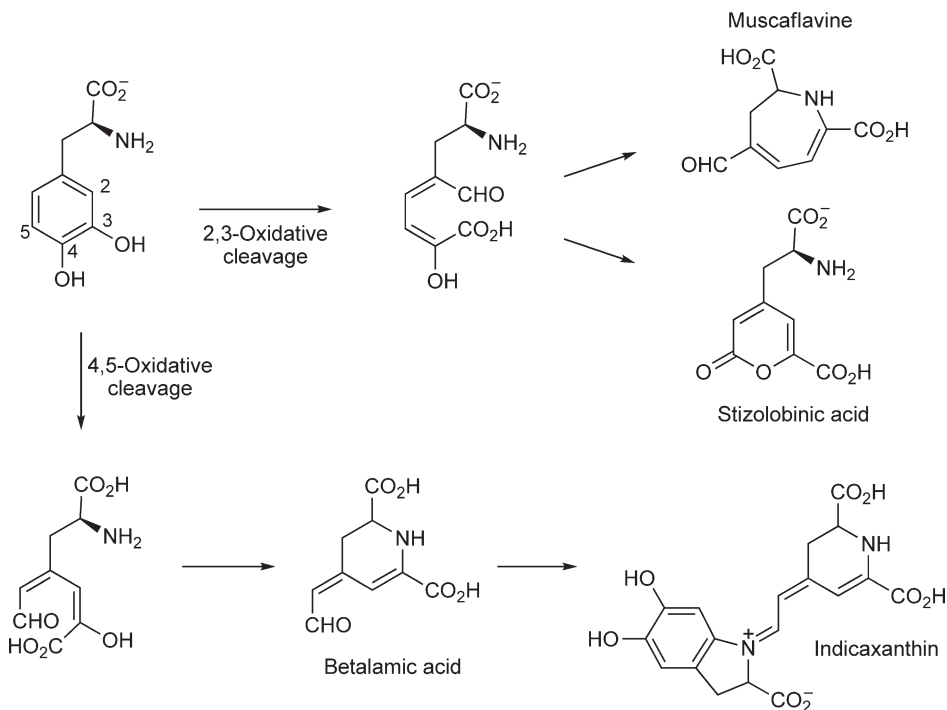
Figure 37   Biosynthesis of betalain pigments via oxidative cleavage of dopa.
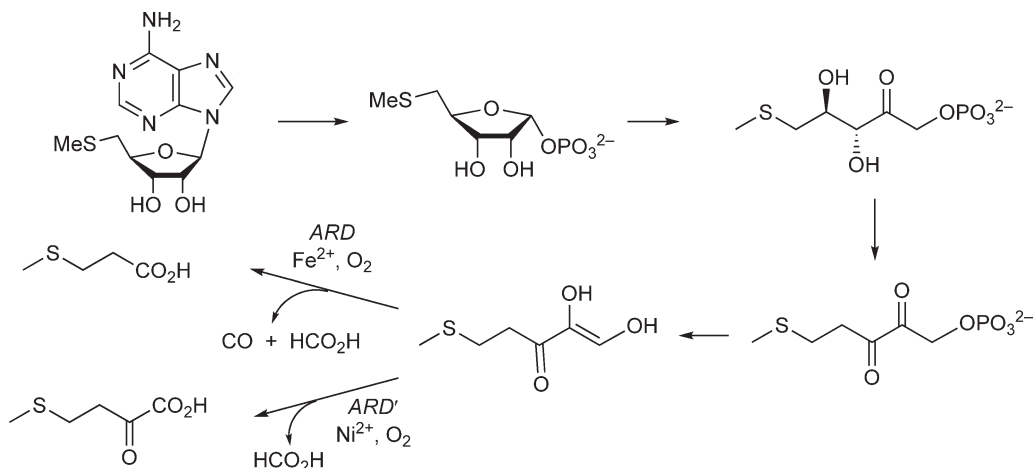
Figure 38 Methionine salvage pathway, illustrating transformations catalyzed by ARD and ARD′.

thiomethylbutyrate and formic acid, while a related transformation converts the same substrate to 3-thiomethylpropionate, formic acid, and CO (**Figure 38**).[190,191]

Abeles and coworkers have purified two dioxygenase enzymes from *Klebsiella pneumoniae* which catalyze these related transformations, and which incorporate two atoms of ^{18}O from $^{18}O_2$ into their respective products.[192] Remarkably, these two enzymes share the same polypeptide sequence, but contain different metal ion cofactors.[193] Dioxygenase ARD contains a single iron(II) cofactor, whereas dioxygenase ARD′ contains a single nickel(II) cofactor.[193] The two oxidative cleavage reactions can be rationalized by the formation of a common hydroperoxide intermediate at C-1, followed either (in b; ARD′) by nucleophilic attack upon C-2 to form a dioxetane intermediate, or (in a; ARD) by nucleophilic attack upon C-3 to form a five-membered endo-peroxide intermediate (**Figure 39**).[194] Evidence for a radical-mediated mechanism has been obtained via enzyme inactivation by a cyclopropyl-containing substrate analogue.[194] The two oxidative cleavages can be observed nonenzymatically, via base-catalyzed auto-oxidation, therefore it seems that the different metal ion cofactors are able to control the choice of reactivity of the hydroperoxide intermediate.

Structure determination of nickel(II)-containing ARD has revealed a mononuclear nickel(II) center ligated by four protein ligands: His-96, His-98, Glu-102, and His-140. The substrate is believed to coordinate to the two vacant coordination sites as the dianion. The protein fold is similar to enzymes of the cupin superfamily.[195]

8.16.4.6   Oxidative Cleavage of Acetylacetone by *Acinetobacter johnsonii*

Acinetobacter johnsonii is able to degrade acetylacetone (2,4-pentanedione), utilizing a novel dioxygenase enzyme, as shown in **Figure 40**. The purified enzyme is dependent upon Fe^{2+} for activity, implying that it is a nonheme iron-dependent dioxygenase.[196] ^{18}O labeling studies and structure–activity relationships have been carried out on this enzyme-catalyzed reaction, which have been interpreted in favor of a mechanism involving a dioxetane intermediate.[197]

In summary, a range of oxidative cleavage reactions, catalyzed by dioxygenase enzymes, have been identified, that are involved in catabolic and biosynthetic pathways. While the majority of these enzymes require nonheme iron as cofactor, several examples of copper-dependent or cofactor-independent dioxygenases have come to light, whose mechanisms for oxygen activation and catalytic mechanism share several features with the nonheme iron-dependent dioxygenases.
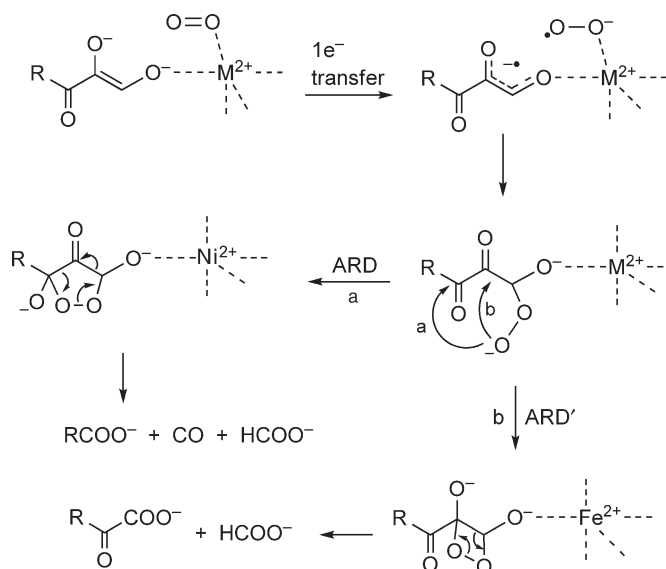
Figure 39 Mechanisms for ARD and ARD′.
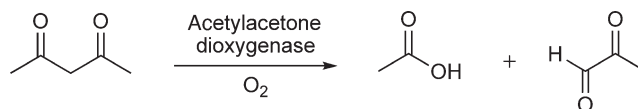
Figure 40 Oxidative cleavage of acetylacetone.

References

1. S. Dagley, *Essays Biochem.* **1975**, *11*, 81–138.
2. E. Diaz; A. Ferrandez; M. A. Prieto; J. L. Garcia, *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 523–569.
3. J. I. Davies; W. C. Evans, *Biochem. J.* **1964**, *91*, 251.
4. D. A. Abramowicz, *Crit. Rev. Biotechnol.* **1990**, *10*, 241–249.
5. S. Harrad; K. Jones, *Chem. Br.* **1992**, *28*, 1110–1112.
6. L. B. Ellis; D. Roe; L. P. Wackett, *Nucleic Acids Res.* **2006**, *34*, D517–D521.
7. F. C. H. Franklin; M. Bagdasarian; M. M. Bagdasarian; K. N. Timmis, *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 7458–7462.
8. W. Reineke; H. J. Knackmuss, *Annu. Rev. Microbiol.* **1988**, *42*, 263–287.
9. J. Heider; G. Fuchs, *Eur. J. Biochem.* **1997**, *243*, 577–596.
10. G. R. Johnson; J. C. Spain, *Appl. Microbiol. Biotechnol.* **2003**, *62*, 110–123.
11. D. L. Crawford; R. L. Crawford, *Enzyme Microb. Technol.* **1980**, *2*, 11–22.
12. M. Ramachandra; D. L. Crawford; G. Hertel, *Appl. Environ. Microbiol.* **1988**, *54*, 3057–3063.
13. E. Masai; Y. Katayama; S. Nishikawa; M. Fukuda, *J. Ind. Microbiol. Biotechnol.* **1999**, *23*, 364–373.
14. K. Kishi; N. Habu; M. Samejima; T. Yoshimoto, *Agric. Biol. Chem.* **1991**, *55*, 1319–1323.
15. S. Kamoda; N. Habu; M. Samejima; T. Yoshimoto, *Agric. Biol. Chem.* **1989**, *53*, 2757–2761.
16. T. D. H. Bugg, *Tetrahedron* **2003**, *59*, 7075–7101.
17. L. Que, Jr.; R. Y. N. Ho, *Chem. Rev.* **1996**, *96*, 2607–2624.
18. B. E. Haigler; D. T. Gibson, *J. Bacteriol.* **1990**, *172*, 457–464.
19. B. E. Haigler; D. T. Gibson, *J. Bacteriol.* **1990**, *172*, 465–468.
20. B. D. Ensley; D. T. Gibson, *J. Bacteriol.* **1983**, *155*, 505–511.
21. W. C. Suen; D. T. Gibson, *J. Bacteriol.* **1993**, *175*, 5877–5881.
22. B. Kauppi; K. Lee; E. Carredano; R. E. Parales; D. T. Gibson; H. Eklund; S. Ramaswamy, *Structure* **1998**, *6*, 571–586.
23. E. Carredano; A. Karlsson; B. Kauppi; D. Choudhury; R. E. Parales; J. V. Parales; K. Lee; D. T. Gibson; H. Eklund; S. Ramaswamy, *J. Mol. Biol.* **2000**, *296*, 701–712.
24. A. Karlsson; J. V. Parales; R. E. Parales; D. T. Gibson; H. Eklund; S. Ramaswarmy, *Science* **2003**, *299*, 1039–1042.
25. R. E. Parales; J. V. Parales; D. T. Gibson, *J. Bacteriol.* **1999**, *181*, 1831–1837.
26. D. T. Gibson; G. E. Cardini; F. C. Maseles; R. E. Kallio, *Biochemistry* **1970**, *9*, 1631–1635.
27. L. P. Wackett; L. D. Kwart; D. T. Gibson, *Biochemistry* **1988**, *27*, 1360–1367.
28. S. M. Resnick; D. S. Torok; K. Lee; J. M. Brand; D. T. Gibson, *Appl. Environ. Microbiol.* **1994**, *60*, 3323–3328.
29. K. Lee, *J. Bacteriol.* **1999**, *181*, 2719–2725.
30. M. B. Neibergall; A. Stubna; Y. Mekmouche; E. Münck; J. D. Lipscomb, *Biochemistry* **2007**, *46*, 8004–8016.

31. M. D. Wolfe; J. V. Parales; D. T. Gibson; J. D. Lipscomb, *J. Biol. Chem.* **2001**, *276*, 1945–1953.
32. K. Chen; L. Que, Jr., *Angew. Chem. Intl. Ed.* **1999**, *38*, 2227–2229.
33. M. Costas; A. K. Tipton; K. Chen; D. H. Jo; L. Que, Jr., *J. Am. Chem. Soc.* **2001**, *123*, 6722–6723.
34. K. Chen; M. Costas; J. Kim; A. K. Tipton; L. Que, Jr., *J. Am. Chem. Soc.* **2002**, *124*, 3026–3035.
35. A. Bassan; M. R. A. Blomberg; P. E. M. Siegbahn, *J. Biol. Inorg. Chem.* **2004**, *9*, 439–452.
36. S. Chakravarty; R. N. Austin; D. Deng; J. T. Groves; J. D. Lipscomb, *J. Am. Chem. Soc.* **2007**, *129*, 3514–3515.
37. O. Hayaishi, *Bacteriol. Rev.* **1966**, *30*, 720–731.
38. O. Hayaishi; M. Katagiri; S. Rothberg, *J. Am. Chem. Soc.* **1955**, *77*, 5450–5451.
39. D. H. Ohlendorf; J. D. Lipscomb; P. C. Weber, *Nature* **1988**, *336*, 403–405.
40. J. W. Pyrz; A. L. Roe; L. J. Stern; L. Que, Jr., *J. Am. Chem. Soc.* **1985**, *107*, 614–620.
41. M. W. Vetting; D. H. Ohlendorf, *Structure* **2000**, *8*, 429–440.
42. A. M. Orville; N. Elango; J. D. Lipscomb; D. H. Ohlendorf, *Biochemistry* **1997**, *36*, 10039–10051.
43. A. M. Orville; J. D. Lipscomb; D. H. Ohlendorf, *Biochemistry* **1997**, *36*, 10052–10066.
44. T. E. Elgren; A. M. Orville; K. A. Kelly; J. D. Lipscomb; D. H. Ohlendorf; L. Que, Jr., *Biochemistry* **1997**, *36*, 11504–11513.
45. R. W. Frazee; A. M. Orville; K. B. Dolbeare; H. Yu; D. H. Ohlendorf; J. D. Lipscomb, *Biochemistry* **1998**, *37*, 2131–2144.
46. D. D. Cox; L. Que, Jr., *J. Am. Chem. Soc.* **1988**, *110*, 8085–8092.
47. H. G. Jang; D. D. Cox; L. Que, Jr., *J. Am. Chem. Soc.* **1991**, *113*, 9200–9204.
48. R. Yamahara; S. Ogo; H. Masuda; Y. Watanabe, *J. Inorg. Biochem.* **2002**, *88*, 284–294.
49. M. Y. M. Pau; M. I. Davis; A. M. Orville; J. D. Lipscomb; E. I. Solomon, *J. Am. Chem. Soc.* **2007**, *129*, 1944–1958.
50. R. J. Mayer; L. Que, Jr., *J. Biol. Chem.* **1984**, *259*, 13056–13060.
51. G. Avigad; S. England, *Fed. Proc.* **1969**, *28*, 345–346.
52. R. V. J. Chari; C. P. Whitman; J. W. Kozarich; K. L. Ngai; L. N. Ornston, *J. Am. Chem. Soc.* **1987**, *109*, 5514–5519.
53. R. Hill; G. W. Kirby; G. J. O'Loughlin; D. J. Robins, *J. Chem. Soc., Perkin Trans. 1,* **1993**, 1967–1993.
54. R. V. J. Chari; C. P. Whitman; J. W. Kozarich; K. L. Ngai; L. N. Ornston, *J. Am. Chem. Soc.* **1987**, *109*, 5520–5521.
55. A. Goldman; D. L. Ollis; T. A. Steitz, *J. Mol. Biol.* **1987**, *194*, 143–153.
56. D. J. Neidhart; G. L. Kenyon; J. A. Gerlt; G. A. Petsko, *Nature* **1990**, *347*, 692–694.
57. S. E. Williams; E. M. Woolridge; S. C. Ransom; J. A. Landro; P. C. Babbitt; J. W. Kozarich, *Biochemistry* **1992**, *31*, 9768–9776.
58. J. Yang; Y. Wang; E. M. Woolridge; V. Arora; G. A. Petsko; J. W. Kozarich; D. Ringe, *Biochemistry* **2004**, *43*, 10424–10434.
59. Y. Kojima; N. Itada; O. Hayaishi, *J. Biol. Chem.* **1961**, *236*, 2223–2230.
60. S. Han; L. D. Eltis; K. N. Timmis; S. W. Muchmore; J. T. Bolin, *Science* **1995**, *270*, 976–980.
61. E. L. Hegg; L. Que, Jr., *Eur. J. Biochem.* **1997**, *250*, 625–629.
62. T. Senda; S. Sugiyama; H. Narita; T. Yamamoto; K. Kimbara; M. Fukuda; M. Sato; K. Yano; Y. Mitsui, *J. Mol. Biol.* **1996**, *255*, 735–752.
63. A. Kita; S. I. Kita; I. Fujisawa; K. Inaka; T. Ishida; K. Horiike; M. Nozaki; K. Miki, *Structure* **1998**, *7*, 25–34.
64. K. Sugimoto; T. Senda; H. Aoshima; E. Masai; M. Fukuda; Y. Mitsui, *Structure* **1999**, *7*, 953–965.
65. Y. Zhang; K. L. Colabroy; T. P. Begley; S. E. Ealick, *Biochemistry* **2005**, *44*, 7632–7643.
66. E. L. Spence; M. Kawamukai; J. Sanvoisin; H. Braven; T. D. H. Bugg, *J. Bacteriol.* **1996**, *178*, 5249–5256.
67. F. H. Vaillancourt; J. T. Bolin; L. D. Eltis, *Crit. Rev. Biochem. Mol. Biol.* **2006**, *41*, 241–267.
68. D. M. Arciero; J. D. Lipscomb, *J. Biol. Chem.* **1986**, *261*, 2170–2178.
69. F. H. Vaillancourt; C. J. Barbosa; T. G. Spiro; J. T. Bolin; M. W. Blades; R. F. B. Turner; L. D. Eltis, *J. Am. Chem. Soc.* **2002**, *124*, 2485–2496.
70. C. Bianchini; F. Frediani; F. Laschi; A. Meli; F. Vizza; P. Zanello, *Inorg. Chem.* **1990**, *29*, 3402–3409.
71. E. L. Spence; G. J. Langley; T. D. H. Bugg, *J. Am. Chem. Soc.* **1996**, *118*, 8336–8343.
72. C. J. Winfield; Z. Al-Mahrizy; M. Gravestock; T. D. H. Bugg, *J. Chem. Soc., Perkin Trans. 1* **2000**, 3277–3289.
73. E. G. Kovaleva; J. D. Lipscomb, *Science* **2007**, *316*, 453–457.
74. T. D. H. Bugg; G. Lin, *Chem. Commun.* **2001**, 941–952.
75. J. Sanvoisin; G. J. Langley; T. D. H. Bugg, *J. Am. Chem. Soc.* **1995**, *117*, 7836–7837.
76. S. L. Groce; J. D. Lipscomb, *J. Am. Chem. Soc.* **2003**, *125*, 11780–11781.
77. J. Schlosrich; K. L. Eley; P. J. Crowley; T. D. H. Bugg, *ChemBiochem* **2006**, *7*, 1899–1908.
78. G. Lin; G. Reid; T. D. H. Bugg, *J. Am. Chem. Soc.* **2001**, *123*, 5030–5039.
79. S. Mendel; A. Arndt; T. D. H. Bugg, *Biochemistry* **2004**, *43*, 13390–13396.
80. N. Sato; Y. Uragami; T. Nishizaki; Y. Takahashi; G. Sazaki; K. Sugimoto; T. Nonaka; E. Masai; M. Fukuda; T. Senda, *J. Mol. Biol.* **2002**, *321*, 621–636.
81. S. L. Groce; J. D. Lipscomb, *Biochemistry* **2005**, *44*, 7175–7188.
82. D.-H. Jo; L. Que, Jr., *Angew. Chem. Intl. Ed.* **2000**, *39*, 4284–4288.
83. P. C. A. Bruijnincx; M. Lutz; A. L. Spek; W. R. Hagen; B. M. Weckhuysen; G. van Koten; R. J. M. Klein Gebbink, *J. Am. Chem. Soc.* **2007**, *129*, 2275–2286.
84. A. K. Whiting; Y. R. Boldt; M. P. Hendrich; L. P. Wackett; L. Que, Jr., *Biochemistry* **1996**, *35*, 160–170.
85. M. R. Harpel; J. D. Lipscomb, *J. Biol. Chem.* **1990**, *265*, 6301–6311.
86. C. P. Whitman; B. A. Aird; W. R. Gillespie; N. J. Stolowich, *J. Am. Chem. Soc.* **1991**, *113*, 3154–3162.
87. C. P. Whitman; G. Hajipour; R. J. Watson; W. H. Johnson, Jr.; M. E. Bembenek; N. J. Stolowich, *J. Am. Chem. Soc.* **1992**, *114*, 10104–10110.
88. H. S. Subramanya; D. I. Roper; Z. Dauter; E. J. Dodson; G. J. Davies; K. S. Wilson; D. B. Wigley, *Biochemistry* **1996**, *35*, 792–802.
89. J. T. Stivers; C. Abeygunawardana; A. S. Mildvan; G. Hajipour; C. P. Whitman; L. H. Chen, *Biochemistry* **1996**, *35*, 803–813.
90. T. K. Harris; R. M. Czerwinski; W. H. Johnson, Jr.; P. M. Legler; C. Abeygunawardana; M. A. Massiah; J. T. Stivers; C. P. Whitman; A. S. Mildvan, *Biochemistry* **1999**, *38*, 12343–12357.
91. N. Metanis; A. Brik; P. E. Dawson; E. Keinan, *J. Am. Chem. Soc.* **2004**, *126*, 12726–12727.
92. H. Lian; C. P. Whitman, *J. Am. Chem. Soc.* **1994**, *116*, 10403–10411.
93. G. Dunn; M. G. Montgomery; F. Mohammed; A. Coker; J. B. Cooper; T. Robertson; J.-L. Garcia; T. D. H. Bugg; S. P. Wood, *J. Mol. Biol.* **2005**, *346*, 253–265.

94. N. Nandhagopal; A. Yamada; T. Hatta; E. Masai; M. Fukuda; Y. Mitsui; T. Senda, *J. Mol. Biol.* **2001**, *309*, 1139–1151.
95. S. Fushinobu; T. Saku; M. Hidaka; S.-Y. Jun; H. Nojiri; H. Yamane; H. Shoun; T. Omori; T. Wakagi, *Protein Sci.* **2002**, *11*, 2184–2195.
96. H. Habe; K. Morii; S. Fushinobu; J.-W. Nam; Y. Ayabe; T. Yoshida; T. Wakagi; H. Yamane; H. Nojiri; T. Omori, *Biochem. Biophys. Res. Commun.* **2003**, *303*, 631–639.
97. W. W. Y. Lam; T. D. H. Bugg, *Biochemistry* **1997**, *36*, 12242–12251.
98. I. M. J. Henderson; T. D. H. Bugg, *Biochemistry* **1997**, *36*, 12252–12258.
99. S. M. Fleming; T. A. Robertson; G. J. Langley; T. D. H. Bugg, *Biochemistry* **2000**, *39*, 1522–1531.
100. J.-J. Li; C. Li; C. A. Blindauer; T. D. H. Bugg, *Biochemistry* **2006**, *45*, 12461–12469.
101. D. M. Speare; S. M. Fleming; M. N. Beckett; J.-J. Li; T. D. H. Bugg, *Org. Biomol. Chem.* **2004**, *2*, 2942–2950.
102. C. Li; M. G. Montgomery; F. Mohammed; J.-J. Li; S. P. Wood; T. D. H. Bugg, *J. Mol. Biol.* **2005**, *346*, 241–251.
103. C. Li; J.-J. Li; M. G. Montgomery; S. P. Wood; T. D. H. Bugg, *Biochemistry* **2006**, *45*, 12470–12479.
104. J.-J. Li; T. D. H. Bugg, *Org. Biomol. Chem.* **2007**, *5*, 507–513.
105. J. R. Pollard; I. M. J. Henderson; T. D. H. Bugg, *Chem. Commun.* **1997**, 1885–1886.
106. J. R. Pollard; T. D. H. Bugg, *Eur. J. Biochem.* **1998**, *251*, 98–106.
107. A. Izumi; D. Rea; T. Adachi; S. Unzai; S. Y. Park; D. I. Roper; J. R. H. Tame, *J. Mol. Biol.* **2007**, *370*, 899–911.
108. J. R. Pollard; D. Rialland; T. D. H. Bugg, *Appl. Environ. Microbiol.* **1998**, *64*, 4093–4094.
109. S. J. Lee; J. H. Ko; H. Y. Kang; Y. Lee, *Biochem. Biophys. Res. Commun.* **2006**, *346*, 1009–1015.
110. D. Rea; V. Fulop; T. D. H. Bugg; D. I. Roper, *J. Mol. Biol.* **2007**, *373*, 866–876.
111. W. J. Wang; S. Y. K. Seah, *Biochemistry* **2005**, *44*, 9447–9455.
112. B. Lindblad; S. Lindstedt; G. Steen, *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 4641–4645.
113. L. Serre; A. Sailland; D. Sy; P. Boudec; A. Rolland; E. Pebay-Peyroula; C. Cohen-Addad, *Structure* **1999**, *7*, 977–988.
114. J. M. Brownlee; K. Johnson-Winters; D. H. T. Harrieson; G. R. Moran, *Biochemistry* **2004**, *43*, 6370–6377.
115. L. Stryer, *Biochemistry*, 4th ed.; Freeman: New York, 1999, Chapter 25.
116. A. E. Garrod, *Lancet* **1902**, *2*, 1616–1620.
117. G. P. Titus; H. A. Mueller; J. Burgner; S. Rodriguez de Cordoba; M. A. Penalva; D. E. Timm, *Nat. Struct. Biol.* **2000**, *7*, 542–546.
118. L. Lack, *J. Biol. Chem.* **1961**, *236*, 2835–2840.
119. D. Pokorny; W. Steiner; D. W. Ribbons, *Trends Biotechnol.* **1997**, *15*, 291–296.
120. M.St. Loius; R. M. Tanguay, *Hum. Mutat.* **1997**, *9*, 291–299.
121. D. E. Timm; H. A. Mueller; P. Bhanumoorthy; J. M. Harp; G. J. Bunick, *Structure* **1999**, *7*, 1023–1033.
122. R. L. Bateman; P. Bhanumoorthy; J. F. Witte; R. W. McClard; M. Grompe; D. E. Timm, *J. Biol. Chem.* **2001**, *276*, 15284–15291.
123. S. L. Fuenmayor; M. Wild; A. L. Boyes; P. A. Williams, *J. Bacteriol.* **1998**, *180*, 2522–2530.
124. N.-Y. Zhou; S. L. Fuenmayor; P. A. Williams, *J. Bacteriol.* **2001**, *183*, 700–708.
125. M. Sono; M. P. Roach; E. D. Coulter; J. H. Dawson, *Chem. Rev.* **1996**, *96*, 2841–2888.
126. H. Sugimoto; S.-I. Oda; T. Otsuki; T. Hino; T. Yoshida; Y. Shiro, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 2611–2616.
127. D. H. Munn; M. Zhou; J. T. Attwood; I. Bondarev; S. J. Conway; B. Marshall; C. Brown; A. L. Mellor, *Science* **1998**, *281*, 1191–1193.
128. D. H. Munn; J. Pressey; A. C. Beall; R. Hudes; M. R. Alderson, *J. Immunol.* **1996**, *156*, 523–532.
129. O. Hayaishi; O. Takikawa; R. Yoshida, *Prog. Inorg. Chem.* **1990**, *38*, 75–95.
130. S. Christen; E. Peterhans; R. Stocker, *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 3998–4006.
131. C. Momany; V. Levdikov; L. Blagova; S. Lima; R. S. Phillips, *Biochemistry* **2004**, *43*, 1193–1203.
132. S. Lima; R. Khristoforov; C. Momany; R. S. Phillips, *Biochemistry* **2007**, *46*, 2735–2744.
133. R. S. Phillips; S. Sundararaju; S. Koushik, *Biochemistry* **1998**, *37*, 8783–8789.
134. J. Breton; N. Avanzi; S. Magagnin; N. Covini; G. Magistrelli; L. Cozzi; A. Isacchi, *Eur. J. Biochem.* **2000**, *267*, 1092–1099.
135. P. Malherbe; C. Kohler; M. Da Prada; G. Lang; V. Kiefer; R. Schwarcz; H. W. Lahm; A. M. Cesura, *J. Biol. Chem.* **1994**, *269*, 13792–13799.
136. J. L. Walsh; W. P. Todd; B. K. Carpenter; R. Schwarcz, *Biochem. Pharmacol.* **1991**, *42*, 985–990.
137. P. Kurnasov; V. Goral; K. Colabroy; S. Gerdes; S. Anantha; A. Osterman; T. P. Begley, *Chem. Biol.* **2003**, *10*, 1195–1204.
138. Y. Zhang; K. L. Colabroy; T. P. Begley; S. E. Ealick, *Biochemistry* **2005**, *44*, 7632–7643.
139. K. L. Colabroy; H. L. Zhai; T. F. Li; Y. Ge; Y. Zhang; A. M. Liu; S. E. Ealick; F. W. McLafferty; T. P. Begley, *Biochemistry* **2005**, *44*, 7623–7631.
140. J. A. Olson; O. Hayaishi, *Proc. Natl. Acad. Sci. U.S.A.* **1965**, *54*, 1364–1370.
141. A. Wyss; G. Wirtz; W. D. Woggon; R. Brugger; M. Wyss; A. Friedlein; H. Bachmann; W. Hunziker, *Biochem. Biophys. Res. Commun.* **2000**, *271*, 334–336.
142. J. von Lintig; K. Vogt, *J. Biol. Chem.* **2000**, *275*, 11915–11920.
143. J. Paik; A. During; E. H. Harrison; C. L. Mendelsohn; K. Lai; W. S. Blaner, *J. Biol. Chem.* **2001**, *276*, 32160–32168.
144. W. D. Woggon, *Pure Appl. Chem.* **2002**, *74*, 1397–1408.
145. G. M. Wirtz; C. Bornemann; A. Giger; R. K. Muller; H. Schneider; G. Schlotterbeck; G. Schiefer; W. D. Woggon, *Helv. Chim. Acta* **2001**, *84*, 2301–2315.
146. S. H. Schwartz; B. C. Tan; D. A. Gage; J. A. Zeevaart; D. R. McCarty, *Science* **1997**, *276*, 1872–1874.
147. B. C. Tan; S. H. Schwartz; J. A. Zeevaart; D. R. McCarty, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12235–12240.
148. B. C. Tan; L. M. Joseph; W. T. Deng; L. Liu; Q. B. Li; K. Cline; D. R. McCarty, *Plant J.* **2003**, *35*, 44–56.
149. G. Giuliano; S. Al-Babili; J. von Lintig, *Trends Plant Sci.* **2003**, *8*, 145–149.
150. M. E. Auldridge; D. R. McCarty; H. J. Klee, *Curr. Opin. Plant Biol.* **2006**, *9*, 315–321.
151. S. H. Schwartz; X. Qin; J. A. Zeevaart, *J. Biol. Chem.* **2001**, *276*, 25208–25211.
152. P. Karrer; K. Miki, *Helv. Chim. Acta* **1929**, *12*, 985–989.
153. F. Bouvier; J.-C. Isner; O. Dogbo; B. Camara, *Trends Plant Sci.* **2005**, *10*, 187–191.
154. K. Sorefan; J. Booker; K. Haurogne; M. Goussot; K. Baibridge; E. Foo; S. Chatfield; S. Ward; C. Beveridge; C. Rameau; O. Leyser, *Gene Dev.* **2003**, *17*, 1467–1474.
155. J. Booker; M. Auldridge; S. Wills; H. Klee; O. Leyser, *Curr. Biol.* **2004**, *14*, 1–20.

156. D. P. Kloer; S. Ruch; S. Al-Babili; P. Beyer; G. E. Schulz, *Science* **2005**, *308*, 267–269.
157. S. Y. Han; N. Kitahata; T. Saito; M. Kobayashi; K. Shinozaki; S. Yoshida; T. Asami, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3033–3036.
158. J. J. Hutton; A. L. Tappel; S. Udenfriend, *Arch. Biochem. Biophys.* **1967**, *118*, 231–238.
159. G. J. Cardinale; R. E. Rhoads; S. Udenfriend, *Biochem. Biophys. Res. Commun.* **1971**, *43*, 537–543.
160. A. G. Prescott; M. D. Lloyd, *Nat. Prod. Rep.* **2000**, *17*, 367–383.
161. C. J. Schofield; Z. Zhang, *Curr. Opin. Struct. Biol.* **1999**, *9*, 722–731.
162. M. Ryle; R. P. Hausinger, *Curr. Opin. Chem. Biol.* **2002**, *6*, 193–201.
163. Z. Zhang; J. Ren; D. K. Stammers; J. E. Baldwin; K. Harlos; C. J. Schofield, *Nat. Struct. Biol.* **2000**, *7*, 127–133.
164. B. Siegel, *Bioorg. Chem.* **1979**, *8*, 2129–2226.
165. H. M. Hanauskeabel; V. Gunzler, *J. Theor. Biol.* **1982**, *94*, 421–455.
166. J. E. Baldwin; R. W. Adlington; N. P. Crouch; I. A. C. Pereira; R. T. Aplin; C. Robinson, *J. Chem. Soc., Chem. Commun.* **1993**, 105–106.
167. B. Lindblad; G. Lindstedt; S. Lindstedt, *J. Am. Chem. Soc.* **1970**, *92*, 7446–7449.
168. P. J. Sabourin; L. L. Bieber, *J. Biol. Chem.* **1982**, *257*, 7468–7471.
169. Y. Kikuchi; Y. Suzuki; N. Tamiya, *Biochem. J.* **1983**, *213*, 507–512.
170. M. Wu; T. P. Begley; J. Myllyharju; K. I. Kivirikko, *Bioorg. Chem.* **2000**, *28*, 261–265.
171. M. Wu; H. S. Moon; T. P. Begley; J. Myllyharju; K. I. Kivirikko, *J. Am. Chem. Soc.* **1999**, *121*, 587–588.
172. D. A. Proshlyakov; T. F. Henshaw; G. R. Monterosso; M. J. Ryle; R. P. Hausinger, *J. Am. Chem. Soc.* **2004**, *126*, 1022–1023.
173. L. M. Hoffart; E. W. Barr; R. B. Guyer; J. M. Bollinger; C. Krebs, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 14738–14743.
174. O. W. Choroba; D. H. Williams; J. B. Spencer, *J. Am. Chem. Soc.* **2000**, *122*, 5389–5390.
175. B. K. Hubbard; M. G. Thomas; C. T. Walsh, *Chem. Biol.* **2000**, *7*, 931–942.
176. P. Jaakola; D. R. Mole; Y.-M. Tian; M. I. Wilson; J. Gielbert; S. J. Gaskell; A. von Kreigsheim; H. F. Hebestreit; M. Mukherji; C. J. Schofield; P. H. Maxwell; C. W. Pugh; P. J. Ratcliffe, *Science* **2001**, *292*, 468–472.
177. J. M. Elkins; K. S. Hewitson; L. A. McNeill; J. F. Seibel; I. Schlemminger; C. W. Pugh; P. J. Ratcliffe; C. J. Schofield, *J. Biol. Chem.* **2003**, *278*, 1802–1806.
178. T. Oka; F. J. Simpson, *Biochem. Biophys. Res. Commun.* **1971**, *43*, 1–5.
179. H. K. Hund; J. Breuer; F. Lingens; J. Hutterman; R. Kappl; S. Fetzner, *Eur. J. Biochem.* **1999**, *263*, 871–878.
180. R. A. Steiner; K. H. Kalk; B. W. Dijkstra, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16625–16630.
181. S. Fetzner, *Appl. Microbiol. Biotechnol.* **2002**, *60*, 243–257.
182. I. Bauer; N. Max; S. Fetzner; F. Lingens, *Eur. J. Biochem.* **1996**, *240*, 576–583.
183. F. Fischer; S. Künne; S. Fetzner, *J. Bacteriol.* **1999**, *181*, 5725–5733.
184. F. Fischer; S. Fetzner, *FEMS Microbiol. Lett.* **2000**, *190*, 21–27.
185. T. D. H. Bugg; C. J. Winfield, *Nat. Prod. Rep.* **1998**, *15*, 513–530.
186. G. Impellizzeri; M. Piattelli, *Phytochemistry* **1972**, *11*, 2499–2502.
187. P. A. Girod; J. P. Zryd, *Phytochemistry* **1991**, *30*, 169–174.
188. L. A. Mueller; U. Hinz; J. P. Zyrd, *Phytochemistry* **1997**, *44*, 567–569.
189. L. Christinet; F. R. X. Burdet; M. Zaiko; U. Hinz; J. P. Zyrd, *Plant Physiol.* **2004**, *134*, 265–274.
190. J. W. Wray; R. H. Abeles, *J. Biol. Chem.* **1993**, *268*, 21466–21469.
191. R. Myers; J. W. Wray; S. Fish; R. H. Abeles, *J. Biol. Chem.* **1993**, *268*, 24785–24794.
192. J. W. Wray; R. H. Abeles, *J. Biol. Chem.* **1995**, *270*, 3147–3153.
193. Y. Dai; P. C. Wensink; R. H. Abeles, *J. Biol. Chem.* **1999**, *274*, 1193–1195.
194. Y. Dai; T. C. Pochapsky; R. H. Abeles, *Biochemistry* **2001**, *40*, 6379–6387.
195. T. C. Pochapsky; S. S. Pochapsky; T. Ju; H. Mo; F. Al-Mjeni; M. Maroney, *Nat. Struct. Biol.* **2002**, *9*, 966–972.
196. G. Straganz; L. Brecker; H.-J. Weber; W. Steiner; D. W. Ribbons, *Biochem. Biophys. Res. Commun.* **2002**, *297*, 232–236.
197. G. D. Straganz; H. Hofer; W. Steiner; B. Nidetzky, *J. Am. Chem. Soc.* **2004**, *126*, 12202–12203.

Biographical Sketch

Timothy D. H. Bugg (born 1965) is professor of biological chemistry at the University of Warwick. Following his Ph.D. studies with Dr. C. Abell at the University of Cambridge, he spent 2 years as a SERC/NATO postdoctoral research fellow in the laboratory of Professor

C. T. Walsh at Harvard Medical School. In 1991 he began his academic career as a lecturer in organic chemistry at the University of Southampton, before moving to Warwick in 1999. His research interests are in the study of enzyme mechanisms, principally enzymes involved in the bacterial degradation of aromatic compounds, and enzymes involved in bacterial peptidoglycan biosynthesis. He enjoys playing and coaching volleyball, cycling, and playing the violin.

8.17 *S*-Adenosylmethionine and Iron–Sulfur Clusters in Biological Radical Reactions: The Radical SAM Superfamily

Eric M. Shepard and Joan B. Broderick, Montana State University, Bozeman, MT, USA

8.17.1   Introduction

8.17.1.1   Iron–Sulfur Clusters and the Origins of Protein-Based Biocatalysts

Iron–sulfur clusters are ubiquitous in modern biological systems, serving multiple purposes including protein structural stabilization, electron transfer, substrate binding and activation, iron storage, donation of sulfide, and regulation of gene expression.[1,2] The structures of these clusters are diverse, ranging from simple [2Fe–2S] diamonds and [4Fe–4S] distorted cubes to the more complex structures found in enzymes such as nitrogenase and hydrogenase; in all cases, these clusters have structural features that resemble naturally occurring inorganic iron sulfide minerals. The structural similarities between biological and inorganic iron sulfide minerals, together with the preponderance of Fe and S in the earth's mineral record, have led many investigators to suggest that simple, abiotic precursor compounds comprised of Fe–S minerals may have acted as some of the earliest catalysts on the primordial earth. Moreover, in a 'metabolism-first' theory on the origin of life, Fe–S-containing compounds could have acted as the basic building blocks of life in the transition from an abiotic to a biotic world via the gradual conversion from mineral-based catalysts to protein-based biocatalysts.[3] Eck and Dayhoff[4] laid a solid foundation for the beginnings of biochemical evolution by hypothesizing that the ancestor of the ferredoxin protein, a short polypeptide comprised of some of the simplest amino acids, was a prototype for metabolism prior to the evolution of the genetic code and protein complexity.

Russell and coworkers have argued that alkaline pH and the development of chemical gradients are two events that may have been critical to the origin of life.[5] In their proposal, mineral complexes such as mackinawite, consisting of a series of [2Fe–2S] rhombs, may have acted as a hydrogenase and an electron transfer agent through the membrane of the hydrothermal mound.[6] Moreover, Wächtershäuser[7,8] has proposed a chemoautotrophic origin of life known as the 'iron–sulfur world' (ISW) based on the premise that prebiotic reactions were carried out by iron–sulfur minerals located near deep sea hydrothermal vents. Huber and Wächtershäuser have shown that Fe/Ni–S phases generate acetic acid through a carbon fixation route under presumed primordial conditions,[9] can activate amino acids to generate peptides,[10] and can even form CO-dependent α-hydroxy and α-amino acids.[11] The observation that basic metal scaffolds have the ability to carry out reactions that are believed to have occurred on early earth is evidence in support of the premise that modern-day protein-based catalysis has its origins in mineral-based catalysis.

Under reducing conditions, simple iron–sulfur clusters can spontaneously assemble from Fe^{2+} and S^{2-} in aqueous solutions. The fundamental unit of iron–sulfur clusters is a [2Fe–2S] rhomb with Fe–S bonds of approximately 2.5 Å and Fe–S–Fe and S–Fe–S bond angles of approximately 75 and 105°, respectively. The rhombs themselves exist as distorted planar structures with each individual iron in an ideal or near-ideal tetrahedral environment. The simplest and most common types of biological Fe–S clusters are [2Fe–2S], [3Fe–4S], and [4Fe–4S] (**Figure 1**), with the iron atoms ligated by protein-derived cysteine thiols and inorganic sulfide ions in tetrahedral geometry, although histidine and arginine ligation has also been observed.[12,13] This environment is ideally suited for the most common function of Fe–S clusters, namely electron transfer, with a large range of reduction potentials (from <-400 to >400 mV) depending on the protein microenvironment and/or coordination environment of the clusters.[14]

Protein-bound Fe–S clusters display a remarkable ability to interconvert with one another, and undergo ligand exchange reactions and oxidative degradation.[1,15] These inherent properties coupled with their ability to serve as 'compact redox catalysts in the range of low potentials' may very well point toward their ancient beginnings as biocatalysts.[1] This chemical versatility may help to explain why more and more examples of distinctive Fe–S clusters are being discovered in modern biological systems. Examples of distinctive Fe–S clusters include metal centers that coordinate substrates via a unique Fe site (members of the radical *S*-adenosylmethionine (SAM or AdoMet) superfamily, aconitase, and others), enzymes that contain Fe–S heteroatomic clusters (nitrogenase iron–molybdenum cofactor (FeMoco), carbon monoxide dehydrogenase (CODH), and acetyl CoA synthase (ACS)), and enzymes that contain unique ligation sets around specialized iron centers ([NiFe] and [FeFe] hydrogenases) (**Figure 1**).[1,16]

In terms of chemical reactions that were likely commonplace on early earth, it is plausible that modified iron–sulfur minerals acted as catalysts to bring about reactions involving H_2 oxidation, N_2 reduction, and
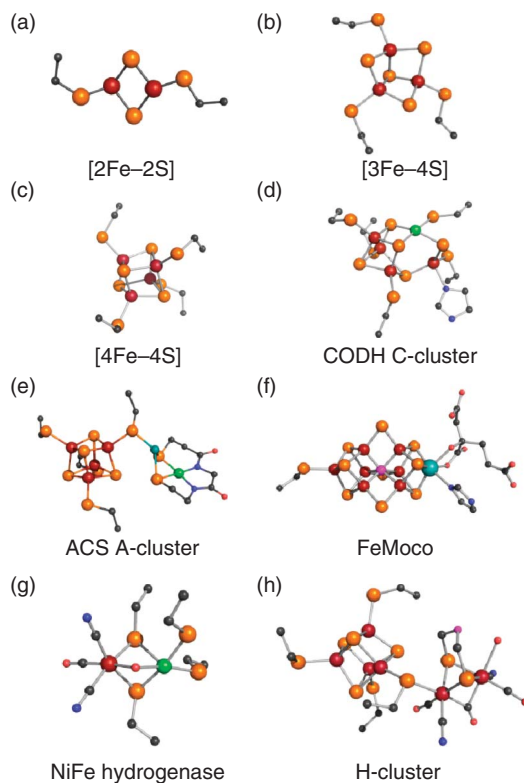
Figure 1 Iron–sulfur clusters observed in proteins. Color coding for the atoms depicted in [Fe–S] clusters is as follows: maroon (iron), orange (sulfur), black (carbon), blue (nitrogen), red (oxygen), pink (unidentified or unknown), and green (nickel). In the ACS A-cluster, cyan represents copper, whereas in FeMoco the cyan sphere represents molybdenum.

the interconversion of CO and CO_2. Presently, these reactions are carried out by complex iron–sulfur (Fe–S)-containing enzymes that have been evolutionarily tuned to carry out their respective functions via modifications of basic Fe–S clusters (**Figure 1**). It has been demonstrated that sulfur vacancies present on iron pyrite act to increase not only the retention time of adsorbed amino acids at the FeS_2–H_2O interface, but also the reactivity of the iron and sulfur atoms at the defect site.[17] The complex metallocofactors presently observed in CODH (**Figure 1(d)**), ACS (**Figure 1(e)**), the Fe–Mo cofactor of nitrogenase (**Figure 1(f)**), and the [NiFe] and [FeFe] hydrogenases (**Figures 1(g) and 1(h)**) can be thought of as highly specialized 'defect' sites. Current research in the metallobiochemistry of such clusters focuses on understanding not only why the modifications associated with the so-called defect site are crucial to the chemical reaction carried out by the enzyme, but also how these modifications arise during metal cluster assembly and why these clusters require a protein environment to carry out their respective chemical reactions. Experimental insights should provide an outline of the steps accompanying life's transition from abiotic precursor compounds to the biotic complexity currently observed.

For the most common types of biological clusters ([2Fe–2S], [3Fe–4S], and [4Fe–4S]), the assembly machinery includes iron chaperones, cysteine desulfurases, electron transfer proteins, molecular chaperones, and scaffold proteins on which the nascent cluster is assembled prior to insertion into a target protein.[14] For the more complex and unusual clusters, such as FeMoco of nitrogenase, the [NiFe] center of hydrogenase, or the H-cluster of [FeFe] hydrogenase, significantly less is known about the cluster assembly process. As we shall discover below, one common theme that these systems share in the synthesis of their respective metallocofactors is the involvement of radical chemistry provided by radical SAM enzymes.

8.17.1.2   Radical SAM Enzymes

8.17.1.2.1   A connection to early biological catalysts?

Pyruvate formate lyase (PFL) is an essential enzyme in bacterial anaerobic metabolism and is responsible for converting pyruvate and coenzyme A (CoA) into formate and acetyl CoA. PFL is likely an ancient enzyme, evolving under the growing complexity of intermediary metabolism in an anaerobic world. PFL shares high structural and functional similarity with the class III anaerobic ribonucleotide reductase (ARR), responsible for ribonucleotide reduction, with both enzymes utilizing a glycyl radical cofactor in catalysis.[18] It has been suggested that RNA preceded DNA during the evolution of life, and with the appearance of an enzyme (ARR) capable of reducing ribonucleotides, DNA could replace RNA as the storage vesicle for genetic information. Currently, ribonucleotide reductases are found in all organisms and are essential for DNA replication and repair, providing the only means for the conversion of nucleotides into deoxynucleotides.[19] Both ARR and PFL require the generation of glycyl radicals in order to carry out their respective reactions, and glycyl radicals in both systems are generated through an activating enzyme utilizing a redox-active [4Fe–4S] cluster and SAM. These radical reactions require a protein environment to shield the radical from solvent, and this suggests that SAM radical-based chemistry may very well have been one of the early functions associated with protein-based biocatalysts.

8.17.1.2.2   Before the superfamily: Early studies of radical SAM enzymes

Although the radical SAM enzymes were first identified as a superfamily in 2001, the seeds for identifying this superfamily were sown in the preceding 30+ years, as novel enzymatic reactions that were dependent on SAM and iron were elucidated. In work emanating from the laboratory of Joachim Knappe in the late 1960s and early 1970s, it was shown that PFL could be isolated in an inactive form and then activated by another protein (then known as 'enzyme II', now the PFL-activating enzyme (PFL-AE)) in a reaction dependent on flavodoxin, SAM, and iron.[20–25] This activation was subsequently demonstrated to involve the generation of a stable glycyl radical on PFL,[26] thus providing the first evidence for a link between SAM-dependent enzymes and biological radical reactions. When ultimately purified in 1984, the PFL-AE was found to contain a visible chromophore; this unidentified chromophore was probably a result of partial binding of the iron–sulfur cluster later identified as essential to this enzyme's activity.[27] The amino acid sequence of the PFL-AE was first published in 1988, and at that time the unusual cluster of three cysteines in the CX_3CX_2C motif, and the possibility that these might be involved in iron binding, was noted.[28] Subsequent publications reported that PFL-AE bound one equivalent of Fe(II), although it was later demonstrated to contain an iron–sulfur cluster.[29,30] In the early 1990s, a similar type of SAM-dependent activating enzyme system was reported to be required for the activation of the ARR,[31] and in subsequent years an iron–sulfur cluster in this enzyme was demonstrated.[32]

Also in the early 1970s, investigations by Barker and coworkers on *Clostridium subterminale*, which is capable of fermenting lysine to acetate, butyrate, and ammonia, revealed that lysine 2,3-aminomutase (LAM) catalyzed the first step of fermentation, the conversion of L-lysine to L-β-lysine, in a reaction dependent on iron, SAM, reducing conditions, and pyridoxal-5′-phosphate (PLP).[33] The isomerization reaction proceeded with the α-amino group of lysine migrating to the β-carbon in β-lysine, a reaction analogous to those catalyzed by adenosylcobalamin-dependent enzymes. In the late 1980s, Frey and coworkers provided the first evidence via label transfer studies that LAM utilized SAM as a precursor to an intermediate 5′-deoxyadenosyl radical, and referred to the apparent parallels to analogous rearrangement reactions catalyzed by adenosylcobalamin (B$_{12}$)-dependent enzymes.[34,35] The enzymatic activity of LAM was subsequently shown to correlate with the iron–sulfur cluster content, pointing to a central role of the iron–sulfur cluster in catalysis.[36] In the early 1990s, a substrate radical intermediate in the LAM-catalyzed reaction was detected and shown to be kinetically competent,[37–39] providing an important new clue that iron–sulfur clusters and SAM were working together in this enzyme to initiate radical-based catalysis reminiscent of the B$_{12}$-dependent aminomutases.

Biotin synthase (BioB) and lipoate synthase were two other enzymes that, prior to the identification of the superfamily, were shown to contain iron–sulfur clusters and to utilize SAM for catalyzing the radical-mediated insertion of sulfur into unactivated C–H bonds.[40–46]

8.17.1.2.3 Radical SAM enzymes: Nature's radical catalysts

The identification of the radical SAM superfamily was made through the use of bioinformatic tools to examine sequence homology, and resulted in the detection of 592 proteins that shared an unusual N-terminal consensus motif comprised of CX_3CX_2C and were predicted to share a common fold for the core domain.[47] Based on the biochemical characterization of a select few members of this family, the CX_3CX_2C motif was predicted to bind a unique Fe–S cluster responsible for the reductive cleavage of SAM into methionine and a highly reactive deoxyadenosyl radical intermediate (**Figure 2**). As we shall see, the deoxyadenosyl radical is responsible for initiation of the chemical reactions catalyzed by this class of enzymes through the abstraction of a hydrogen atom from a substrate molecule, thereby generating a substrate radical that is involved in subsequent radical reactions. Radical SAM enzymes catalyze a vast array of reactions including the introduction of protein-based glycyl radicals, the insertion of sulfur into unreactive carbon–hydrogen bonds, the internal transfer of hydrogen atoms, and DNA repair.

The critical functionality of radical SAM enzymes arises through the generation of the 5′-deoxyadenosyl radical intermediate. Nature has exploited this chemical reactivity in one other system, that of the radical B_{12}-dependent enzymes, which utilize 5′-deoxyadenosylcobalamin (AdoCbl), also known as coenzyme B_{12}, to catalyze isomerization reactions (**Figure 3**).[48] Homolytic cleavage of the cobalt(III)–deoxyadenosine bond results in a cobalt(II) center and the 5′-deoxyadenosyl radical, which generates a substrate-derived radical via the abstraction of a hydrogen atom from the substrate molecule, just as it does in the $[4Fe–4S]^+$ reductive
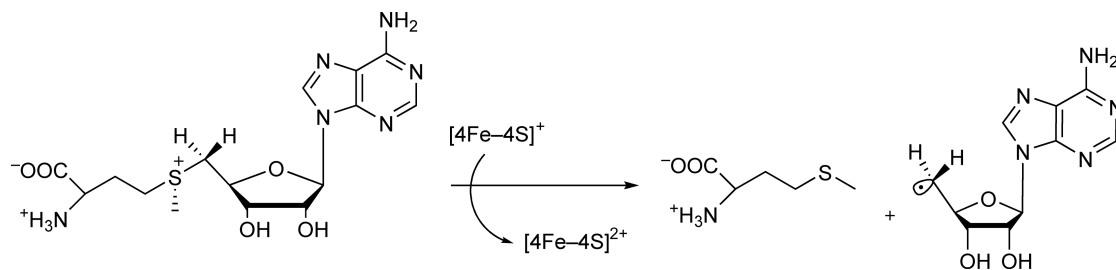
Figure 2 The reductive cleavage of SAM by a $[4Fe–4S]^+$ cluster, forming methionine and the 5′-deoxyadenosyl radical.
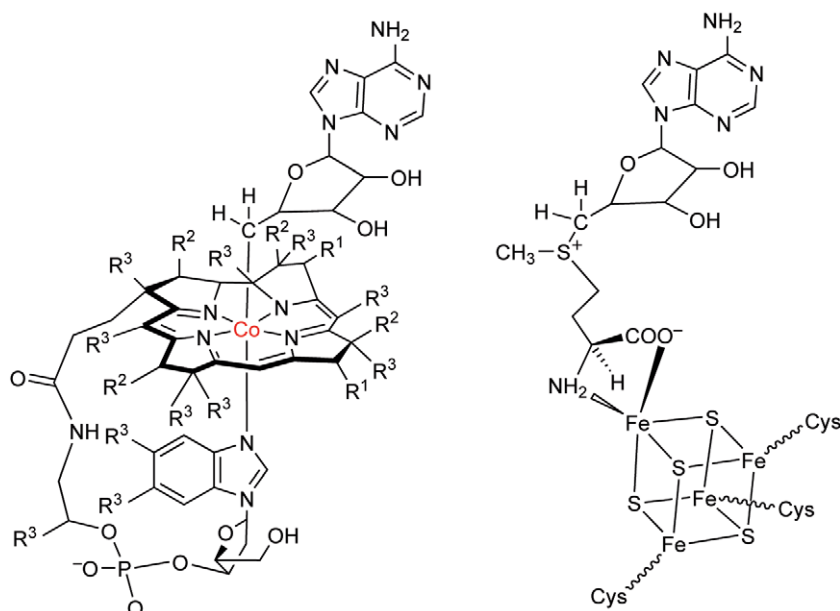
Figure 3 Adenosylcobalamin (left) and SAM coordinated to a [4Fe–4S] cluster (right). The chemical nature of the R-groups decorating the corrin ring of adenosylcobalamin are CH_2CH_2CONH (R^1), CH_2CONH (R^2), and CH_3 (R^3).

cleavage of SAM in radical SAM enzymes. Another similarity between these two classes of enzymes relates to the cofactor binding domain, which for the AdoCbl enzymes methylmalonyl-CoA mutase and glutamate mutase consists of an $(\alpha/\beta)_8$ triose phosphate isomerase (TIM) barrel fold. However, one difference between these classes of enzymes is the chemical diversity of the reactions in which the 5′-deoxyadenosyl radicals are involved. For the AdoCbl class, the 5′-deoxyadenosyl radical generated through cleavage of the cobalt–carbon bond is responsible for initiating the reduction of ribonucleotides and catalyzing difficult 1,2 interchange reactions involving a hydrogen atom and an R group, which can be a carbon skeleton fragment, an alcohol, or an amine group. In contrast, the reductive cleavage of SAM in radical SAM enzymes initiates a broad range of downstream processes from protein-based radical generation, to sulfur insertion, to DNA repair, among others, as will be discussed further in a later section of this chapter. Indeed, though both Baker and Stadtman[49] and Frey[50] once referred to SAM as a 'poor man's adenosylcobalamin' due to its simpler chemical structure than adenosylcobalamin coupled with similar reactivity, the extraordinary versatility of SAM in initiating a wide array of biochemical transformations led Frey to redub SAM as a 'rich man's adenosylcobalamin'.[51]

8.17.2   Radical SAM Enzymes: A Common Mechanistic Start

Despite the prevalence of the radical SAM proteins, until recently little was understood regarding the mechanism(s) by which these reactions proceed. Early work, including label transfer studies on PFL-AE and LAM, implicated an AdoMet-derived 5′-deoxyadenosyl radical as an intermediate in catalysis.[52,53] The 5′-deoxyadenosyl radical, generated by reductive cleavage of SAM (**Figure 2**), is now thought to be a common intermediate responsible for H-atom abstraction from substrates in the diverse reactions catalyzed by the radical SAM superfamily.[47,51,54–57] Once reductive cleavage of AdoMet occurs, the resulting 5′-deoxyadenosyl radical (dAdo·) can initiate a remarkably diverse set of reactions, as exemplified by those shown in **Figure 4**. The generation of a 5′-deoxyadenosyl radical intermediate from AdoMet is novel chemistry; prior to the discovery of its role in LAM[34] and PFL-AE,[26,52] AdoMet was known primarily for its central role in biological methylation reactions.[58] Based on the widespread occurrence of the radical SAM proteins, however, it is clear that initiation of radical reactions is another significant role of AdoMet in biology. This is also novel chemistry for iron–sulfur clusters. Although iron–sulfur clusters are commonly involved in biological electron transfer and associated redox reactions, unusual chemistry was predicted for the iron–sulfur clusters in the radical SAM enzymes based on the poor match of redox potentials for long-range electron transfer from cluster to AdoMet: protein-bound [4Fe–4S] clusters typically have redox potentials in the range of −400 to −600 mV,[1,59] while sulfoniums similar to AdoMet exhibit reduction potentials more negative than −1 V.[60] It was therefore predicted that the role of the iron–sulfur cluster was not simply to reduce AdoMet by long-range electron transfer, but rather to be intimately involved in the chemistry of reductive cleavage of the S–C bond.[61]

8.17.2.1   S-Adenosylmethionine as a Cofactor in Radical Reactions

SAM plays a number of important roles in living organisms, including transmethylation, trans-sulfuration, and polyamine biosynthesis.[58,62] Methyl group donation is the most recognized function of this molecule and involves heterolytic cleavage of the S–C bond such that a methyl cation is the donated species. In the radical SAM enzymes, however, evidence points to homolytic S–C bond cleavage in SAM as a source of 5′-deoxyadenosyl radicals, which perform the key hydrogen atom abstraction noted above. This role of SAM was completely unprecedented at the time it was first proposed.[26,34] In fact, at the time, only adenosylcobalamin (B_{12}) was known to generate 5′-deoxyadenosyl radicals in biology. The B_{12} cofactor, with its relatively weak Co–C bond, is believed to undergo Co–C bond homolysis to produce cob(II)alamin and an adenosyl radical, with the energy for homolysis provided, in part, by the binding interactions of the cofactor and substrate to the enzyme.[63,64] How a relatively simple molecule such as SAM, which has only the adenosyl moiety in common with B_{12} (**Figure 3**), could produce the same highly reactive primary radical intermediate was a perplexing question. Because both LAM and PFL-AE, two of the earliest studied radical SAM enzymes, were reported to require metals for catalysis,[27,33,36,65] significant speculation surrounded the possibility of organometallic
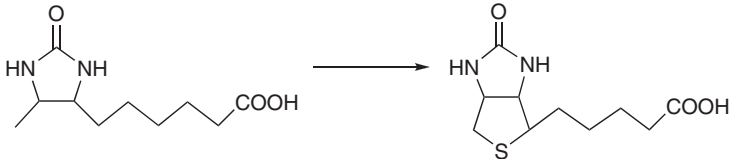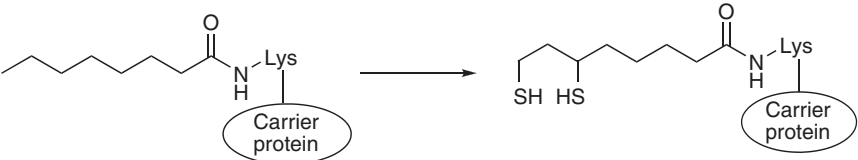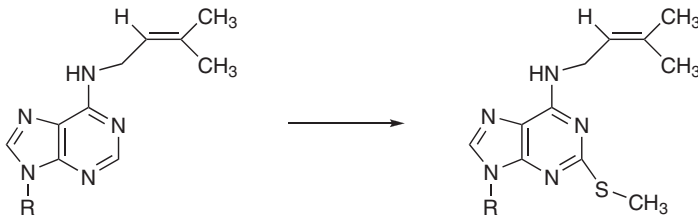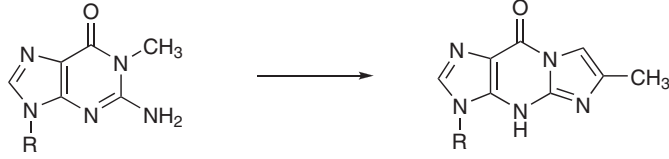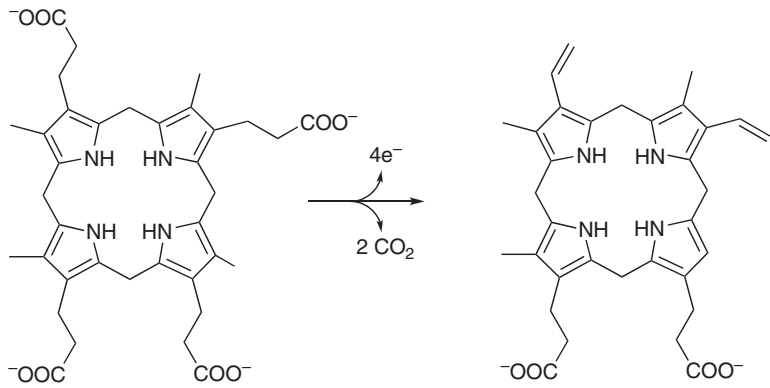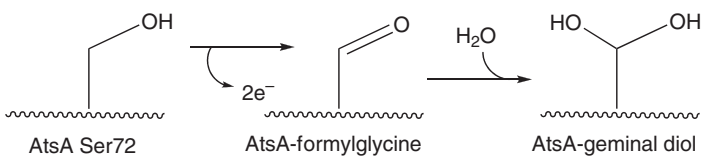
Enzyme	Reaction catalyzed	Role of SAM
BioB		Substrate
LipA		Substrate
MiaB		Substrate
TYW1		Substrate
HemN		Substrate
AtsB		Substrate

Figure 4   (Continued)

Enzyme	Reaction catalyzed	Role of SAM
GRE-AE		Substrate
LAM		Cofactor
MoaA		Substrate
ThiC		Substrate
ThiH		Substrate
SPL		Cofactor

Figure 4   Representative reactions catalyzed by radical SAM enzymes. BioB, biotin synthase; LipA, lipoyl synthase; MiaB, enzyme catalyzing methylthiolation of tRNA; TYW1, enzyme involved in wybutosine biosynthesis; HemN, oxygen-independent coproporphyrinogen synthase; AtsB, arylsulfatase-activating enzyme; GRE-AE, glycyl radical-activating enzyme; LAM, lysine 2,3-aminomutase; MoaA, molybdopterin cofactor biosynthesis enzyme; ThiC, HMP-P synthase; ThiH, dehydroglycine-generating enzyme in thiamine biosynthesis; SPL, spore photoproduct lyase.

intermediates involving metal–adenosyl complexes.[26,50] As discussed below, however, the radical SAM enzymes appear to use a completely unprecedented mechanism for generating the extremely reactive 5′-deoxyadenosyl radical intermediate.

8.17.2.2 Properties of the Iron–Sulfur Clusters

Although in general there is little sequence homology among the diverse members of the radical SAM superfamily, all harbor a conserved CX_3CX_2C motif; for several members of the superfamily, this motif has been shown to provide the ligands for a catalytically essential iron–sulfur cluster.[66–71] Although different types of iron–sulfur clusters have been identified in several of the radical SAM enzymes, it is now clear that the prototypical radical SAM cluster is a [4Fe–4S] cluster. As suggested by the presence of only three cysteines in the conserved motif, this cluster has only three cysteinal ligands, one to each of three irons of the cluster, with the fourth iron being coordinated by a noncysteinal ligand; such a cluster is referred to as a site-differentiated cluster because one of the four iron sites is chemically distinct. The fourth iron is coordinated by SAM in the E-SAM complex (the enzyme-SAM bound complex, further details are given below); however, the ligand(s) to the unique iron in the absence of SAM remains unknown. The site-differentiated nature of the [4Fe–4S] clusters of radical SAM enzymes was suggested not only by the three cysteine motif, but also by the propensity of some of these enzymes to undergo oxidation of the [4Fe–4S] cluster with loss of one iron to produce a [3Fe–4S]$^+$ state. Spectroscopic evidence for site differentiation was subsequently confirmed by Mössbauer spectroscopy of site-specifically labeled PFL-AE in which either the unique site alone or the three nonunique sites were labeled with the Mössbauer-active isotope ^{57}Fe; the spectroscopic studies demonstrated that the unique iron had distinct spectroscopic properties that were dramatically altered in the presence of SAM, while the other three irons were unperturbed by the presence of SAM.[72] These Mössbauer studies provided the first spectroscopic evidence for coordination of the unique iron by SAM. A more detailed picture emerged from electron-nuclear double resonance (ENDOR) studies, first of PFL-AE[73,74] and then of LAM,[75] complexed to site-specifically labeled SAMs, which provided the first direct evidence for the coordination of the unique iron of the [4Fe–4S] cluster by the amino and carboxyl groups of the methionine portion of SAM (**Figure 5**). In addition, the ENDOR studies revealed the presence of an orbital overlap between the iron–sulfur cluster and the sulfonium of SAM, a feature proposed to be central to the mechanism of initiation of radical chemistry. The unprecedented coordination of SAM to the iron–sulfur cluster, as discovered using ENDOR spectroscopy, was subsequently confirmed by every reported crystal structure of a radical SAM enzyme (see Section 8.17.4).

The catalytically active state of the radical SAM cluster was first clearly demonstrated via single turnover experiments performed on the PFL-AE.[76] In these experiments, PFL-AE was reduced from the [4Fe–4S]$^{2+}$ state to the [4Fe–4S]$^+$ state by photoreduction with 5-deazariboflavin; by removing the source of illumination, the two specific states of the cluster could be examined for their ability to generate the glycyl radical on PFL in the absence of exogenous reductant. It was found that the quantity of glycyl radical generated on PFL was
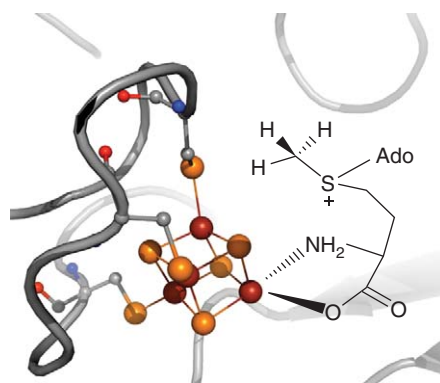
Figure 5 Coordination of SAM to the unique site of the site-differentiated radical SAM cluster, as first elucidated by ENDOR studies on PFL-AE and LAM.

directly correlated with the amount of $[4Fe–4S]^+$ cluster present on PFL-AE, suggesting that one reduced $[4Fe–4S]^+$ cluster was capable of generating a single glycyl radical on PFL. Furthermore, the data demonstrated that the $[4Fe–4S]^+$ cluster was oxidized by one electron to an EPR-silent state upon generation of the glycyl radical. Together, these data provided support for a mechanism in which a reduced $[4Fe–4S]^+$ cluster provides the electron required for reductive cleavage of SAM to generate a 5′-deoxyadenosyl radical intermediate, which in this case abstracts a hydrogen atom from G734 of PFL to generate the catalytically active glycyl radical state of the enzyme. Evidence for the catalytic relevance of the $[4Fe–4S]^+$ state has been obtained for other radical SAM enzymes as well, including LAM[77] and the ARR-activating enzyme (ARR-AE).[78]

Although the experimental evidence just described supported a role for the [4Fe–4S] clusters of radical SAM enzymes in providing the electron needed for reductive cleavage of SAM, it was immediately clear that the chemistry must be more involved than a simple long-range electron transfer reaction. While trialkylsulfoniums like SAM have reduction potentials of approximately −1.8 V,[60,79,80] the radical SAM clusters for which potentials have been determined have potentials of approximately −450 mV. These potentials would suggest that the reduction of SAM by one of the radical SAM iron–sulfur clusters would be uphill by approximately 1.4 V, or 32 kcal mol^{-1}; such a barrier is unrealistic given the rates of the radical SAM enzyme-catalyzed reactions. Recent work on LAM has provided clues as to how the radical SAM enzymes may overcome this thermodynamic barrier, with the binding of SAM and substrate providing significant perturbation of the redox potentials of both the SAM sulfonium and the iron–sulfur cluster, such as to render the reaction more accessible thermodynamically.[81,82]

8.17.2.3   A Mechanism for Radical SAM Chemistry

A general mechanistic scheme that is expected to reflect the initial steps in catalysis for all the diverse radical SAM enzymes is illustrated in **Figure 6**. In this mechanism, a $[4Fe–4S]^{2+}$ cluster with SAM coordinated via the
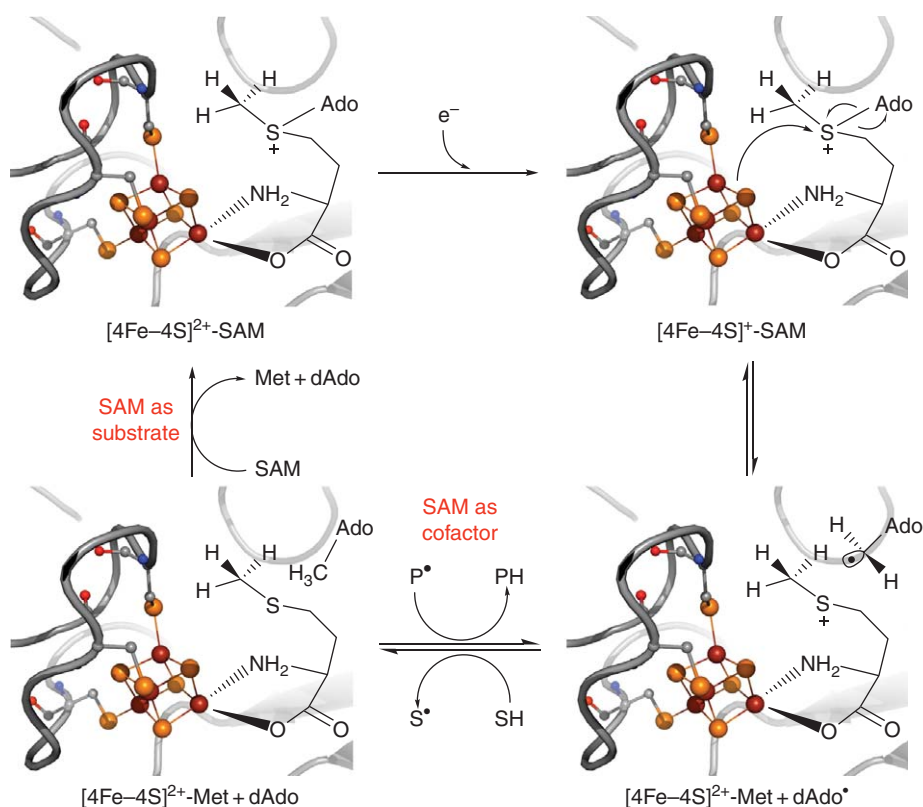
Figure 6   A general mechanism for reactions catalyzed by radical SAM enzymes, where SAM is utilized as either a cofactor or a substrate. The protein backbone and iron–sulfur cluster were generated using PyMol and PDB id 3cb8.pdb.

amino and carboxyl moieties is reduced by one electron to the catalytically active state. In this state and in the presence of substrate, the $[4Fe–4S]^+$ cluster transfers an electron via an inner-sphere mechanism to the sulfonium sulfur of SAM, thereby promoting homolytic S–C5′ bond cleavage to generate methionine (still bound at the unique Fe site) and the 5′-deoxyadenosyl radical intermediate. The 5′-deoxyadenosyl radical intermediate then abstracts H· from substrate, and the resulting substrate radical intermediate may either be the end product (as in the glycyl radical enzymes (GREs)) or undergo further transformation to produce a product or a product radical. In most of the radical SAM enzymes characterized to date, SAM acts as a cosubstrate and is converted stoichiometrically to methionine and 5′-deoxyadenosine (**Figure 4**). In some, however, including both LAM and spore photoproduct lyase (SPL), SAM is used as a cofactor, and a product radical generated during turnover reabstracts H· from 5′-deoxyadenosine to regenerate the 5′-deoxyadenosyl radical, which reacts with the methionine with loss of an electron to the $[4Fe–4S]^{2+}$ cluster to regenerate SAM and the reduced cluster.

8.17.3   Diverse Reactions Catalyzed by Radical SAM Enzymes

8.17.3.1   Activating Enzymes and the Generation of Glycyl Radicals

PFL was the first enzyme discovered to require a glycyl radical as an essential cofactor in catalysis.[26] This homodimer of 85 kDa subunits is produced in an inactive state *in vivo* and undergoes posttranslational activation by the PFL-AE, which we now know utilizes radical SAM chemistry to carry out this reaction. Specifically, PFL-AE utilizes SAM, an iron–sulfur cluster, and a source of electrons (flavodoxin *in vivo*) to generate a 5′-deoxyadenosyl radical intermediate; this key intermediate stereospecifically abstracts the pro-*S* hydrogen from G734 of PFL.[83] Interestingly, G734 is buried in the X-ray crystal structures of PFL, 8 Å from the nearest surface,[84] and yet the biochemical evidence points clearly to a direct H· abstraction from G734 as the activation event catalyzed by PFL-AE. Such a direct H-atom abstraction requires that the site of generation of the 5′-deoxyadenosyl radical at the PFL-AE active site must be in close proximity to the G734 of PFL; insight into how this close proximity is achieved is discussed further in the section on the PFL-AE crystal structure later in this chapter.

Despite the novelty at the time of the first identified stable and catalytically essential glycyl radical in PFL, increasing numbers of these GREs are being found in diverse areas of anaerobic metabolism. As mentioned previously in this chapter, the ARR is a GRE and is activated by a radical SAM-activating enzyme. Both the PFL-AE and ARR-AE have been characterized fairly extensively. Both harbor a single [4Fe–4S] cluster that can be reduced to the catalytically active $[4Fe–4S]^+$ state, and both utilize SAM as a substrate. Differences in the propensity for 'nonproductive' cleavage of SAM (cleavage that occurs in the absence of substrate) have been reported for these two activating enzymes; however, it is not clear whether such differences are merely a result of differences in protein preparation methods. To summarize, these two activating enzymes appear to be quite similar in nearly all aspects, but they do not cross-react: PFL-AE does not activate ARR and ARR-AE does not activate PFL. It is fascinating to consider that two enzymes so central to anaerobic metabolism, one for glucose metabolism and one for the reduction of ribonucleotides, both rely on glycyl radicals for catalysis.

It is now clear that PFL and ARR are part of a large group of enzymes harboring glycyl radicals that are found in both strict and facultative anaerobes.[85] The prevalence of glycyl radical chemistry in anaerobic metabolism is presumably a result of its effectiveness as a cofactor in difficult chemical transformations that must be carried out in the absence of oxygen. Examples of other GREs include benzylsuccinate synthase (BSS), which catalyzes the first step in anaerobic toluene degradation, the adenosylcobalamin-independent glycerol dehydratase (GDH), which catalyzes a key step in glycerol fermentation, and the hydroxyphenylacetate decarboxylase (HPD), which catalyzes a difficult decarboxylation reaction to produce *p*-cresol. All of these GREs harbor a glycyl radical motif, a short region of sequence homology surrounding the site of the glycyl radical, and all are thought to function by transferring the glycyl radical to a thiol in the active site, with the resulting thiyl radical participating directly in catalysis. Although all three of these GREs have subunits that are known or predicted to be structurally very similar to PFL and ARR (10-stranded α/β barrels with finger loops harboring the site of the glycyl radical and the active site cysteine(s)), the BSS and HPD have additional small subunits of unknown function and HPD has a bound iron–sulfur cluster.[86] Each of these GREs also has a

putative radical SAM AE gene in the same operon; in the case of the HPD-activating enzyme, two [4Fe–4S] clusters are bound and the enzyme has been reported to activate as well as deactivate HPD.[85]

It is also of interest to note that the GRE-activating enzymes are not the only radical SAM proteins that act on a protein substrate; AtsB catalyzes the formation of a Cα-formylglycine (2-amino-3-oxopropionic acid, FGly) on AtsA, thereby activating this arylsulfatase (see Section 8.17.3.8).

8.17.3.2   LAM and Related Mutases

As noted in an earlier section of this chapter, LAM was among the earliest enzymes recognized to utilize an iron–sulfur cluster and SAM to initiate radical catalysis. LAM is also the enzyme that prompted the most insightful analogies between adenosylcobalamin- and SAM-mediated radical reactions. LAM catalyzes a classic B_{12}-type reaction, the reversible interconversion of L-α-lysine and L-β-lysine, involving abstraction of H· from an unactivated C–H bond of substrate, followed by rearrangement and reabstraction of H· from 5′-deoxyadenosine by a product radical. In addition to SAM and a [4Fe–4S] cluster, this reaction requires PLP, which forms a covalent adduct with lysine and the radical intermediates (**Figure 7**). By use of the AdoMet analog 3′,4′-anhydroadenosylmethionine (anSAM) in the LAM reaction, Frey and coworkers were able to provide the first direct evidence for an intermediate 5′-deoxyadenosyl radical in the radical SAM enzymes.[87,88] By incubation of this SAM analog with reduced enzyme and substrate, a new EPR signal was observed that could be assigned to an allylic analog of the 5′-deoxyadenosyl radical. ENDOR spectroscopy has also been utilized in conjunction with substrates and analogs to monitor the motions of active site components during catalytic turnover; the results provide a rare glimpse into a working enzyme active site and provide evidence for control of the radical chemistry by maintenance of van der Waals contacts between radicals and their reacting partners.[89] Although no such studies have been conducted on the other radical SAM enzymes, it is reasonable to suspect that other enzymes will exhibit similar means to control the extremely reactive intermediates, particularly the 5′-deoxyadenosyl radical, generated during turnover. Glutamate 2,3-aminomutase is another [4Fe–4S]-, SAM-, and PLP-dependent aminomutase, which appears to have similar properties to LAM.[90]

8.17.3.3   Sulfur Insertion: Biotin Synthase and Lipoate Synthase

BioB and lipoate synthase represent a group of radical SAM enzymes that catalyze sulfur insertion reactions (**Figure 4**) and were among the group of enzymes identified early on to utilize iron–sulfur clusters and SAM in
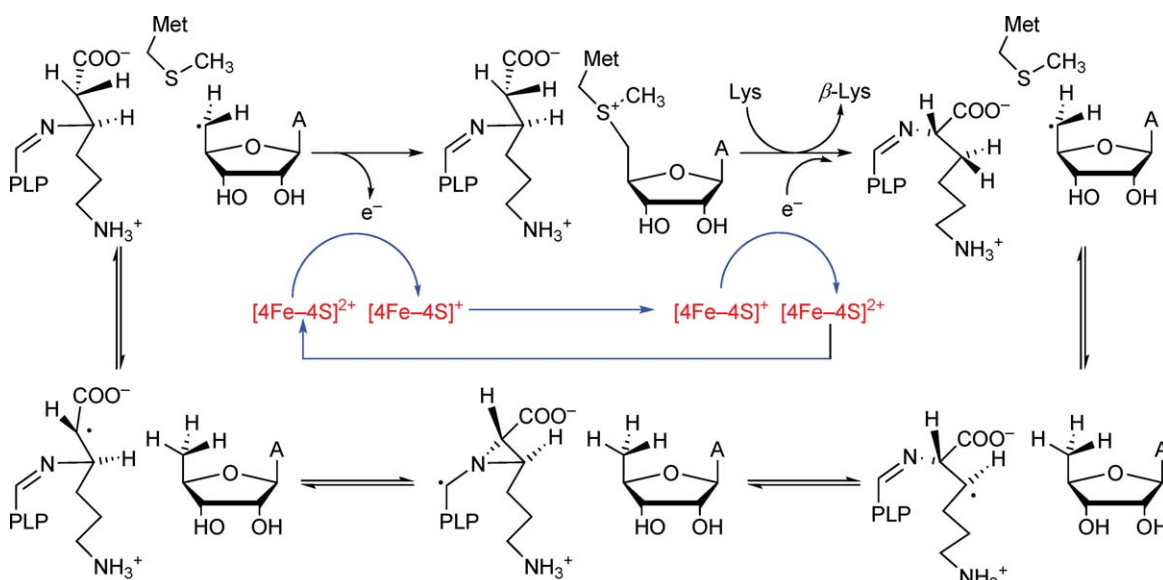
Figure 7   Mechanism for the rearrangement reaction catalyzed by lysine 2,3-aminomutase.

catalysis. Spectroscopic studies on BioB provided evidence for both [4Fe–4S] and [2Fe–2S] clusters and for facile interconversions between these clusters.[42,91,92] The cluster interconversions were ultimately revealed to involve cluster disassembly and reassembly under appropriate conditions, rather than dimerization of [2Fe–2S] clusters to generate [4Fe–4S] clusters.[93] The source of the sulfur required for the synthesis of biotin has been the subject of considerable debate; while one report classifies BioB as a PLP-dependent cysteine desulfurase that utilizes cysteine as the sulfur source,[94] considerably greater evidence points to a [2Fe–2S] cluster bound to BioB as the source of sulfur. First, chemical, spectroscopic, and spectroelectrochemical analysis provided evidence for two distinct iron–sulfur cluster binding sites on BioB,[95,96] and subsequent spectroscopic studies demonstrated that a [2Fe–2S] cluster is destroyed during turnover.[97,98] The crystal structure of BioB revealed the presence of one radical SAM [4Fe–4S] cluster, as well as a [2Fe–2S] cluster positioned close to the bound dethiobiotin (DTB), such that one could envision a DTB radical intermediate reacting with the [2Fe–2S] cluster to produce the first sulfur insertion (see Section 8.17.4.1). It has recently been demonstrated that the putative product of the first sulfur insertion reaction, 9-mercaptodethiobiotin, is a competent catalytic intermediate in the reaction catalyzed by BioB.[99] The synthesis of biotin would then require a second reductive cleavage of SAM to initiate the second sulfur insertion event, and indeed two equivalents of SAM have been shown to be cleaved per biotin synthesized by BioB.[99] If a [2Fe–2S] cluster in BioB was serving as the source of sulfur in biotin biosynthesis, then one would expect that only a single turnover of the enzyme would be possible in the absence of the machinery required to rebuild the [2Fe–2S] cluster, and this expectation is in general agreement with the experimental results published over the years on BioB; that is, BioB *in vitro* does not function as a true catalyst, but rather as a reagent, in the synthesis of biotin. *In vivo*, BioB has been shown to function catalytically, although catalysis renders the protein susceptible to proteolytic degradation.[100] A similar susceptibility of BioB to proteolytic degradation *in vitro* after loss of either the [4Fe–4S] or [2Fe–2S] clusters has also been reported, pointing to an important role of these clusters not only in catalysis, but also in stabilizing the protein structure.[101]

Although the gene encoding lipoyl synthase (LipA) was cloned in 1992,[102] and the identification of LipA as an iron–sulfur protein followed,[45,103] no catalytic activity was observed for this enzyme until the finding that octanoylated proteins (e.g., octanoyl-acyl carrier protein (octanoyl-ACP)[46] and the H-protein of the glycine cleavage system[104]), and not free octanoic acid, served as the substrate for LipA. Subsequent studies have revealed striking similarities to BioB. Like BioB, LipA binds two distinct iron–sulfur clusters; however, in the case of LipA, both clusters are [4Fe–4S] clusters: one of these is a radical SAM cluster while the other may play a role in sulfur insertion, as proposed above for the [2Fe–2S] cluster of BioB.[105] Also analogous to BioB, LipA utilizes two equivalents of SAM to synthesize one equivalent of the lipoyl cofactor.[104] It has been shown that both sulfurs inserted to synthesize the lipoyl cofactor arise from the same LipA molecule[106] and that the sulfurs are inserted sequentially, with insertion at C6 occurring first.[107] A large isotope effect has been observed for the H-atom abstraction from C8.[107]

8.17.3.4   DNA Repair: SPL

The major photoproduct in UV-irradiated *Bacillus* spore DNA is a unique thymine dimer called spore photoproduct (SP, 5-thyminyl-5,6-dihydrothymine).[108–110] In contrast, UV irradiation of DNA in most growing cells produces primarily cyclobutane pyrimidine dimers as well as the 6,4-photoproduct.[111] The unusual UV photochemistry of *Bacillus* spores appears to be largely associated with the presence in spores of large quantities of a family of proteins known as small, acid-soluble proteins (SASPs).[112–115] It has been proposed that binding of SASPs to DNA promotes a structural change and a change in the level of hydration of the DNA that results in the formation of SP rather than cyclobutane thymine dimers.[116–118] Pyrimidine dimers such as SP are damaging to cells, as they can block replication and transcription or can result in mutations if transcription proceeds past the region of the dimer. Repair of these dimers, therefore, is critical in order to avoid mutations, and thus is the key to UV resistance. Although pyrimidine dimers can be excised and replaced, the only well-characterized example of direct pyrimidine dimer reversal is the photoreactivation catalyzed by DNA photolyase.[111,119] However, photoreactivation has been shown to be absent in many species, including *Bacillus*, suggesting that alternate means of pyrimidine dimer repair might be found.[109,120]

The enzyme SPL is the first identified nonphotoactivatable pyrimidine dimer lyase and it specifically targets SP and cleaves it into two thymines by a light-independent mechanism.[121,122] Early publications[123,124] provided

evidence that SPL utilized SAM and contained an iron–sulfur cluster, suggesting that it was a member of the Fe–S/AdoMet family of enzymes. Subsequent work has shown that the $[4Fe–4S]^+$ state of SPL is active in SP repair and that SP repair is initiated by direct H-atom abstraction from the C6 of SP.[125,126] Evidence that SP utilized SAM catalytically, and did not generate 5′-deoxyadenosine and methionine as products of turnover, placed SPL alongside LAM as the radical SAM enzyme that utilizes SAM as a cofactor. Following on the initial reports of SP synthesis by Begley and coworkers,[127–129] Carell and coworkers[130] have reported the synthesis and assay of 5*R*- and 5*S*-dinucleoside SP. Although the extent of turnover observed was extremely small, they concluded that SPL repairs only the 5*S* isomer of SP; this result was quite a surprise, as the 5*S* isomer would be formed in A-DNA only via interstrand cross-links and not by cross-linking adjacent thymines on the same DNA strand. A more recent analysis of *in vitro* enzymatic assays on stereochemically defined SP substrates demonstrated that SPL specifically repairs only the 5 *R* isomer of SP. The observation that 5 *R*-SP, but not 5 *S*-SP, is a substrate for SPL is consistent with the expectation that 5 *R* is the SP isomer produced *in vivo* upon UV irradiation of bacterial spore DNA.[131]

8.17.3.5 Synthesis of Complex Organic Cofactors

8.17.3.5.1 *Molybdenum cofactor biosynthesis*

Molybdenum is found in biology primarily either in nitrogenase, in which the molybdenum ion is part of an iron–sulfur–molybdenum cluster known as the FeMoco (see Section 8.17.3.6), or in certain redox enzymes that catalyze oxo-transfer reactions; in the latter enzymes, molybdenum is coordinated to a tricyclic pyranopterin known as molybdopterin in an active site metal center referred to as the molybdopterin cofactor (MoCo). The biosynthesis of molybdopterin is an evolutionarily conserved pathway encompassing several reactions that are not well understood. The enzymes MoaA and MoaC catalyze the first step in MoCo synthesis involving the radical rearrangement of the guanine C8 atom in 5′-GTP into precursor Z, a tetrahydropyranopterin containing a cyclic phosphate moiety (**Figure 4**). MoaA was first identified as an iron–sulfur protein in 1996, and was identified as a member of the radical SAM superfamily by Sofia *et al.*[47] Subsequent detailed spectroscopic and biochemical characterization of the human version of MoaA, referred to as MOCS1A, revealed that the purified protein contained two iron–sulfur clusters: an N-terminal 'radical SAM' cluster coordinated by the characteristic CX_3CX_2C motif and an additional site-differentiated [4Fe–4S] cluster in the C-terminal region coordinated by a $CX_2CX_{13}C$ motif.[71] Both clusters were found to be essential for enzymatic activity; however, beyond a role in enzyme stabilization, no specific function was identified for the C-terminal cluster in these studies. Further insight into the mechanism by which MoaA catalyzes the synthesis of precursor Z was provided by crystal structures of the enzyme, which are discussed in Section 8.17.4.4 of this chapter.

8.17.3.5.2 *Heme biosynthesis*

The conversion of coproporphyrinogen III to protoporphyrinogen IX via the oxidative decarboxylation of the propionate side chains of rings A and B to vinyl groups during heme biosynthesis can be catalyzed by two unrelated enzymes. In eukaryotes, HemF catalyzes this reaction by using O_2 as an electron acceptor. In bacteria, HemN catalyzes this reaction by using SAM as a one-electron oxidizing agent, while also utilizing an additional unidentified molecule as a second electron acceptor (**Figure 4**). The overall reaction involves four electrons, two of which are used to cleave SAM via the $[4Fe–4S]^+$ clusters and two which are transferred to the unidentified small molecule oxidizing agent X.

When purified under anaerobic conditions, HemN contains a single $[4Fe–4S]^{2+}$ cluster with a unique site that binds AdoMet, as identified by Mössbauer spectroscopy. As will be discussed further in the section on the HemN crystal structure later in this chapter, there is evidence that HemN binds two molecules of SAM simultaneously, with one of these bound to the unique site of the iron–sulfur cluster and the other bound further away in an active site channel. Similar to ARR-AE and BioB, HemN-catalyzed SAM cleavage in the absence of substrate (i.e., nonproductive cleavage) occurred only to a minor extent and only when the strong reducing agent $Na_2S_2O_4$ was used.[132] However, when *Escherichia coli* cell-free extract containing the physiological electron donor system was used as the reductant in assays, SAM cleavage occurred only in the presence of coproporphyrinogen[133]. This observation suggests that the energetic barrier accompanying the cleavage of SAM is partly overcome by coupling electron transfer from the cluster to SAM to the reaction of the dAdo radical produced with substrate. Moreover, mutagenesis studies of the amino acids responsible for binding SAM2 (the SAM not bound to the cluster in the

crystal structure) provide strong evidence for the involvement of both SAM1 and SAM2 in the catalytic mechanism of protoporphyrinogen IX formation, invoking two active sites for decarboxylation of each propionate side chain relative to each SAM molecule.[132] This proposed '2 SAM' mechanism for HemN is discussed further in the section on the HemN crystal structure Section 8.17.4.6 in this chapter.

Direct evidence for a radical mechanism for decarboxylation catalyzed by HemN has been obtained by EPR spectroscopy.[134] When SAM and substrate coproporphyrinogen III are added to reduced HemN, an organic radical EPR signal is observed at $g = 2.002\ 9$, which has a complex pattern of hyperfine couplings from at least five different hydrogen atoms. Identification of this EPR signal as arising from a substrate radical was made by use of regiospecifically labeled (^{15}N or ^{2}H) substrates; the results demonstrated that the unpaired electron was delocalized over carbons 3′, 3, and 4, and thus experienced allylic stabilization. The location of this substrate radical in the porphyrin ring is consistent with the proposed mechanism in which the 5′-deoxyadenosyl radical abstracts the pro-*S* hydrogen at the β position of the propionate side chain to initiate the oxidative decarboxylation reaction.

8.17.3.5.3   Thiamine biosynthesis: ThiC and ThiH

Thiamine pyrophosphate (TPP) plays a vital role in carbohydrate and amino acid metabolism and is an essential cofactor for all living organisms. The final step of TPP biosynthesis involves the cross-linking of two differentially synthesized heterocyclic precursor molecules, 4-amino-5-hydroxymethyl-2-methylpyrimidine pyrophosphate (HMP-PP) with 4-methyl-5-β-hydroxyethylthiazole phosphate (thiazole phosphate, TMP), by the ThiE enzyme. The biosynthesis of TPP follows distinct pathways for aerobic and anaerobic organisms, given the presence of ThiO in aerobes and ThiH in anaerobes. However, these pathways converge with the synthesis of the common intermediate dehydroglycine, which is formed from glycine by the action of ThiO and from tyrosine by the action of ThiH (**Figure 8**). Dehydroglycine is subsequently
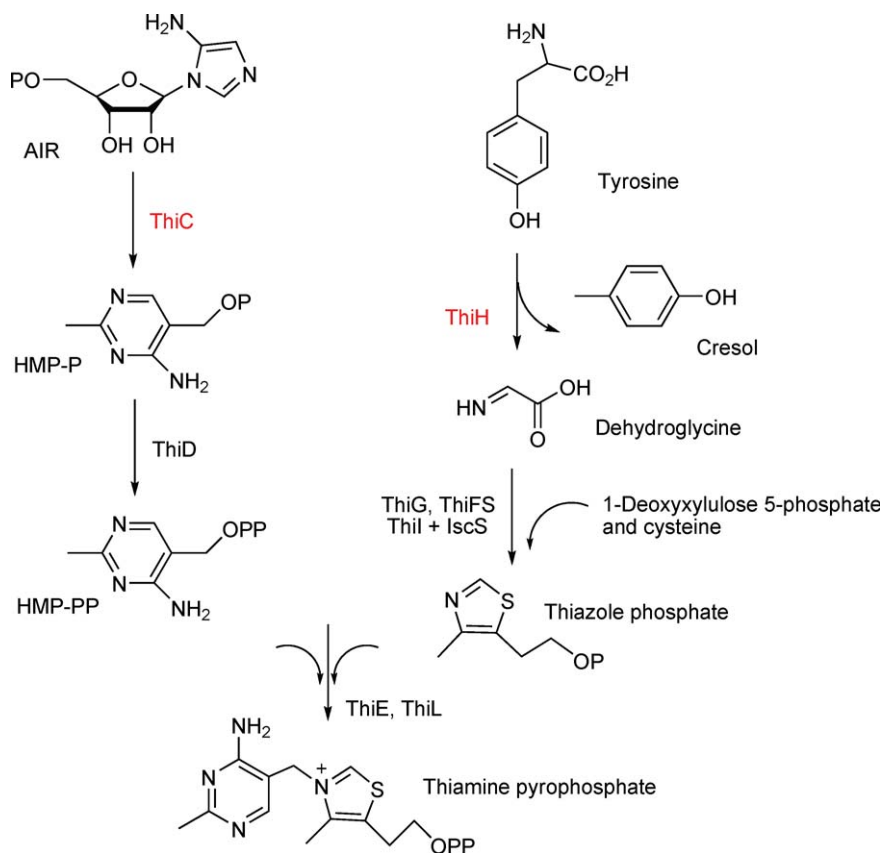
Figure 8   Biosynthesis of thiamine pyrophosphate in *Escherichia coli*. The two radical SAM enzymes involved in this process are highlighted in red. P = phosphate group.

incorporated into TMP in a multistep process requiring thiazole synthase (ThiG), the sulfur donor ThiFS thiocarboxylate, and 1-deoxyxylulose 5-phosphate.

HMP-PP is derived from 5-aminoimidazole ribonucleotide (AIR) in bacteria, while in yeast it is derived from histidine and pyridoxal 5′-phosphate. In both systems, only a single gene product is required for HMP-PP formation: ThiC in bacteria, algae, and plants, and THI5 in yeast. In bacteria, the conversion of AIR to 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMP-P) has been well studied and involves an amazingly complex rearrangement reaction involving the breaking and reformation of multiple bonds. *In vivo* labeling studies have conclusively demonstrated that all the carbon and nitrogen atoms present in 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) are derived from AIR (**Figure 9**).[135,136]

Recently, the successful reconstitution of ThiC from *Salmonella enterica*[137] and *Caulobacter crescentus*[138] has been reported. These biochemical studies have shown that ThiC reductively cleaves SAM in the presence of AIR, generating HMP-P and deoxyadenosine as reaction products. Mössbauer and EPR spectroscopic characterization of the enzyme from *C. crescentus* show the presence of a [4Fe–4S] cluster. Structurally, *C. crescentus* ThiC reveals a remarkable twist on what we have observed for the other structurally characterized radical SAM enzymes.[138] While ThiC has been biochemically shown to act as a radical SAM enzyme, the sequence intriguingly does not contain the canonical CX_3CX_2C motif common to all other members of the radical SAM superfamily. Among known radical SAM sequences, the CX_3CX_2C motif is most commonly found near
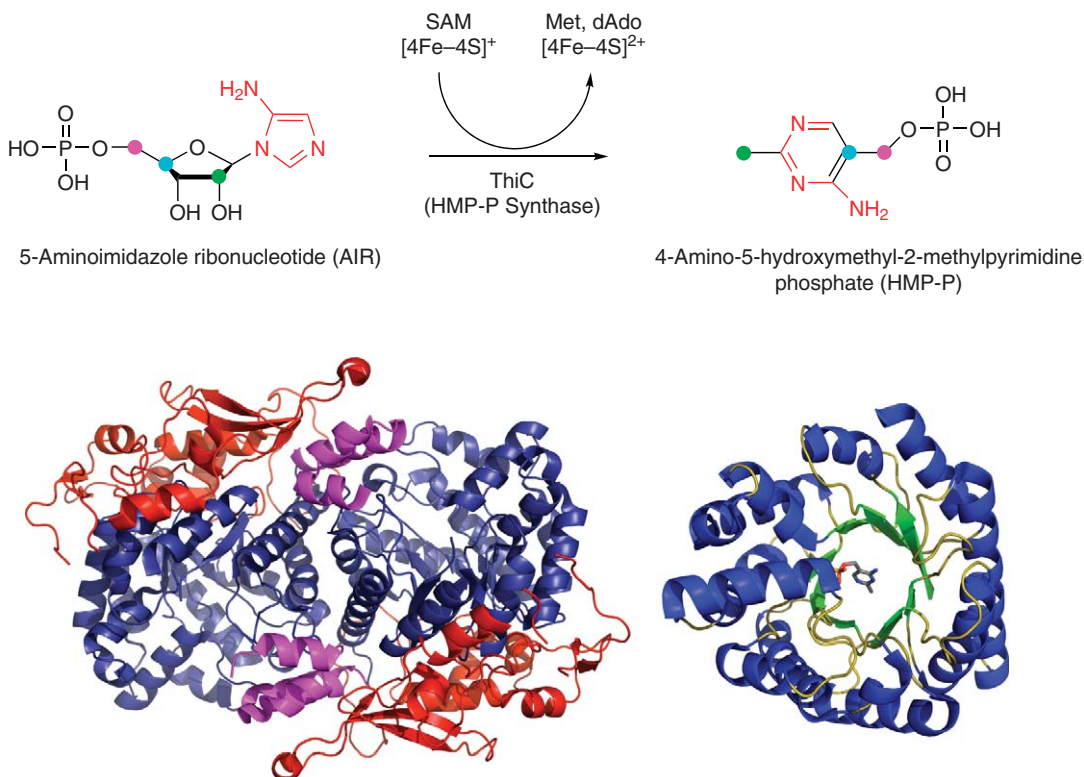
Figure 9   (Top) Mechanism catalyzed by the radical SAM enzyme ThiC in 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate (HMP-P) formation in bacteria and plants. The color coding shows the source of the nonhydrogen atoms in HMP-P, as derived from 5-aminoimidazole ribonucleotide (AIR) labeling studies. (Bottom left) Crystal structure of the bacterial ThiC enzyme from *Caulobacter crescentus* (PDB code: 3epo). The homodimeric structure is shown. Domains 1 are colored red, domains 2 are dark blue, and domains 3 are magenta. (Bottom right) The catalytic domain 2 from one subunit is shown. Secondary structural elements are colored blue for α-helices, green for β-sheets, and olive for random coil. The cysteine residues comprising the C-terminal CX_3CX_2C motif are disordered in electron density maps and cannot be definitively located. However, the putative active site region is denoted by the presence of the bound product HMP-P (shown as a stick figure).

the N-terminus, and is occasionally present toward the middle of the peptide sequence. In the case of ThiC, a CX_2CX_4C motif is present in the C-terminus of the protein. ThiC is homodimeric, with each monomer being comprised of three domains (**Figure 9**). Domain 1 makes up the N-terminal portion of the protein and contains a novel protein fold consisting of six α-helices and five β-strands. Domain 2 is the catalytic domain and has the full $(\alpha/\beta)_8$ TIM barrel fold that was similarly observed for BioB and HydE. Domain 3, the smallest domain of the protein, is the C-terminal portion of ThiC and contains three α-helices arranged in an antiparallel fashion.

Unfortunately, no [Fe–S] cluster was bound in the X-ray structure but assignment of the active site cavity could be made by two additional structures solved in the presence of the HMP-P product and the substrate analog imidazole ribonucleotide (IMR).[138] The active site cavity is located toward the C-terminal end of the $(\alpha/\beta)_8$ TIM barrel and the HMP-P binding site is comprised of several hydrophobic residues. The N1 and N2 atoms of the HMP-P form hydrogen bonds with a glutamate and a serine residue, while the phosphate group is stabilized by a helix dipole and several hydrogen bonds with adjacent amino acids. The phosphate binding site of the substrate analog IMR involves many of the same residues as observed for the HMP-P product. Moreover, the imidazole and ribose groups of IMR overlap the pyrimidine binding site of HMP-P. Modeling of the TIM barrel fold of ThiC on the BioB structure allowed for the theoretical positioning of a SAM-bound [4Fe–4S] cluster within the active site. In this model, SAM was found to be closely positioned to the ribose ring of IMR, establishing a putative radical transfer region within the active site cavity. While the exact nature of the rearrangement reaction catalyzed by ThiC remains to be elucidated, future studies will undoubtedly clarify how this unique radical SAM enzyme carries out one of the most difficult rearrangements in primary metabolism.

As mentioned above, ThiH is responsible for dehydroglycine formation from tyrosine in anaerobic organisms (**Figure 4**). Dehydroglycine is then incorporated into thiazole phosphate, which is subsequently covalently linked to HMP-PP to yield thiamine phosphate (**Figure 8**). A final phosphorylation reaction catalyzed by ThiL generates TPP, an essential cofactor for multiple enzymes involved in amino acid and carbohydrate metabolism. Sequence annotation identified ThiH as belonging to the radical SAM superfamily given the presence of the canonical CX_3CX_2C motif, and amino acid substitutions in this motif of *S. enterica* ThiH provided evidence that these residues were critical to the *in vivo* function of this enzyme.[139]

The initial characterization of the ThiH enzyme revealed that it could be purified in approximate 1:1 stoichiometry with ThiG and the as-isolated enzyme contained a $[3Fe–4S]^+$ cluster. Consistent with expectations for a putative radical SAM enzyme, reduced ThiH contains an O_2-sensitive $[4Fe–4S]^+$ cluster.[140,141] Addition of tyrosine to the reduced enzyme resulted in no change in the EPR *g*-values of the reduced cluster; however, addition of SAM resulted in a remarkable increase in the amount of reduced cluster present. Conversely, in the presence of both SAM and tyrosine; only a very small amount of EPR-active clusters was observed, supporting a mechanism in which the presence of tyrosine enables the $[4Fe–4S]^+$ cluster to reductively cleave SAM, producing an EPR-silent $[4Fe–4S]^{2+}$ cluster. Importantly, using radiolabeled substrates, the activity of ThiGH was shown to require both SAM and tyrosine, and data suggest that a step following the initial ThiH-dependent tyrosine cleavage is rate-limiting in thiazole formation.[141,142] Moreover, 5′-deoxyadenosine, cresol, and glyoxylate (the expected product of dehydroglycine hydrolysis) have been confirmed to be produced during ThiGH catalysis.[143] Taken together, the available data provide a picture of ThiH catalysis in which the 5′-deoxyadenosyl radical resulting from reductive cleavage of SAM abstracts H· from tyrosine to generate a tyrosyl radical, which subsequently decomposes to dehydroglycine and *p*-cresol. Additional data on the ThiH enzyme aimed at identifying the nature of active site residues should help to delineate whether the tyrosine radical generated upon extraction of the phenolic hydrogen reacts via homolytic or heterolytic cleavage of the $C\alpha$–$C\beta$ bond (**Figure 10**). Structural information should also help to explain how the dehydroglycine product is protected against hydrolysis.

8.17.3.5.4   Bacteriochlorophyll biosynthesis: BchQ and BchR

Photosynthetic organisms use a variety of pigments, variations of magnesium tetrapyrrole structures, to harvest light.[144] Green sulfur bacteria have evolved unique, self-aggregating pigment structures called chlorosomes, which are mainly comprised of bacteriochlorophyll (BChl) *c*, *d*, or *e* pigments, differing in their esterification nature.[145] In the green sulfur bacterium *Chlorobaculum tepidum*, 97% of the total pigment content is BChl c_F (BChl *c* esterified with farnesol), with each chlorosome being constructed of approximately 215 000 of these pigment molecules.[146] The BChl *c* in *C. tepidum* is a mixture of four homologs that carry different modifications
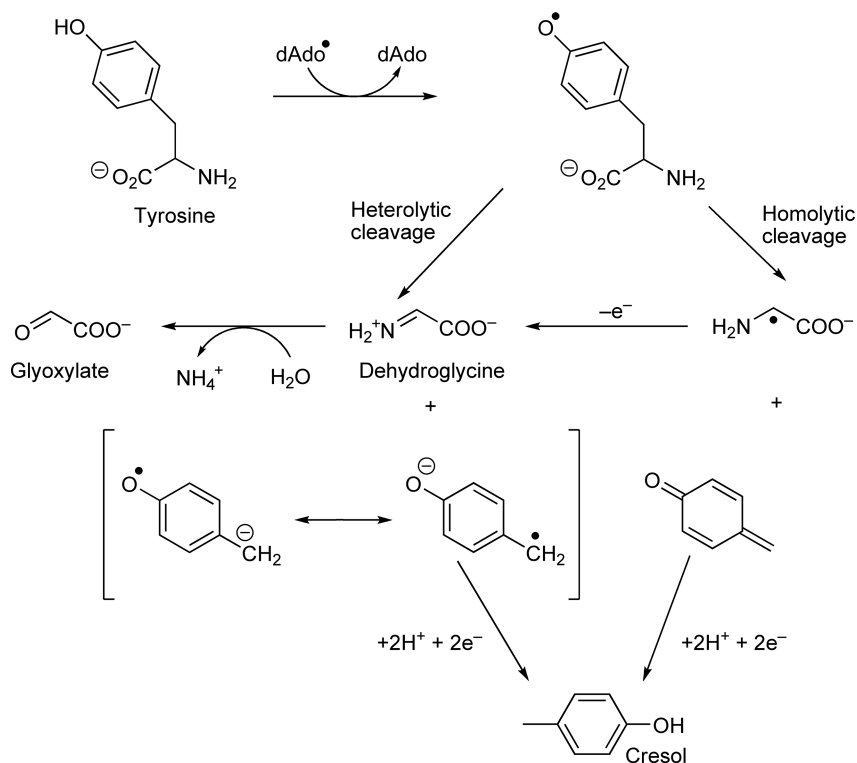
Figure 10   The proposed reaction pathways for the tyrosyl radical generated upon phenolic hydrogen atom abstraction by the 5′-deoxyadenosyl radical in ThiH. Heterolytic and homolytic Cα–Cβ bond cleavage events are shown. Both pathways converge in the ultimate formation of glyoxylate and cresol.

at the C8 and C12 positions (R = ethyl, methyl, *n*-propyl, iso-butyl, or neo-pentyl groups),[147] and it has been demonstrated that these side chains are derived from methylation reactions involving SAM.[148]

Two gene products belonging to the radical SAM family (BchQ and BchR) act as methyltransferases in the modification of BChl pigments. BchQ was shown to modify the C8 position with BchR modifying the C12 position.[147] Methyltransferase mutant cell lines in *C. tepidum* produced pigments that varied from those observed in wild-type cells, and it was concluded that the C8 and C12 modifications contribute to the optimization of the light-harvesting ability of the BChl c_F pigment. In addition, absence of the modifications appeared to affect pigment packing within the chlorosome. Methylations at the C8 and C12 positions likely require radical activation of the carbon atoms and a role for a putative 5′-deoxyadenosyl radical seems plausible, although at this time no detailed mechanistic information is available for these enzymes. Regardless, the available evidence shows that the presence of these radical SAM methyltransferase enzymes is critical for achieving wild-type growth phenotype levels in *C. tepidum*. Given that green sulfur bacteria have evolved to live in low-light environments, the methyltransferases provide vital modifications that enable these organisms to flourish in these niches.

8.17.3.6   Synthesis of Complex Metal Cofactors

8.17.3.6.1   *Biosynthesis of the FeMoco of nitrogenase*

Biological fixation of N_2 into NH_3 is a vital process to the global nitrogen cycle and is necessary for sustaining life on earth. The fixation of nitrogen into ammonia follows according to the equation $N_2 + 8e^- + 16MgATP + 8H^+ \rightarrow 2NH_3 + H_2 + 16MgADP + 16P_i$.[149] The molybdenum-containing nitrogenase enzymes of N_2-fixing bacteria are primarily responsible for catalyzing this reaction, although vanadium and iron-only nitrogenase enzymes also exist. Nitrogenase enzymes of the molybdenum class contain one of the most complex metalloclusters characterized to date: an iron and molybdenum cofactor (FeMoco) comprised of

seven Fe, one Mo, nine S, one homocitrate molecule, and one unidentified atom that is most likely C, N, or O. The synthesis of FeMoco requires the participation of 10 gene products (those of *nifB*, *nifS*, *nifU*, *nifE*, *nifN*, *nifV*, *nifQ*, *nifZ*, *nihH*, *nifD*, and *nifK*) in a remarkably complex series of reactions.[150–152]

Recent advancements in the *in vitro* synthesis of FeMoco have defined the minimal system necessary to achieve formation of this complex cofactor.[153,154] The identification from sequence annotation of NifB as a putative radical SAM protein based on the presence of an N-terminal CX_3CX_2C motif opened up the exciting possibility that radical SAM chemistry may be required for FeMoco synthesis. In addition to the SAM-binding motif, NifB contains nine additional conserved cysteine residues. Subsequent purification and characterization of the *nifB* gene product revealed that the as-isolated protein bound 12 iron atoms per dimer and could be reconstituted to contain 18 iron atoms per dimer. The protein showed UV–visible spectroscopic features characteristic of [Fe–S] clusters, which showed bleaching of the ligand-metal charge transfer (LMCT) features upon addition of sodium dithionite.[153] Moreover, *in vitro* assays showed that FeMoco synthesis required SAM, Fe^{2+}, S^{2-}, molybdate, and homocitrate in an NifB-dependent fashion. NifB activity was observed to be highly sensitive to O_2, and the addition of the SAM analog *S*-adenosylhomocysteine inhibited FeMoco synthesis. These observations coupled with the NifB-dependent incorporation of ^{55}Fe into apodinitrogenase suggest that NifB synthesizes a metal cluster that is some type of FeMoco precursor.

The C-terminal domain of NifB is a NifX-like domain found in proteins capable of binding FeMoco and NifB cofactor (NifB-co). It has been proposed that this domain may either (1) act as a scaffold for the SAM-dependent formation of NifB-co via the actions of the N-terminal radical SAM domain or (2) bind and transfer the NifB-co to the next set of proteins in FeMoco synthesis.[153] Only additional biochemical studies will shed light on the exact role of SAM in NifB-co synthesis and clarify if the actions of this radical SAM protein result in the incorporation of the unidentified C, N, or O atom into FeMoco.

8.17.3.6.2 *Biosynthesis of the H-cluster of hydrogenase*

Hydrogenase enzymes are responsible for H_2 metabolism in microorganisms and catalyze the reaction $2H^+ + 2e^- \rightleftharpoons H_2$. Microorganisms that express hydrogenase enzymes are able to utilize H_2 as a source of reducing equivalents for downstream processes that yield energy and by producing H_2 they can regenerate reduced electron carrier proteins like ferredoxin that accumulate during fermentation.[155,156] Three types of hydrogenase enzymes exist: [FeFe], [NiFe], and [Fe] hydrogenase enzymes, so named for their metal cluster content. [FeFe] hydrogenases are restricted to Bacteria and Eucarya, and contain an active site metal cluster, known as the H-cluster, which is composed of a [4Fe–4S] cubane linked to a 2-Fe unit coordinated by CN^-, CO, and a unique bridging dithiolate ligand (**Figure 11**). The composition of the H-cluster was determined
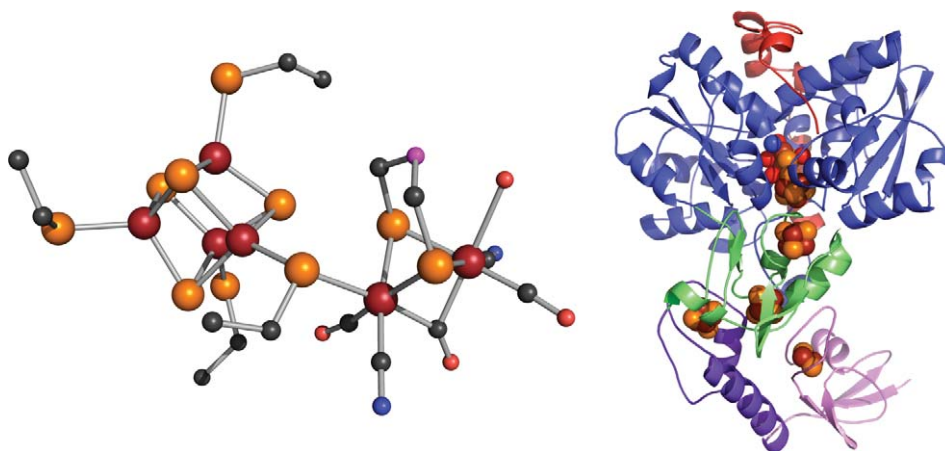
Figure 11   (Left) The H-cluster of [FeFe] hydrogenases. Atoms are colored maroon (Fe), orange (S), black (C), red (O), blue (N), and pink (unidentified). (Right) *Cp*I [FeFe] hydrogenase. The accessory F-cluster domains are two ferrodoxin-like [4Fe–4S] cubanes (green), a ferredoxin-like [2Fe–2S] cluster (light purple), and a histidine coordinated [4Fe–4S] cluster (dark purple). The catalytic domain containing the H-cluster is colored blue and the C-terminus of the protein (shown in red) forms a cap over the active site cavity.

through X-ray structural analysis of the [FeFe] hydrogenase enzymes from *Clostridium pasteurianum* (*Cp*I)[157] and *Desulfovibrio desulfuricans*.[158] Each enzyme is comprised of the active site H-cluster domain (**Figure 11**, blue) and accessory Fe–S cluster domains (F-clusters) believed to function in the shuttling of electrons to the H-cluster.[156] The F-cluster domains in *Cp*I HydA contain two ferrodoxin-like [4Fe–4S] cubanes (**Figure 11**, green), a ferredoxin-like [2Fe–2S] cluster (light purple), and a histidine coordinated [4Fe–4S] cluster (dark purple). The [4Fe–4S] cubane of the H-cluster is coordinated by protein-derived cysteine thiolates, and is bridged to the 2-Fe subcluster by a cysteinyl thiolate (Cys503, **Figure 11**). Cys503 is the only ligand linking the 2-Fe subcluster to HydA; other ligands include five diatomic CO and CN⁻ molecules and a nonprotein dithiolate.

As free molecules, CO and CN⁻ are toxic to metal-containing enzymes due to their ability to act as π-acids and form strong metal–ligand bonds. These molecules are synergistic ligands, undergoing metal-to-ligand back bonding, which acts to stabilize metals in low oxidation states. The 2-Fe subcluster in as-isolated HydA has been described as a low spin $S = 1/2$ state with an Fe(II)/Fe(I) pair,[159,160] and the distal Fe atom has been proposed to be the site for H_2 oxidation and H^+ reduction.[161] Also of significant outstanding interest is the nature of the bridging dithiolate ligand. It was initially proposed that this ligand may be propanedithiolate (PDT); however, this was revised to dithiomethylamine (DTMA) given the ability of a secondary amine functionality to serve as a proton donor/acceptor.[162] Recently, this assignment was challenged in a density functional theory (DFT) study analyzing O, CH_2, NH, NH_2^+, and S as respective bridgehead atoms.[160] The latter study determined that dithiomethylether (DTME) provided the lowest energy stabilization and gave the least deviation from H-cluster crystal structure coordinates.

Studies analyzing mutants incapable of H_2 production in the green algae *Chlamydomonas reinhardtii* determined that the gene products of *hydEF* and *hydG*, which are present in all organisms containing [FeFe] hydrogenases, are critical for obtaining mature HydA.[163] While *hydEF* exists as a gene fusion product in *C. reinhardtii*, these genes can be independently arranged in other organisms.[164] Further studies have shown that coexpression of HydE, HydF, and HydG from the anaerobic soil bacterium *Clostridium acetobutylicum* with the [FeFe] hydrogenase structural gene product from three different microbial sources results in the formation of active HydA.[165] Sequence annotation identified the HydE and HydG proteins as belonging to the radical SAM superfamily of enzymes, given the presence of the canonical CX_3CX_2C sequence motif, whereas HydF contains an N-terminal GTPase domain comprised of the Walker A P-loop and Walker B Mg^{2+} binding motifs, and five putative C-terminal Fe–S cluster ligands. Preliminary characterization of HydE, HydF, and HydG from *Thermotoga maritima* corroborated the functional inferences made from genomic annotation and demonstrated that all three proteins could bind Fe–S clusters, and showed that upon chemical reconstitution the HydE and HydG enzymes could bind a maximum of two [4Fe–4S] clusters.[166–168] Further work has demonstrated an *in vitro* methodology for determining activation of HydA heterologously expressed in a genetic background devoid of HydE, HydF, and HydG (HydA$^{\Delta EFG}$).[169] Activation of HydA$^{\Delta EFG}$ occurs only when it is simultaneously coexpressed with all three maturase proteins (HydE, HydF, and HydG), indicating that a protein-derived intermediate in H-cluster biosynthesis is formed under these conditions and is then transferred to the *hydA* gene product. Insightful results obtained from single amino acid point mutations showed that the site-differentiated SAM binding cluster in both HydE and HydG, the P-loop motif, and [Fe–S] binding motif of HydF were all essential for the maturation of the HydA enzyme in whole-cell extracts.[165]

These experimental observations led to a hypothetical biosynthesis pathway proposed for H-cluster biosynthesis (**Figure 12**).[170] DFT studies of the energetic steps accompanying radical decomposition of glycine led to the proposal that the two radical SAM enzymes HydE and/or HydG are responsible for the formation of a dithiolate-bridged 2-Fe cluster, by the alkylation of the sulfides on a [2Fe–2S] rhomb, in analogy to the reactions carried out by BioB and LipA. This reaction may be carried out by HydG alone, as recent data suggest that the second [Fe–S] cluster bound by HydE plays no catalytic role, and may instead either serve as a reservoir for iron and sulfide or may be an artifact from an ancient HydE enzyme.[168] In a second step, HydE and/or HydG are proposed to generate a glycine radical, which then interacts with the iron centers of the dithiolate-bridged 2-Fe cluster, decomposing into CO and CN⁻ ligands. HydF was proposed to serve as a scaffolding protein in this process, with cluster translocation from HydF to immature HydA resulting in the formation of holo [FeFe] hydrogenase enzyme.[170] Subsequent to this proposal, the analysis of purified HydE, HydF, and HydG proteins from coexpressions in which all three proteins were present (HydE^{FG}, HydF^{EG}, and
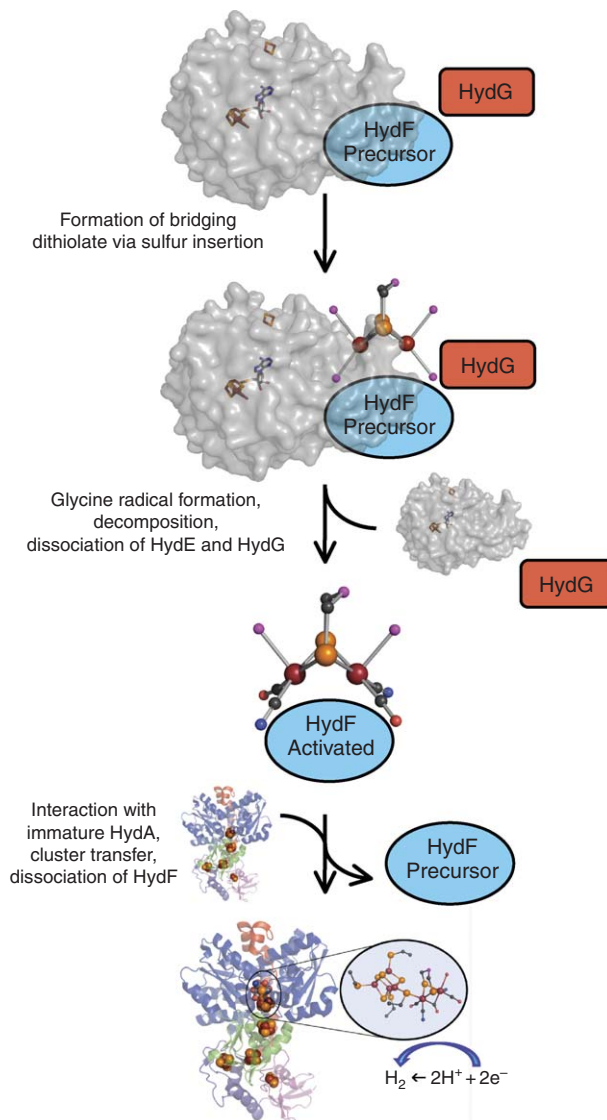
Figure 12 Proposed biosynthetic pathway for the H-cluster of [FeFe] hydrogenases. The crystal structure of HydE (3cix.pdb) is shown as a transparent surface plot to highlight the bound [Fe–S] clusters and SAM. In the proposed mechanism, HydE and HydG build a putative H-cluster precursor on the protein scaffold HydF. The H-cluster precursor is believed to be highly similar to the 2-Fe subcluster of the H-cluster. It is not known whether HydE and HydG act sequentially or simultaneously on HydF. Regardless, the activated HydF protein has been shown to activate the immature HydA protein, which is believed to be comprised of the accessory [Fe–S] clusters and the cubane portion of the H-cluster. Following cluster transfer, HydF is presumably returned to its precursor state where it may potentially undergo another activation event that may require the assistance of iron–sulfur cluster assembly proteins. Color coding for the atoms depicted in [Fe–S] cluster formation is as follows: maroon (iron), orange (sulfur), black (carbon), blue (nitrogen), red (oxygen), and pink (unidentified or unknown). Domains of the [FeFe] structural protein are colored differentially, with blue indicating the catalytic domain that houses the H-cluster.

$HydG^{EF}$) revealed that as-purified HydF from this genetic background ($HydF^{EG}$) mediates activation of $HydA^{\Delta EFG}$, whereas HydF expressed in the absence of HydE and HydG ($HydF^{\Delta EG}$) was not able to effect activation.[171] Importantly, the activation of $HydA^{\Delta EFG}$ by $HydF^{EG}$ does not require the presence of HydE and HydG during the activation process, nor does it require the addition of exogenous small molecules. Together, these data strongly suggest that a 2-Fe subcluster is formed on HydF by the actions of HydE and HydG and

that this H-cluster precursor is then transferred to HydA to generate the active holoenzyme. Additional biochemical experiments will allow for the elucidation of the specific reactions catalyzed by HydE and HydG, the identification of the substrates for these enzymes, as well as the characterization of the cluster harbored by HydF^{EG} that is transferred to HydA$^{\Delta EFG}$. Lastly, the role of the GTPase functionality of HydF during the maturation process is currently under investigation. An interesting new report has demonstrated that HydG catalyzes the cleavage of tyrosine, in a reaction reminiscent of ThiH (8.17.3.5.3), although the exact role of the degradation product(s) in the [FeFe] maturation process remain to be elucidated.[172]

8.17.3.7   Modification of tRNA

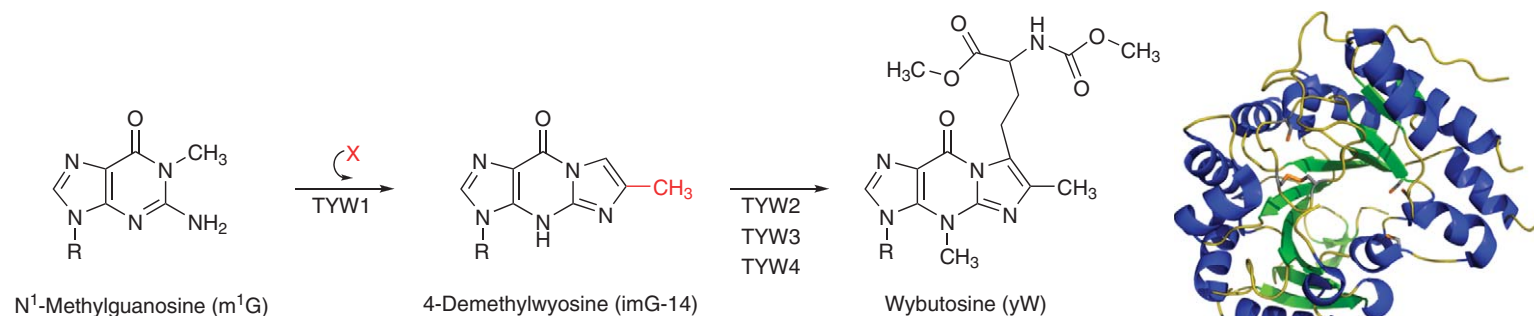#### 8.17.3.7.1   Biosynthesis of wybutosine

Tricyclic hypermodified nucleosides are found in archaeal and eukaryotic tRNAs and are frequently observed at position 34 (wobble base) or position 37 (adjacent to the anticodon). Position 37 typically contains a hypermodified nucleoside such as N^6-threonylcarbamoyladenosine (t^6A), 2-methylthio-N^6-isopentenyl-adenosine (ms^2i^6A-37), or wybutosine (yW).[173] yW and its derivatives occur at position 37 in archaeal and eukaryotic phenylalanine tRNA (tRNA$_{Phe}$). The modifications serve to maintain the correct translational reading frame via hydrophobic interactions, which reinforce codon–anticodon pairing and prevent incorrect Watson–Crick base-pairing. Studies have shown that unmodified tRNA leads to translational defects that have been implicated in different pathological states.[173]

The biosynthesis of yW from its guanine precursor in *Saccharomyces cerevisiae* involves five gene products. TRM5 uses SAM as a methyl donor and methylates G37 of tRNA$_{Phe}$ producing N^1-methylguanosine (m^1G). TYW1, identified through sequence annotation as a radical SAM enzyme, utilizes m^1G as a substrate and catalyzes ring cyclization by incorporation of a two-carbon fragment between the N^5-amine and N^8-methyl groups, producing 4-demethylwyosine (imG-14) (**Figure 13**)**. The source of the two-carbon fragment is heretofore unidentified. The imG-14 product of TYW1 catalysis is subsequently alkylated with α-amino-α-carboxypropyl (derived from the methionine portion of SAM) by TYW2 producing yW-86. TYW3 then catalyzes SAM-dependent methylation of the guanine ring of yW-86, producing yW-72. The final step in yW biosynthesis involves SAM-dependent methylation of the α-carboxylate group and methoxycarbonylation of the α-amino group of yW-72, as catalyzed by TYW4.[173]

TYW1 from *S. cerevisial* is comprised of an N-terminal flavodoxin domain and a C-terminal radical SAM domain that contains the canonical CX_3CX_2C motif. BLAST searches revealed that every archaeal genome except *Haloquadratum walsby* contained a homolog of the *S. cerevisial* radical SAM domain, but the N-terminal flavodoxin domain was discovered to have no archaeal homologs, suggesting that archaea utilize a different redox partner.[174] Recently, the crystal structure of the archaeal TYW1 enzymes from *Methanocaldococcus jannaschii* and *Pyrococcus horikoshii* have been determined (**Figure 13**).[174,175] The three-dimensional structure shows that these TYW1 enzymes share an incomplete $(\alpha/\beta)_6$ TIM barrel, as has been observed for other radical SAM enzymes. In addition to the site-differentiated CX_3CX_2C [Fe–S] binding site, the sequences and structures of these enzymes reveal the presence of three additionally conserved cysteine residues that may be responsible for binding an additional [Fe–S] cluster. Unfortunately, difficulties were encountered with crystallizing the enzymes with a full complement of [Fe–S] clusters, preventing definitive cluster content assignment at this time. Regardless, both enzymes were shown to exhibit some electron density in the CX_3CX_2C motif site, and the *P. horikoshii* structure showed density in the second site as well. Mutational studies of the cysteine residues corresponding to the second [Fe–S] cluster binding site in *S. cerevisial* TYWI indicated that this alternate [Fe–S] cluster is required for catalysis.[175] Both structures exhibit a large positively charged patch on the protein surface near the putative active site pocket, which presumably functions in binding the phosphate backbone of the anticodon stem of the tRNA substrate, aiding in the positioning of position 37 from the stem–loop. More definitive characterization of the [Fe–S] cluster content of TYW1 should help to shed light on the mechanistic details of this modification reaction.

8.17.3.7.2   Methylthiolation of tRNA

As mentioned above, ms^2i^6A is one of the modifications observed at position 37 in tRNA molecules. Unlike yW synthesis, the formation of ms^2i^6A involves only two enzymes. In the first reaction catalyzed by MiaA, an isopentenyl group is added to the N^6 position of adenosine, yielding the product N^6-isopentenyl-adenosine

Figure 13 (Left) Pathway for the synthesis of wybutosine (R = ribose sugar). The radical SAM enzyme TYW1 catalyzes the carbon insertion/cyclization reaction using an unknown substrate molecule termed X. The carbon atoms derived from this molecule are shown in red. Complete synthesis of yW requires three additional modifications, subsequently catalyzed in three separate SAM-dependent reactions by TYW2, TYW3, and TYW4. (Right) Crystal structure of the archaeal TYW1 enzyme from *Pyrococcus horikoshii* (PDB code: 2yx0). Secondary structural elements are colored blue for α-helices, green for β-sheets, and olive for random coil. The cysteine residues comprising the CX$_3$CX$_2$C motif are shown on the left-hand side of the barrel structure, while the cysteine residues comprising the putative second [Fe–S] binding site are shown on the right-hand side of the β-barrel.
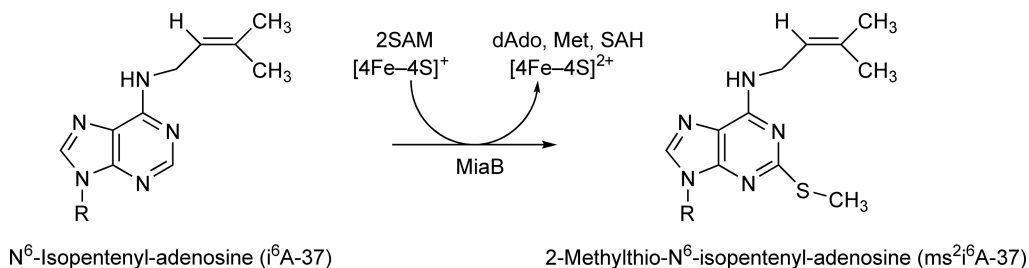
Figure 14 Reaction catalyzed by the radical SAM enzyme MiaB in 2-methylthio-N^6-isopentenyl-adenosine formation (R = ribose sugar). MiaB catalyzes the thiolation and methylation of the N^6-isopentenyl-adenosine precursor, apparently consuming two molecules of SAM in the process. SAH, *S*-adenosylhomocysteine.

(i^6A-37). In the second and final step of ms^2i^6A synthesis, catalyzed by the radical SAM enzyme MiaB, the substrate i^6A-37 undergoes both a sulfur insertion and a methylation reaction, thus forming ms^2i^6A (**Figure 14**).

The MiaB enzyme from *T. maritima* has been well characterized and contains, in addition to the SAM binding [4Fe–4S] cluster, an additional N-terminal [4Fe–4S] cluster bound by three conserved cysteine ligands. The clusters appear to be separated by 12–20 Å and the radical SAM cluster has a lower redox potential than the N-terminal [Fe–S] cluster.[176] Labeling experiments with [^{3}H$_3$C]-SAM conclusively demonstrated that SAM serves as the source of the methyl group in the ms^2i^6A product.[177] Moreover, selenium-substituted MiaB comprised of [Fe–Se] clusters showed selenium incorporation into mse^2i^6A, demonstrating that MiaB, and not SAM, serves as the source of sulfur atom in the product. Activity assays of wild-type and mutant MiaB proteins suggest that both [4Fe–4S] clusters are required for ms^2i^6A production.[176] A mechanism has been put forth whereby the thiolation and methylation of one molecule of i^6A-37 require the consumption of two molecules of SAM.[177] In this proposal, the first molecule of SAM is reductively cleaved, presumably generating a radical at the C2 position of the modified adenosine ring, at which position a sulfur atom is subsequently incorporated. The second molecule of SAM acts as the source of the methyl group. While the mechanistic details remain to be experimentally confirmed, available biochemical data tentatively indicate that the N-terminal cluster may serve as the source of the sulfur atom that is incorporated into the product. Only further studies will indicate whether the accessory cluster in MiaB plays the same role as the [2Fe–2S] cluster in BioB.

8.17.3.8 Formylglycine Generation: AtsB

Organosulfate ester cleavage is catalyzed by sulfatase enzymes, resulting in the formation of the corresponding alcohol and inorganic sulfate.[178] Three classes of sulfatase enzymes have been discovered to date and are differentiated by the types of cofactors utilized for catalysis. Group I enzymes, termed arylsulfatases, are found in archaea, bacteria, and eukaryotes and require either a calcium or a magnesium ion and an FGly cofactor. Eukaryotic sulfatase enzymes are involved in a multitude of processes, such as hydrolysis of sulfate ester linkages in steroids, lipids, and glycosaminoglycans.

The FGly cofactor is contained within a highly conserved signature sequence (C/S-X-P/A-S/X-R-X-X-X-L/X-T/X-G/X-R/X) in which the initial Cys or Ser residue is modified, resulting in the further subclassification of Cys-type or Ser-type sulfatases.[179,180] Crystal structures of the arylsulfatase enzymes from human and bacterial sources reveal that the FGly cofactor is present as a geminal diol. Proposed mechanisms for organosulfate ester cleavage involve nucleophilic attack on the sulfur atom of substrate organosulfate by one of the alcohol groups of FGly, resulting in sulfate-bound FGly and release of the corresponding alcohol. Elimination of the sulfate via deprotonation of the remaining cofactor alcohol group then generates the formylglycine aldeyhyde, which is hydrated to the geminal diol.[181,182]

Two pathways for FGly formation exist. The first occurs in higher eukaryotes by an unknown mechanism involving sulfatase modifying factor 1 in a process requiring O_2 and reducing equivalents.[183] The second pathway, catalyzed by AtsB in prokaryotes, is oxygen independent. Studies of the AtsB enzyme from *Klebsiella pneumonia* identified this protein as belonging to the radical SAM superfamily and demonstrated that SAM was required to form an active AtsB–AtsA complex.[184] Moreover, mutation of the conserved CX_3CX_2C residues

prevented AtsB from generating FGly, and it was proposed that the 5′-deoxyadenosyl radical generated upon SAM cleavage was responsible for initiating FGly formation via abstraction of a hydrogen atom from Cβ of Ser72. AtsB has a total of 13 cysteine residues and *in vitro* spectroscopic characterization of the purified protein shows that AtsB contains two $[4Fe-4S]^{2+}$ clusters per monomer, with reconstitution yielding three bound $[4Fe-4S]^{2+}$ clusters.[185] AtsB was shown to undergo multiple turnover events and high-performance liquid chromatography (HPLC) and matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) analyses demonstrated that the rate of formation of 5′-deoxyadenosine is equivalent to the rate of FGly production.

A mechanism for AtsB has been proposed whereby the target seryl residue becomes poised for modification via its coordination through the hydroxyl group to a vacant site on one of the irons of one of the accessory clusters. This coordination acts to decrease the pK_a for alcohol deprotonation and allows for inner-sphere electron transfer to the [Fe–S] cluster. Abstraction of hydrogen from Cβ by the deoxyadenosyl radical, followed by inner-sphere electron transfer and radical recombination, results in AtsA formylglycine production.[185] Lastly, available evidence suggests that Cys-type and Ser-type anaerobic sulfatase enzymes likely function by similar mechanisms, which can be rationalized as coordination of either serine or cysteine to an [Fe–S] cluster. Intriguingly, the recent report on two anaerobic sulfatase maturase enzymes from *Bacteroides thetaiotaomicron* suggests that these are dual-substrate enzymes able to oxidize both cysteine and serine residues to formylglycine, although substrate preferences may exist.[186] Additional studies will hopefully shed light on the factors controlling substrate preference and the potential coordination of cysteine and serine substrates to accessory [Fe–S] clusters.

8.17.3.9 A Radical SAM Dehydrogenase in Butirosin Biosynthesis: BtrN

Aminoglycosides that contain 2-deoxystreptamine constitute a significant group of antibiotic agents that are used to combat bacterial and protozoal infections.[187] One such aminoglycosidic antibiotic, known as butirosin, is synthesized by way of the radical SAM enzyme BtrN, which oxidizes an alcohol group to a carbonyl in the synthesis of 2-deoxystreptamine. We have already observed an analogous chemical reaction in the formation of the AtsA-formylglycine residue by AtsB, where the $[4Fe-4S]^{2+}$ cluster is proposed to undergo two sequential reduction and oxidation steps during turnover: the first event results in SAM cleavage and a $[4Fe-4S]^{2+}$ cluster, which is then capable of accepting the electron from the organic radical intermediate, resulting in formylglycine generation.[185] A similar mechanism has been proposed for BtrN (see below).[188] Thus, despite the lack of sequence homology between BtrN and AtsB, it appears as though these enzymes share the ability to catalyze the same chemical transformation. Accordingly, they have been defined as 'radical SAM dehydrogenases'.[189]

In the case of butirosin synthesis, BtrN has been shown to bind a $[4Fe-4S]^+$ cluster and the enzyme uses SAM as a substrate in the oxidation of an alcohol group in 2-deoxy-scyllo-inosamine (DOIA) to a ketone group in the 3-amino-2,3-dideoxy-scyllo-inosose (amino–DOI) product.[188] BtrN shows remarkable substrate specificity for DOIA, able to differentiate DOIA from a variety of sugars and cyclitols. Moreover, the use of deuterated substrate showed that hydrogen atom abstraction from DOIA contributes to the kinetic isotope effect but is not the rate-limiting step – ketone formation appears to be the rate-limiting step. In a similar manner to the proposed AtsB mechanism, the 5′-deoxyadenosyl radical abstracts a hydrogen atom from the DOIA substrate, generating a DOIA radical intermediate. The oxidized $[4Fe-4S]^{2+}$ cluster then accepts the electron from the organic radical to produce amino-DOI. Changes to the EPR spectrum of reduced BtrN in the presence of SAM are consistent with direct coordination of SAM to the $[4Fe-4S]^+$ cluster. Addition of DOIA resulted in *g*-value shifts in the reduced BtrN spectrum, providing evidence that the substrate interacts with the [4Fe–4S] cluster in a manner that may promote the reductive cleavage of SAM. Importantly, an EPR signal characteristic of an organic radical was observed upon incubation of BtrN with SAM, reductant, and either deuterated or nondeuterated DOIA allowing for the definitive assignment of the C3 DOIA radical species and providing strong support for the mechanism outlined above.[189] Additional studies may help delineate the effects of DOIA on the [Fe–S] cluster, probing whether or not the substrate molecule alters the redox potential of the cluster to promote SAM cleavage. Structure determination of BtrN should explain the molecular basis for DOIA specificity.

The involvement of radical SAM enzymes in the synthesis of antibiotic components can be expanded to include DesII. DesII is a SAM-dependent deaminase responsible for synthesizing a key intermediate in desosamine biosynthesis. Desosamine, 3-(dimethylamino)-3,4,6-trideoxyhexose, is a component of certain

antibiotic compounds synthesized by *Streptomyces venezuelae.*[190] DesII is proposed to generate the 5′-deoxyadenosyl radical, which abstracts a hydrogen atom from the C3 sugar of thymidine diphosphate-4-amino-4,6-dideoxy-D-glucose. This may result in radical-induced deamination followed by readdition of ammonia, forming an aminol radical. If SAM acts as a cofactor, then the recycling of the hydrogen atom from deoxyadenosine to the aminol radical may subsequently result in elimination of ammonia to yield the 3-keto-6-deoxyhexose product.[190] Future work will determine if SAM is utilized as a cofactor or substrate in the deaminase reaction.

8.17.3.10   Radical Methylation Reactions: Utilization of Two Distinct Cofactors

As observed in the section discussing BtrN and DesII, radical SAM enzymes have already been shown to play a critical role in the synthesis of certain elements of antibiotic compounds. Moreover, we observed in the MiaB section that a radical SAM enzyme is responsible for catalyzing a methylation reaction in the synthesis of ms^{2}i^{6}A. Methyl transfer reactions in biology commonly utilize DNA methylase enzymes or SAM-dependent methyltransferases; the latter system has been touched upon in the multiple SAM-dependent methyl transfer reactions involved in yW synthesis. However, a unique methylation reaction has recently been proposed that invokes the use of both SAM and methylcobalamin.

Fosphomycin, fortimicin A, and bialaphos are three antibiotics that appear to utilize radical SAM enzymes in methylation reactions.[191] The enzymes that are involved in these putative methylation reactions (Fom3, Fms7, and BcpD) are comprised of a radical SAM domain in which the CX_3CX_2C motif is present and a B$_{12}$-like binding domain that accommodates methylcobalamin.[192] Unlike the methylation reaction catalyzed by MiaB, in which SAM is the source of the methyl group, methylcobalamin acts as the methyl donor during the synthesis of these respective antibiotics.[193–195] In the synthesis of fosfomycin (epoxypropylphosphonic acid), Fom3 cleaves SAM and the 5′-deoxyadenosyl radical is proposed to abstract the pro-*R* hydrogen atom from the alcohol group at the C2 position of the hydroxyethylphosphonate substrate. The resulting organic radical then reacts with the methylcobalamin moiety, which presumably is in close proximity to the [4Fe–4S] SAM binding cluster, generating the 2-hydroxypropylphosphonate (HPP) product, which then undergoes Fom4-catalyzed epoxidation to yield fosfomycin.[192] Additional biochemical studies will shed light on this proposed mechanism. Moreover, Fom3 structure determination not only provides direct mechanistic clues, in terms of defining the proximity of the radical SAM and methylcobalamin binding domains, but may also add to the proposed evolutionary relationship between the radical SAM and adenosylcobalamin enzyme families.

8.17.4   Insights from Structural Studies of Radical SAM Enzymes

The characterized radical SAM enzymes act on a variety of substrates in numerous biochemical pathways, including vitamin and cofactor biosynthesis, facilitation of lysine fermentation and biosynthesis of β-lysyl antibiotics, repairing UV-induced DNA damage, and generating protein-derived glycyl radicals. Each of the above functions involves widely different substrate molecules including examples from amino acids to DTB to DNA to protein. Identifying the structural elements that help to define the molecular factors governing substrate recognition is one of the important tasks in the field.

Structures of nine radical SAM enzymes have been solved to date (**Figures 9, 12, 13, and 15**). All display partial or full TIM barrel subunit folds whereby a barrel-like structure is formed out of β-strands that are surrounded by α-helices. The [4Fe–4S] cluster that binds SAM is located within the barrel. Of the solved structures, only BioB and HydE have a complete TIM barrel fold comprised of eight β-strands and eight α-helices $(\alpha/\beta)_8$. All other structures are $(\alpha/\beta)_6$ in composition, and it has been suggested that the most primitive members of the family, like ARR-AE, probably have $(\alpha/\beta)_4$ folds, possibly indicating the evolution for this subunit fold from $(\alpha/\beta)_2$ precursor building blocks.[196] Beyond the CX_3CX_2C motif, the conserved regions involved in SAM binding (GGE and GXIXGXXE motifs) are spread throughout the sequence and substantial sequence divergence occurs beyond strand $\beta 5$, meaning that the core of SAM radical proteins contains about 200 amino acids. The divergent C-terminal portion of these enzymes, which shows no sequence homology between subclasses of radical SAM proteins, is probably responsible for conferring substrate specificity.[196]

Interestingly, among the structurally characterized radical SAM enzymes with known substrates, the size of the enzyme tends to increase as the substrate size decreases. The structures in **Figure 15** show the monomeric X-ray crystal structures of BioB, HydE, HemN, LAM, MoaA, and PFL-AE.

8.17.4.1 Biotin Synthase

BioB catalyzes the conversion of DTB to biotin by inserting a sulfur atom between the nonactivated C6 and C9 carbons of DTB (see Section 8.17.3.3). The crystal structure of BioB shows that the enzyme has a full $(\alpha/\beta)_8$ TIM barrel fold.[70] Two Fe–S clusters are bound within the barrel and define the active site. The first cluster is the canonical [4Fe–4S] cluster coordinated by the CX_3CX_2C motif and is responsible for ligating SAM through the N/O chelation of the amino and carboxylate groups to the unique Fe. Consistent with spectroscopic results on BioB and PFL-AE, the sulfonium of SAM is not directly ligated to the [4Fe–4S] cluster, but is in close proximity to it, being $\approx 4\,\text{Å}$ away from the nearest cluster Fe. SAM is bound in an extended conformation across the top of the barrel, interacting with residues in several β-strands, which effectively serve to completely shield SAM from solvent, a necessary precaution for radical-based chemistry.
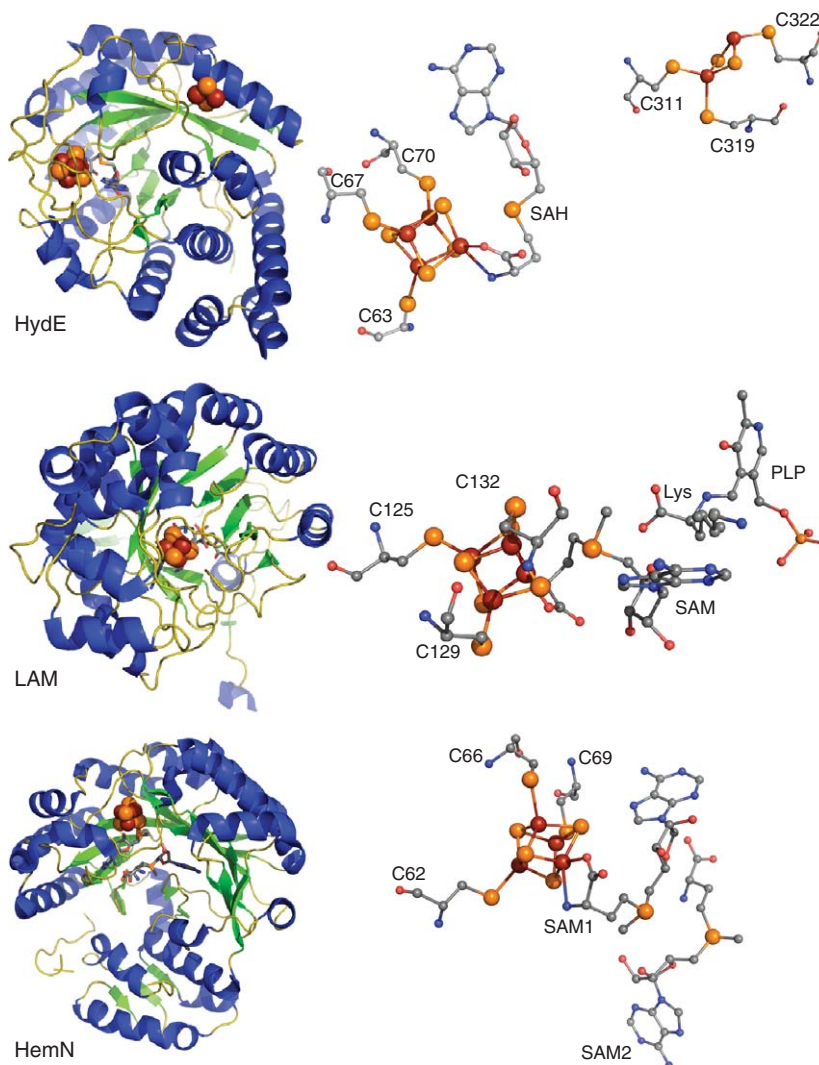
Figure 15 (Continued)
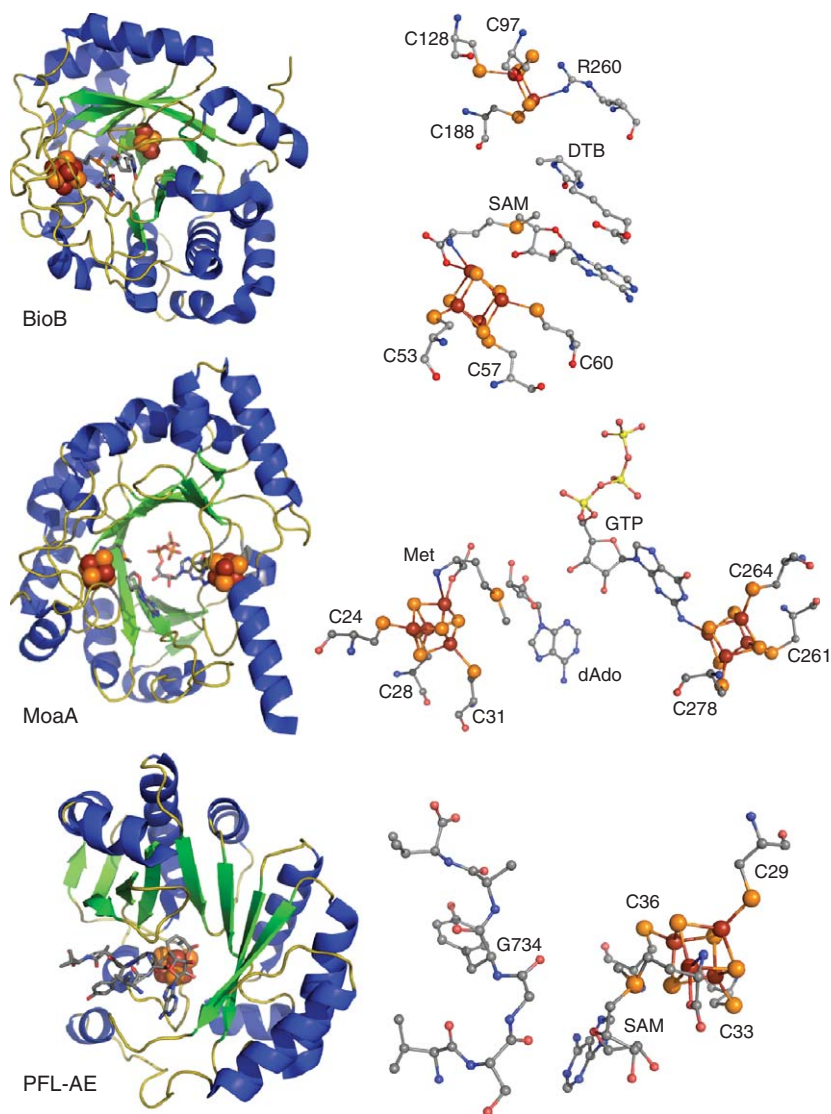
Figure 15 Radical SAM X-ray crystal structures. Left panels: HydE (3cix.pdb), LAM (2a5h.pdb), HemN (1olt.pdb), BioB (1r30.pdb), MoaA (2fb3.pdb), and PFL-AE (3cb8.pdb) are shown. Secondary structural elements colored according to α-helices (blue), β-sheets (green), and coils (olive). Right panels: Active site residues of interest are labeled. Atoms are colored gray (carbon), red (oxygen), blue (nitrogen), yellow (phosphorus), orange (sulfur), and maroon (iron). See text for discussion of active site properties and abbreviations. Note: In the HemN structure, only one conformation of each SAM molecule is shown for clarity, although both conformations are shown in the PDB coordinates due to disordered electron density maps, which may indicate that these molecules have some flexibility in the active site environment.

The second Fe–S cluster is a [2Fe–2S] cluster and is positioned $\approx 15\,\text{Å}$ from the [4Fe–4S] cluster on the opposite side of the barrel. The [2Fe–2S] cluster is ligated by Cys97, Cys128, Cys188, and Arg260. Coordination of a metal center by an arginine residue is striking and has little precedence in metallobiochemistry, and initially led to the suggestion that this residue played a critical function in modulating the properties of the cluster, especially given its strict conservation among BioB sequences from various organisms. Mutagenesis studies of this residue in the *E. coli* protein have since strongly suggested that it plays no role in the catalytic formation of biotin, indicating instead that its role may be linked to the acceptance of the cluster from the iron sulfur cluster (ISC) machinery *in vivo*.[13]

Substrate DTB is observed to bind in the core of the TIM barrel between SAM and the [2Fe–2S] cluster. DTB forms extensive van der Waals interactions with SAM, covering nearly 50% of its surface area, and explains why SAM binding is enhanced by a factor of 20 in the presence of DTB. Key features of these contact regions include the stacking of the DTB carboxylate tail against the adenine ring of SAM and the stacking of the ureido ring of DTB over the ribose of SAM. The bidentate coordination of DTB to Asn222 likely serves to orient DTB for hydrogen atom abstraction, with C6 positioned 4.1 Å and C9 positioned 3.9 Å from the 5′ carbon of SAM. Moreover, the [2Fe–2S] cluster is ideally suited to play a role in thiophane ring formation, as the closest bridging sulfur of the [2Fe–2*S*] cluster is 4.6 Å away from the C9 of DTB.

8.17.4.2   Pyruvate Formate Lyase-Activating Enzyme

The PFL-AE, as its name suggests, activates PFL, the enzyme that converts pyruvate and CoA into acetyl CoA and formate under anaerobic conditions in bacteria. PFL is a central metabolic enzyme that provides the sole source of acetyl CoA for the citric acid cycle during bacterial fermentation. GREs, including PFL and ARR, share a common subunit fold comprised of two sets of five-stranded β-sheets running in an antiparallel fashion. The critical glycine residue that must be primed for activation by the activating enzyme is located buried within the protein at the tip of the second of two β-finger motifs.[84,197] Monomeric PFL-AE (28 kDa in size) interacts with homodimeric PFL (170 kDa) and abstracts the C2 pro-*S* hydrogen atom from Gly734 of PFL thereby generating the catalytically pertinent glycyl radical species.[52] Gly734 is located $\approx$8 Å from the protein surface in PFL; this, coupled with the evidence for direct H· abstraction from G734, suggests that the formation of the glycyl radical must involve significant structural rearrangement. The recent X-ray crystal structure of PFL-AE complexed with a 7-mer peptide substrate provides unique insights into this protein–protein activation event (see below).

The X-ray crystal structure of the 245 amino acid PFL-AE is the smallest radical SAM structure currently available and therefore embodies the minimal protein machinery required to reductively cleave SAM, at least among the available repertoire of radical SAM structures. The structure of PFL-AE reveals that the enzyme has a partial $(\alpha/\beta)_6$ TIM barrel subunit fold, with the N-terminal end of the barrel (the bottom) being highly solvent exposed. The enzyme binds a site-differentiated [4Fe–4S] cluster with the CX_3CX_2C motif and in the peptide-bound structure, SAM is observed to coordinate to the unique iron site through the chelation of the amino and carboxylate groups. The 7-mer peptide substrate (RVSGYAV) binds in a bent conformation across the lateral opening of the barrel, efficiently sealing the active site from solvent, thus creating an environment for radical transfer. In the absence of substrate, SAM binds to the [4Fe–4S] cluster, although it does so with significant disorder for much of the electron density associated with the ribose and adenine portions of SAM. Peptide binding stabilizes SAM, locking it into a conformation that likely primes the molecule for reductive cleavage. This observation is interesting in light of the potential ability of substrate binding to control the unproductive cleavage of SAM; it is possible that *in vivo* SAM cleavage can only be directly coupled to glycyl radical generation.

Several PFL-AE van der Waals contacts to the RVSGYAV peptide are made upon binding, predominantly with the motif DGXGXR located on loop A. These interactions provide the keys to peptide specificity, peptide conformation, and orientation of the glycine in the active site. A Gly734 Cα to SAM C5′ distance of 4.1 Å is optimized by three specific interactions: the stacking of His37 over the adenine ring of SAM, orientation of the glycine residue, and Asp16 of the DGXGXR motif coupled with the absolutely conserved Asn38 that hydrogen bonds to the amino and carboxyl groups of Gly734. Comparison of the substrate-free and substrate-bound PFL-AE crystal structures reveals that the primary conformational change is associated with loop A containing the DGXGXR motif. The loop region must swing in an upward motion to make contact with the peptide and in doing so Asp16 is displaced by $\approx$10 Å. This large motion is likely critical to the reductive cleavage of SAM and activation of PFL, either by enabling a conformational change in PFL itself or by simply properly positioning the peptide in the active site.

The Gly734 radical is quite stable under anaerobic conditions and catalyzes several turnover events via an active site thiyl radical.[198] If exposed to O_2 however, the radical is destroyed resulting in irreversible cleavage of the polypeptide chain into two fragments, thereby permanently inactivating the enzyme. In this scenario, PFL functionality can be rescued in some organisms through the expression of a short protein that contains high sequence homology to the last 59 residues of PFL, which includes Gly734. This protein, termed YfiD, is

activated by PFL-AE and then associates with the cleaved PFL thus restoring its activity.[199] A portion of PFL corresponding to YfiD and termed the radical domain (RD) was used in docking studies to examine the potential interactions with the full PFL substrate. The best docking model has the glycine loop of the PFL RD oriented similarly to the peptide in the peptide-bound PFL-AE crystal structure, with a distance between the C5′ of SAM and the Cα of Gly734 of 4.6 Å. Binding of the RD of PFL in the PFL-AE active site in a similar manner to that illustrated in the docking model would require movement of the RD out of the core of PFL; remarkably, this presumed conformational change does not require the input of an outside energy source, such as hydrolysis of ATP or GTP. Modeling suggests that the rotation of the RD on PFL can occur about a single hinge point enabling the glycine to become accessible for binding by PFL-AE. Some outstanding issues relate to how the glycyl radical is properly shielded from solvent upon its return to PFL and what molecular forces drive the movement of loop A on PFL-AE and the movement of the glycine finger loop out of PFL. It is possible that these two events are coupled to one another.

8.17.4.3   Lysine 2,3-Aminomutase

LAM catalyzes the first step of fermentation involving the conversion of L-lysine to L-β-lysine,[33] via an isomerization reaction of the α-amino group of lysine migrating to the β-carbon in β-lysine. LAM crystallizes as a homotetramer of essentially identical subunits and solution studies suggest that the enzyme may exist in dimeric, tetrameric, and hexameric states. The monomeric subunit of the tetramer is comprised of an $(\alpha/\beta)_6$ TIM barrel fold. Se-adenosyl-L-selenomethionine (SeSAM) is observed to directly ligate the [4Fe–4S] cluster, which is bound in close proximity to the L-α-lysine, bound as the external aldimine adduct of PLP. The PLP cofactor is bound at the N-terminus of the crescent-shaped TIM barrel. Nine residues form a total of 10 hydrogen bonds with the phosphate group, the phenolic oxygen, and the pyridine nitrogen of the PLP cofactor. Three active site hydrogen bonds hold lysine in place, formed between the ε-amino group and Asp293 and Asp330 and between the substrate carboxylate and Arg134. The interactions between lysine, active site residues, and the PLP cofactor act to position the α- and β-carbons of lysine for the NH$_2$ and H-atom interchange at these positions that accompanies the conversion to β-lysine. The structure reveals that the pro-*R* hydrogen at the β-carbon position is ≈ 4 Å from the reactive C5′ position of SAM, ideally positioned for proton abstraction following the reductive cleavage of SAM.[200]

8.17.4.4   MoaA: Molybdenum Cofactor Biosynthesis Enzyme

MoaA has an $(\alpha/\beta)_6$ subunit fold and binds two [4Fe–4S] clusters by three-cysteine motifs.[71] The clusters are approximately 17 Å apart and are on opposite sides of a hydrophilic channel that runs through the center of the TIM barrel. The N-terminal [4Fe–4S] cluster is responsible for anchoring SAM and binds SAM through the methionine as a N/O chelate at the unique Fe site **Figure 15**. The C-terminal [4Fe–4S] cluster is responsible for anchoring 5′-GTP via coordination from the purine N1 nitrogen and the exocyclic amino group of the guanine ring to the unique Fe site.[201] The triphosphate of 5′-GTP is stabilized by 12 hydrogen bonds mainly contributed by R17, R71, R192, K69, and K163, which effectively neutralize the negative charges of the phosphate groups. Unlike typical NTP binding enzymes, no cation binding site exists in MoaA and 5′-GTP binding is actually reduced by approximately 40% in the presence of Mg^{2+}. While MoaA is unable to use other purine nucleotides as substrates in the production of precursor Z, other NTPs can bind to MoaA, although with decreased affinity. The structure reveals that binding of other NTPs is primarily governed by the large number of electrostatic interactions between the protein and phosphate groups.

Interestingly, the crystal structure of the R17/266/268A triple mutant was solved and it was revealed that SAM could still coordinate to the N-terminal [4Fe–4S] cluster, although the mutant was no longer able to reductively cleave SAM. This suggests that the redox potential of the [4Fe–4S] cluster was perturbed by the reduced positive charge in the R $\rightarrow$ A triple mutant, which speaks toward the importance of the protein microenvironment in defining the properties of Fe–S clusters.

While much remains to be learned about the mechanism of precursor Z synthesis, the structure of MoaA obtained with SAM and GTP in the presence of Na$_2$S$_2$O$_4$ provided some insight into how the mechanism of radical transfer from the [4Fe–4S]$^+$ cluster to SAM to 5′-GTP may occur. The structure reveals that SAM has

been cleaved into methionine and dAdo. The reactive C5 group of dAdo was found to be disordered, indicating that the bond cleavage event renders a radical group that has greater mobility in the active site, presumably increasing its ability to react with the radical acceptor species. Possible sites of H-atom abstraction on 5′-GTP are the imidazole C8 atom and either the C2′ or C3′ position of the ribose. The intermolecular distances between these groups and the C5 carbon of the putative radical on dAdo in the structure are between 5.3 and 8.1 Å. These distances are too far for direct radical transfer and indicate that significant rearrangements of substrate and/or product must occur, probably involving movement of the ribose moiety. This is in unique contrast to the structures of the radical SAM enzymes LAM and BioB, where the reactive C5 carbon group of SAM lies within 3.8–3.9 Å of the site of substrate H-atom abstraction and no movement of substrate or product must be invoked for the radical transfer reaction to occur. Regardless, the H-atom abstraction from either the C2′ or C3′ position of the sugar appears feasible and would be similar to the reaction catalyzed by ribonucleotide reductases.

The most difficult chemical step in precursor Z synthesis is the cleavage of the C2′–C3′ bond with subsequent insertion of the C8 atom. While the mechanistic details of this process are unknown, it appears that strict radical transfer from dAdo to 5′-GTP must occur with significant rearrangement of active site molecules in a manner that prevents unwanted side reactions. The C-terminal [4Fe–4S] cluster may play an important, heretofore unidentified role during catalysis, either functioning as an electron donor/acceptor system or coordinating reaction intermediates.

8.17.4.5   HydE: Hydrogenase H-Cluster Biosynthesis Enzyme

The H-cluster of [FeFe] hydrogenase is one of the most remarkable clusters in metallobiochemistry, as it contains a [4Fe–4S] cubane linked to a [2Fe] unit decorated by cyanide, carbon monoxide, and a unique bridging dithiolate ligand. The biosynthesis of this cluster requires the actions of three proteins, HydE, HydF, and HydG. HydF is a GTPase and HydE and HydG are putative radical SAM enzymes. The exact function(s) of HydE and HydG in the maturation of [FeFe] hydrogenase is unknown.

The structure of HydE from *T. maritima* is very similar to that of BioB from *E. coli*, displaying a relatively small r.m.s. deviation of 2.5 Å between the Cα atoms of both enzymes.[168] HydE, like BioB, belongs to a subset of radical SAM enzymes having a full $(\alpha/\beta)_8$ TIM barrel subunit fold. HydE binds a site-differentiated [4Fe–4S] cluster responsible for binding SAM. Depending on the conditions used to grow crystals, a second Fe–S cluster can bind to the enzyme as well. This [2Fe–2S] cluster is located 20 Å away from the conserved [4Fe–4S] cluster, near the protein surface, and is ligated by three cysteine residues and a water molecule. Interestingly, the positions of the two clusters in HydE are very similar to the positions of the two clusters in MoaA. Unlike the second Fe–S cluster in MoaA, which directs the unique Fe site toward the center of the active site cavity, the unique Fe site of the [2Fe–2S] cluster is directed toward the solvent medium, suggesting that it does not play a role in substrate binding. Moreover, mutations of the three cysteine ligands of the [2Fe–2S] cluster have no adverse effects on the ability to produce active [FeFe] hydrogenase, providing evidence against the notion that this cluster serves as a source of sulfur in the synthesis of the bridging dithiolate compound. The exciting role of HydE in H-cluster assembly and the identification of the substrate molecule that presumably accepts the radical from SAM cleavage remain to be elucidated.

8.17.4.6   Oxygen-Independent Coproporphyrinogen Oxidase

The structure of HemN is an $(\alpha/\beta)_6$ TIM barrel fold that binds a [4Fe–4S] cluster and two molecules of SAM.[68] Interestingly, the [4Fe–4S] cluster is bound in a highly polarized environment within the barrel, with half of the cluster surrounded by hydrophobic residues and the other half of the cluster surrounded by hydrophilic residues. The first molecule of SAM (SAM1) coordinates the unique iron atom of the [4Fe–4S] cluster as seen in other structures. The second molecule of SAM (SAM2) binds in a position adjacent to the [4Fe–4S] coordinated SAM1 and appears to be held in place by five amino acids that are conserved to varying degrees among HemN sequences. A hydrophilic pocket lined with charged residues sits symmetrically adjacent to the sulfoniums of SAM1 and SAM2 and appears to be positioned near the propionate side chains in the

substrate-bound model, making this the putative binding site for the small molecule X that serves as the unknown additional electron acceptor during the HemN-catalyzed reaction.

The role of SAM2 in the HemN mechanism is unclear; however, it is apparent that direct reduction of SAM2 by the [4Fe–4S] cluster is not feasible. It has been proposed that substrate coproporphyrinogen binding may induce rotation around the C5′–S bond of SAM2, moving the sulfonium of SAM2 closer to SAM1. In this case, the first electron transfer event to SAM1 (resulting in SAM cleavage) might be immediately transferred to SAM2, with the resulting SAM2-derived dAdo radical then abstracting a hydrogen atom from the β-carbon of the substrate propionate side chain to produce an allylic, coproporphyrinogen substrate radical.[202] Decarboxylation then would ensue, with the subsequent electron transfer to the unknown small molecule oxidizing agent occurring. Subsequent re-reduction of the [4Fe–4S]$^{2+}$ cluster would result in cleavage of SAM1 and the second decarboxylation of substrate, this time via the SAM1-derived dAdo radical. This proposed sequence of events is consistent with the appearance of two molecules of SAM in the structure, and with the previously discussed biochemical evidence for the catalytic relevance of SAM2.

One potential advantage of this proposed mechanism is that it would at least partially negate some of the energetic cost associated with the decarboxylation event occurring only at the site of SAM1. If decarboxylation occurred only at the SAM1 site, then following the first reductive cleavage of SAM1 and subsequent oxidative decarboxylation of the first propionate side chain, another molecule of SAM would have to replace the methionine and dAdo products of SAM cleavage, undoubtedly requiring significant rearrangement of active site molecules and partial or full release of the porphyrinogen intermediate. An alternative mechanism involves SAM2 serving as the initial electron acceptor following the first decarboxylation reaction (initiated via the cleavage of SAM1), thereby generating the second 5′-deoxyadenosyl radical. The latter mechanism bypasses the need to transfer an electron from SAM1 to SAM2 for the initial cleavage event, although it still requires the presence of an unidentified electron acceptor molecule for the second decarboxylation reaction.

8.17.4.7   HMP-P Synthase: A Link between Radical SAM and Adenosylcobalamin Enzymes?

The structure of HMP-P synthase is shown in **Figure 9** and discussed in Section 8.17.3.5.3.[138] Three-dimensional structural homolog searches with the ThiC homodimeric structure performed with DALI (http://www.ebi.ac.uk/dali/) revealed no similar structures. A similar query performed with the catalytic domain 2 (the $(\alpha/\beta)_8$ TIM barrel) identified BioB, LAM, and HemN as structural homologs of ThiC. Additionally, the adenosylcobalamin-dependent enzymes glutamate mutase from *Clostridium cochlearium* and lysine 5,6-aminomutase from *Clostridium sticklandii* were identified.[138] Moreover, as pointed out by Chatterjee *et al.*,[138] HMP-P synthase sequence alignments reveal that enzymes derived from anaerobes are shorter than enzymes from aerobes, lacking as many as 160 N-terminal and C-terminal residues combined. HMP-P synthases from aerobic organisms appear to have additional amino acids that are C-terminal to the [4Fe–4S] SAM binding cluster. A similar observation is also noted for residues on the N-terminal portion of the enzyme family, collectively suggesting that the longer sequences present in HMP-P synthase enzymes from aerobic organisms act in part to shield the active site environment from the radical quencher O_2.

Both radical SAM and adenosylcobalamin enzyme families share some variation of $(\alpha/\beta)_8$ TIM barrel folds. This observation coupled with the fact that both enzyme families generate the highly reactive 5′deoxyadenosyl radical strongly suggests that some type of evolutionary relationship exists. The identification of HMP-P synthase as an enzyme that shares the $(\alpha/\beta)_8$ TIM barrel fold but has a separate SAM binding domain denoted by a unique [Fe–S] cluster binding motif provides additional support for the relationship between these enzyme families. An explanation offered by Chatterjee *et al.*[138] relates to a common ancestor that may have had separate SAM and adenosylcobalamin binding domains that could interchange. Subsequent gene fusion events would result in two enzymes that bound either SAM or adenosylcobalamin, and through time could give rise to the domain architectures we currently observe for the adenosylcobalamin-, CX_3CX_2C SAM-dependent and CX_2CX_4C SAM-dependent enzyme families.

8.17.5   Conclusions

The radical SAM enzymes catalyze remarkably diverse reactions, presumably starting however with the same basic mechanistic steps: coordination of SAM to the unique iron of a site-differentiated $[4Fe–4S]^{2+}$ cluster, reduction of the cluster to the catalytically active $[4Fe–4S]^{+}$ state, and then inner-sphere electron transfer from the cluster to the SAM sulfonium, which promotes homolytic S–C5′ bond cleavage to generate the 5′-deoxyadenosyl radical intermediate believed to be common to all enzymes in the superfamily. At this point, the reaction mechanisms diverge, giving rise to the broad range of reactions summarized in this chapter, as well as many more that have not been discussed here and that we have yet to discover. While the adenosylcobalamin-dependent enzymes were once considered the prototypical (in fact, the only) enzymes utilizing adenosyl radical intermediates, it is now quite clear that the radical SAM superfamily is significantly more widespread in biology and more diverse in its chemistry. Although most of the radical SAM enzymes studied in detail to date are from facultative or strict anaerobic microbes, enzymes in this superfamily are found throughout the phylogenetic kingdom, including in critical biochemical pathways in higher eukaryotes. Our understanding of this super-family is rapidly accelerating, and newly discovered functions of radical SAM enzymes and new chemical mechanisms will continue to fuel excitement in the field for years to come.

Acknowledgements

Abbreviations

AdoCbl or B$_{12}$	adenosylcobalamin
AdoMet or SAM	*S*-adenosylmethionine
AE	activating enzyme
BioB	biotin synthase
dAdo	5′-deoxyadenosine
HydE	hydrogenase H-cluster biosynthesis enzyme
LAM	lysine aminomutase
LipA	lipoyl synthase
MoaA	molybdenum cofactor biosynthesis enzyme
PFL-AE	pyruvate formate lyase activating enzyme
SPL	spore photoproduct lyase

References

1. H. Beinert, *J. Biol. Inorg. Chem.* **2000**, *5*, 2–15.
2. D. Johnson; D. Dean; A. Smith; M. Johnson, *Annu. Rev. Biochem.* **2005**, *74*, 247–281.
3. F. Anet, *Curr. Opin. Chem. Biol.* **2004**, *8*, 654–659.
4. R. Eck; M. Dayhoff, *Science* **1966**, *152* (3720), 363–366.
5. W. Martin; J. Baross; D. Kelley; M. Russell, *Nat. Rev. Microbiol.* **2008**, *6* (11), 805–814.
6. M. Russell, *Acta Biotheor.* **2007**, *55* (2), 133–179.
7. G. Wächtershäuser, *Microbiol. Rev.* **1988**, *52* (4), 452–484.
8. G. Wächtershäuser, *Prog. Biophys. Mol. Biol.* **1992**, *58* (2), 85–201.
9. C. Huber; G. Wächtershäuser, *Science* **1997**, *276* (5310), 245–247.
10. C. Huber; G. Wächtershäuser, *Science* **1998**, *281* (5377), 670–672.
11. C. Huber; G. Wächtershäuser, *Science* **2006**, *314* (5799), 630–632.
12. Y. Zu; M. Couture; D. Kolling; A. Crofts; L. Eltis; J. Fee; J. Hirst, *Biochemistry* **2003**, *42* (42), 12400–12408.
13. R. Broach; J. Jarrett, *Biochemistry* **2006**, *45* (47), 14166–14174.

14. E. M. Walters; R. Garcia-Serres; G. N. L. Jameson; D. A. Glauser; F. Bourquin; W. Manieri; P. Schürmann; M. K. Johnson; B. H. Huynh, *J. Am. Chem. Soc.* **2005**, *127*, 9612–9624.
15. H. Beinert; R. Holm; E. Munck, *Science* **1997**, *277* (5326), 653–659.
16. C. L. Drennan; J. W. Peters, *Curr. Opin. Struct. Biol.* **2003**, *13* (2), 220–226.
17. N. Nair; E. Schreiner; D. Marx, *J. Am. Chem. Soc.* **2006**, *128* (42), 13815–13826.
18. J. Stubbe, *Curr. Opin. Struct. Biol.* **2000**, *10* (6), 731–736.
19. P. Nordlund; P. Reichard, *Annu. Rev. Biochem.* **2006**, *75*, 681–706.
20. J. Knappe; E. Bohnert; W. Brummer, *Biochim. Biophys. Acta* **1965**, *107* (3), 603–605.
21. J. Knappe; J. Schacht; W. Möckel; T. Höpner; H. Vetter; R. Edenharder, *Eur. J. Biochem.* **1969**, *11*, 316–327.
22. J. Knappe; J. Schacht; W. Mockel; T. Hopner; H. J. Vetter; R. Edenharder, *Eur. J. Biochem.* **1969**, *11* (2), 316–327.
23. J. Knappe; H. P. Blaschkowski; P. Gröbner; T. Schmitt, *Eur. J. Biochem.* **1974**, *50*, 253–263.
24. J. Knappe; H. P. Blaschkowski, *Meth. Enzymol.* **1975**, *41*, 508–518.
25. J. Knappe; T. Schmitt, *Biochem. Biophys. Res. Commun.* **1976**, *71*, 1110–1117.
26. J. Knappe; F. A. Neugebauer; H. P. Blaschkowski; M. Gänzler, *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 1332–1335.
27. H. Conradt; M. Hohmann-Berger; H.-P. Hohmann; H. P. Blaschkowski; J. Knappe, *Arch. Biochem. Biophys.* **1984**, *228* (1), 133–142.
28. W. Rödel; W. Plaga; R. Frank; J. Knappe, *Eur. J. Biochem.* **1988**, *177*, 153–158.
29. K. K. Wong; B. W. Murray; S. A. Lewisch; M. K. Baxter; T. W. Ridky; L. Ulissi-DeMario; J. W. Kozarich, *Biochemistry* **1993**, *32*, 14102–14110.
30. J. B. Broderick; R. E. Duderstadt; D. C. Fernandez; K. Wojtuszewski; T. F. Henshaw; M. K. Johnson, *J. Am. Chem. Soc.* **1997**, *119*, 7396–7397.
31. R. Eliasson; M. Fontecave; H. Jörnvall; M. Krook; E. Pontis; P. Reichard, *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3314–3318.
32. E. Mulliez; M. Fontecave; J. Gaillard; P. Reichard, *J. Biol. Chem.* **1993**, *268* (4), 2296–2299.
33. T. P. Chirpich; V. Zappia; R. N. Costilow; H. A. Barker, *J. Biol. Chem.* **1970**, *245* (7), 1778–1789.
34. M. Moss; P. A. Frey, *J. Biol. Chem.* **1987**, *262* (31), 14859–14862.
35. J. Baraniak; M. L. Moss; P. A. Frey, *J. Biol. Chem.* **1989**, *264* (3), 1357–1360.
36. R. M. Petrovich; F. J. Ruzicka; G. H. Reed; P. A. Frey, *J. Biol. Chem.* **1991**, *266* (12), 7656–7660.
37. M. D. Ballinger; P. A. Frey; G. H. Reed, *Biochemistry* **1992**, *31* (44), 10782–10789.
38. M. D. Ballinger; G. H. Reed; P. A. Frey, *Biochemistry* **1992**, *31* (4), 949–953.
39. C. H. Chang; M. D. Ballinger; G. H. Reed; P. A. Frey, *Biochemistry* **1996**, *35*, 11081–11084.
40. I. Sanyal; G. Cohen; D. H. Flint, *Biochemistry* **1994**, *33*, 3625–3631.
41. I. Sanyal; K. J. Gibson; D. H. Flint, *Arch. Biochem. Biophys.* **1996**, *326* (1), 48–56.
42. E. C. Duin; M. E. Lafferty; B. R. Crouse; R. M. Allen; I. Sanyal; D. H. Flint; M. K. Johnson, *Biochemistry* **1997**, *36*, 11811–11820.
43. D. Guianvarc'h; D. Florentin; B. T. S. Bui; F. Nunzi; A. Marquet, *Biochem. Biophys. Res. Commun.* **1997**, *236*, 402–406.
44. F. Escalettes; D. Florentin; B. T. S. Bui; D. Lesage; A. Marquet, *J. Am. Chem. Soc.* **1999**, *121* (15), 3571–3578.
45. R. W. Busby; J. P. M. Schelvis; D. S. Yu; G. T. Babcock; M. A. Marletta, *J. Am. Chem. Soc.* **1999**, *121*, 4706–4707.
46. J. R. Miller; R. W. Busby; S. W. Jordan; J. Cheek; T. F. Henshaw; G. W. Ashley; J. B. Broderick; J. E. Cronan, Jr.; M. A. Marletta, *Biochemistry* **2000**, *39*, 15166–15178.
47. H. J. Sofia; G. Chen; B. G. Hetzler; J. F. Reyes-Spindola; N. E. Miller, *Nucleic Acids Res.* **2001**, *29* (5), 1097–1106.
48. R. Banerjee, *Chem. Rev.* **2003**, *103* (6), 2083–2094.
49. J. J. Baker; T. C. Stadtman, Aminomutases. In *B12. Biochemistry and Medicine*; D. Dolphin, Ed.; Wiley-Interscience: New York, 1982; Vol. 2, pp 203–232.
50. P. A. Frey, *FASEB J.* **1993**, *7*, 662–670.
51. P. A. Frey; O. T. Magnusson, *Chem. Rev.* **2003**, *103*, 2129–2148.
52. M. Frey; M. Rothe; A. F. V. Wagner; J. Knappe, *J. Biol. Chem.* **1994**, *269* (17), 12432–12437.
53. M. L. Moss; P. A. Frey, *J. Biol. Chem.* **1990**, *265*, 18112–18115.
54. J. Cheek; J. B. Broderick, *J. Biol. Inorg. Chem.* **2001**, *6*, 209–226.
55. M. Fontecave; E. Mulliez; S. Ollagnier-de-Choudens, *Curr. Opin. Chem. Biol.* **2001**, *5*, 506–511.
56. J. T. Jarrett, *Curr. Opin. Chem. Biol.* **2003**, *7*, 174–182.
57. J. B. Broderick, Iron–Sulfur Clusters in Enzyme Catalysis. In *Comprehensive Coordination Chemistry II*; J. A. McCleverty, T. J. Meyer, Eds.; Elsevier: Oxford, 2004; Vol. 8, pp 739–757.
58. G. D. Markham *S*-Adenosylmethionine. In *In Nature Encyclopedia of Life Sciences*; John Wiley & Sons: London, 2003, http://mrw. interscience.wiley.com/emrw/9780470015902/home/EditorsContributors.html
59. M. K. Johnson, Iron–Sulfur Proteins. In *Encyclopedia of Inorganic Chemistry*; R. B. King, Ed.; John Wiley & Sons: Chichester, 1994; pp 1896–1915.
60. J. Grimshaw, Electrochemistry of the sulfonium group. In *Chemistry of the Sulphonium Group*; C. J. M. Stirling, Ed.; John Wiley & Sons: Chichester, 1981; pp 141–155.
61. R. M. Petrovich; F. J. Ruzicka; G. H. Reed; P. A. Frey, *Biochemistry* **1992**, *31*, 10774–10781.
62. M. Fontecave; M. Atta; E. Mulliez, *Trends Biochem. Sci.* **2004**, *29* (5), 243–249.
63. R. Banerjee, *Chemistry and Biochemistry of B12*; Wiley-Interscience: New York, 1999.
64. R. Banerjee, *Biochemistry* **2001**, *40* (21), 6191–6198.
65. K. B. Song; P. A. Frey, *J. Biol. Chem.* **1991**, *266* (12), 7651–7655.
66. R. Külzer; T. Pils; R. Kappl; J. Hüttermann; J. Knappe, *J. Biol. Chem.* **1998**, *273* (9), 4897–4903.
67. J. Tamarit; C. Gerez; C. Meier; E. Mulliez; A. Trautwein; M. Fontecave, *J. Biol. Chem.* **2000**, *275* (21), 15669–15675.
68. G. Layer; J. Moser; D. W. Heinz; D. Jahn; W.-D. Schubert, *EMBO J.* **2003**, *22* (23), 6214–6224.
69. S. Ollagnier-de-Choudens; Y. Sanakis; K. S. Hewitson; P. Roach; E. Münck; M. Fontecave, *J. Biol. Chem.* **2002**, *277* (16), 13449–13454.
70. F. Berkovitch; Y. Nicolet; J. T. Wan; J. T. Jarrett; C. L. Drennan, *Science* **2004**, *303*, 76–79.
71. P. Hänzelmann; H. Schindelin, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (35), 12870–12875.
72. C. Krebs; W. E. Broderick; T. F. Henshaw; J. B. Broderick; B. H. Huynh, *J. Am. Chem. Soc.* **2002**, *124* (6), 912–913.

73. C. J. Walsby; W. Hong; W. E. Broderick; J. Cheek; D. Ortillo; J. B. Broderick; B. M. Hoffman, *J. Am. Chem. Soc.* **2002**, *124* (12), 3143–3151.
74. C. J. Walsby; D. Ortillo; W. E. Broderick; J. B. Broderick; B. M. Hoffman, *J. Am. Chem. Soc.* **2002**, *124* (38), 11270–11271.
75. D. Chen; C. Walsby; B. M. Hoffman; P. A. Frey, *J. Am. Chem. Soc.* **2003**, *125*, 11788–11789.
76. T. F. Henshaw; J. Cheek; J. B. Broderick, *J. Am. Chem. Soc.* **2000**, *122*, 8331–8332.
77. K. Lieder; S. Booker; F. J. Ruzicka; H. Beinert; G. H. Reed; P. A. Frey, *Biochemistry* **1998**, *37*, 2578–2585.
78. D. Padovani; F. Thomas; A. X. Trautwein; E. Mulliez; M. Fontecave, *Biochemistry* **2001**, *40* (23), 6713–6719.
79. E. L. Colichman; D. L. Love, *J. Org. Chem.* **1953**, *18*, 40–46.
80. F. D. Saeva; B. P. Morgan, *J. Am. Chem. Soc.* **1984**, *106*, 4121–4125.
81. G. T. Hinckley; P. A. Frey, *Biochemistry* **2006**, *45*, 3219–3225.
82. S. C. Wang; P. A. Frey, *Biochemistry* **2007**, *46* (45), 12889–12895.
83. A. F. V. Wagner; M. Frey; F. A. Neugebauer; W. Schäfer; J. Knappe, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 996–1000.
84. A. Becker; K. Fritz-Wolf; W. Kabsch; J. Knappe; S. Schultz; A. F. V. Wagner, *Nature* **1999**, *6* (10), 969–975.
85. T. Selmer; A. J. Pierik; J. Heider, *Biol. Chem.* **2005**, *386*, 981–988.
86. P. I. Andrei; A. J. Pierik; S. Zauner; L. C. Andrei-Selmer; T. Selmer, *Eur. J. Biochem.* **2004**, *271*, 2225–2230.
87. O. T. Magnusson; G. H. Reed; P. A. Frey, *J. Am. Chem. Soc.* **1999**, *121*, 9764–9765.
88. O. T. Magnusson; G. H. Reed; P. A. Frey, *Biochemistry* **2001**, *40*, 7773–7782.
89. N. S. Lees; D. Chen; C. Walsby; E. Behshad; P. A. Frey; B. M. Hoffman, *J. Am. Chem. Soc.* **2006**, *128*, 10145–10154.
90. F. J. Ruzicka; P. A. Frey, *Biochim. Biophys. Acta* **2007**, *1774*, 286–296.
91. B. Tse Sum Bui; D. Florentin; A. Marquet; R. Benda; A. X. Trautwein, *FEBS Lett.* **1999**, *459*, 411–414.
92. S. Ollagnier-de Choudens; Y. Sanakis; K. S. Hewitson; P. Roach; J. E. Baldwin; E. Münck; M. Fontecave, *Biochemistry* **2000**, *39*, 4165–4173.
93. N. B. Ugulava; B. R. Gibney; J. T. Jarrett, *Biochemistry* **2000**, *39*, 5206–5214.
94. S. Ollagnier-de-Choudens; E. Mulliez; K. S. Hewitson; M. Fontecave, *Biochemistry* **2002**, *41*, 9145–9152.
95. N. B. Ugulava; B. R. Gibney; J. T. Jarrett, *Biochemistry* **2001**, *40* (28), 8343–8351.
96. N. B. Ugulava; K. K. Surerus; J. T. Jarrett, *J. Am. Chem. Soc.* **2002**, *124* (31), 9050–9051.
97. N. B. Ugulava; C. J. Sacanell; J. T. Jarrett, *Biochemistry* **2001**, *40* (28), 8352–8358.
98. B. Tse Sum Bui; R. Benda; V. Schünemann; D. Florentin; A. X. Trautwein; A. Marquet, *Biochemistry* **2003**, *42*, 8791–8798.
99. A. M. Taylor; C. Farrar; J. T. Jarrett, *Biochemistry* **2008**, *47*, 9309–9317.
100. E. Choi-Rhee; J. E. Cronan, *Chem. Biol.* **2005**, *12*, 461–468.
101. M. R. Reyda; R. Dippold; M. E. Dotson; J. T. Jarrett, *Arch. Biochem. Biophys.* **2008**, *471*, 32–41.
102. M. A. Hayden; I. Huang; D. E. Bussiere; G. W. Ashley, *J. Biol. Chem.* **1992**, *267* (14), 9512–9515.
103. S. Ollagnier-de Choudens; M. Fontecave, *FEBS Lett.* **1999**, *453*, 25–28.
104. R. M. Cicchillo; D. F. Iwig; A. D. Jones; N. M. Nesbitt; C. Baleanu-Gogonea; M. G. Souder; L. Tu; S. J. Booker, *Biochemistry* **2004**, *43*, 6378–6386.
105. R. M. Cicchillo; K.-H. Lee; C. Baleanu-Gogonea; N. M. Nesbitt; C. Krebs; S. J. Booker, *Biochemistry* **2004**, *43*, 11770–11781.
106. R. M. Cicchillo; S. J. Booker, *J. Am. Chem. Soc.* **2005**, *127*, 2860–2861.
107. P. Douglas; M. Kriek; P. Bryant; P. L. Roach, *Angew. Chem. Int. Ed. Engl.* **2006**, *45* (31), 5197–5199.
108. J. E. Donnellan, Jr.; R. B. Setlow, *Science* **1965**, *149*, 308–310.
109. J. E. Donnellan, Jr.; R. S. Stafford, *Biophys. J.* **1968**, *8*, 17–28.
110. A. J. Varghese, *Biochem. Biophys. Res. Commun.* **1970**, *38*, 484–490.
111. A. Sancar, *Chem. Rev.* **2003**, *103*, 2203–2237.
112. P. Setlow, *Annu. Rev. Microbiol.* **1988**, *42*, 319–338.
113. B. Setlow; D. Sun; P. Setlow, *J. Bacteriol.* **1992**, *174* (7), 2312–2322.
114. P. Setlow, *Trends Microbiol.* **2007**, *15* (4), 172–180.
115. T. Douki; B. Setlow; P. Setlow, *Photochem. Photobiol.* **2005**, *81* (1), 163–169.
116. W. L. Nicholson; B. Setlow; P. Setlow, *J. Bacteriol.* **1990**, *172* (12), 6900–6906.
117. S. C. Mohr; N. V. H. A. Sokolov; C. He; P. Setlow, *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 77–81.
118. C. S. Hayes; Z.-Y. Peng; P. Setlow, *J. Biol. Chem.* **2000**, *275* (45), 35040–35050.
119. A. Sancar, *Biochemistry* **1994**, *33*, 2–9.
120. P. Fajardo-Cavazos; C. Salazar; W. L. Nicholson, *J. Bacteriol.* **1993**, *175*, 1735–1744.
121. E. C. Freidberg; G. C. Walker; W. Siede, *DNA Repair and Mutagenesis*; American Society for Microbiology: Washington, DC, 1995.
122. R. E. Yasbin; D. Cheo; D. Bol, DNA Repair Systems. In *Bacillus subtilis and Other Gram Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*; A. L. Sonenshein, J. A. Hoch, R. Losick, Eds.; American Society for Microbiology: Washington, DC, 1993.
123. R. Rebeil; Y. Sun; L. Chooback; M. Pedraza-Reyes; C. Kinsland; T. P. Begley; W. L. Nicholson, *J. Bacteriol.* **1998**, *180* (18), 4879–4885.
124. R. Rebeil; W. L. Nicholson, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98* (16), 9038–9043.
125. J. Cheek; J. B. Broderick, *J. Am. Chem. Soc.* **2002**, *124* (12), 2860–2861.
126. J. M. Buis; J. Cheek; E. Kalliri; J. B. Broderick, *J. Biol. Chem.* **2006**, *281* (36), 25994–26003.
127. S. J. Kim; C. Lester; T. P. Begley, *J. Org. Chem.* **1995**, *60*, 6256–6257.
128. R. Nicewonger; T. P. Begley, *Tetrahedron Lett.* **1997**, *38* (6), 935–936.
129. R. A. Mehl; T. P. Begley, *Org. Lett.* **1999**, *1* (7), 1065–1066.
130. M. G. Friedel; O. Berteau; J. C. Pieck; M. Atta; S. Ollagnier-de-Choudens; M. Fontecave; T. Carell, *Chem. Commun.* **2006**, 445–447.
131. T. C. Chandra; S. C. Silver; E. Zilinskas; E. M. Shepard; W. E. Broderick; J. B. Broderick, *J. Am. Chem. Soc.* **2009**, *131* (7), 2420–2421.
132. G. Layer; K. Grage; T. Teschner; V. Schunemann; D. Breckau; A. Masoumi; M. Jahn; P. Heathcote; A. Trautwein; D. Jahn, *J. Biol. Chem.* **2005**, *280* (32), 29038–29046.
133. G. Layer; K. Grage; T. Teschner; V. Schünemann; D. Breckau; A. Masoumi; M. Jahn; P. Heathcote; A. X. Trautwein; D. Jahn, *J. Biol. Chem.* **2005**, *280*, 29038–29046.
134. G. Layer; A. J. Pierik; M. Trost; S. E. Rigby; H. K. Leech; K. Grage; D. Breckau; I. Astner; L. Jänsch; P. Heathcote; M. J. Warren; D. W. Heinz; D. Jahn, *J. Biol. Chem.* **2006**, *281*, 15727–15734.

135. B. Estramareix; S. David, *Biochim. Biophys. Acta* **1990**, *1035* (2), 154–160.
136. B. Lawhorn; R. Mehl; T. Begley, *Org. Biomol. Chem.* **2004**, *2* (17), 2538–2546.
137. N. Martinez-Gomez; D. Downs, *Biochemistry* **2008**, *47* (35), 9054–9056.
138. A. Chatterjee; Y. Li; Y. Zhang; T. Grove; M. Lee; C. Krebs; S. Booker; T. Begley; S. Ealick, *Nat. Chem. Biol.* **2008**, *4* (12), 758–765.
139. N. Martinez-Gomez; M. Robers; D. Downs, *J. Biol. Chem.* **2004**, *279* (39), 40505–40510.
140. R. Leonardi; S. Fairhurst; M. Kriek; D. Lowe; P. Roach, *FEBS Lett.* **2003**, *539* (1–3), 95–99.
141. M. Kriek; F. Martins; R. Leonardi; S. Fairhurst; D. Lowe; P. Roach, *J. Biol. Chem.* **2007**, *282* (24), 17413–17423.
142. R. Leonardi; P. Roach, *J. Biol. Chem.* **2004**, *279* (17), 17054–17062.
143. M. Kriek; F. Martins; M. Challand; A. Croft; P. Roach, *Angew. Chem. Int. Ed. Engl.* **2007**, *46* (48), 9223–9226.
144. J. Suzuki; D. Bollivar; C. Bauer, *Annu. Rev. Genet.* **1997**, *31*, 61–89.
145. N. Frigaard; D. Bryant, *Chlorosomes: Antenna Organelles in Photosynthetic Green Bacteria*; Springer: Berlin, Heidelberg, 2006; Vol. 2, pp 79–114.
146. G. Montano; B. Bowen; J. LaBelle; N. Woodbury; V. Pizziconi; R. Blankenship, *Biophys. J.* **2003**, *85* (4), 2560–2565.
147. C. Gomez Maqueo; N. Frigaard; D. Bryant, *J. Bacteriol.* **2007**, *189* (17), 6176–6184.
148. M. Huster; K. Smith, *Biochemistry* **1990**, *29* (18), 4348–4355.
149. R. Y. Igarashi; L. C. Seefeldt, *Crit. Rev. Biochem. Mol. Biol.* **2003**, *38* (4), 351–384.
150. P. Dos Santos; D. Dean; Y. Hu; M. Ribbe, *Chem. Rev.* **2004**, *104* (2), 1159–1173.
151. Y. Hu; A. Fay; C. Lee; J. Yoshizawa; M. Ribbe, *Biochemistry* **2008**, *47* (13), 3973–3981.
152. L. Rubio; P. Ludden, *J. Bacteriol.* **2005**, *187* (2), 405–414.
153. L. Curatti; P. Ludden; L. Rubio, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (14), 5297–5301.
154. L. Curatti; J. A. Hernandez; R. Y. Igarashi; B. Soboh; D. Zhao; L. M. Rubio, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (45), 17626–17631.
155. P. Vignais; B. Billoud; J. Meyer, *FEMS Microbiol. Rev.* **2001**, *25* (4), 455–501.
156. J. W. Peters, *Curr. Opin. Struct. Biol.* **1999**, *9* (6), 670–676.
157. J. Peters; W. Lanzilotta; B. Lemon; L. Seefeldt, *Science* **1998**, *282* (5395), 1853–1858.
158. Y. Nicolet; C. Piras; P. Legrand; C. Hatchikian; J. Fontecilla-Camps, *Structure* **1999**, *7* (1), 13–23.
159. J. Peters, *Curr. Opin. Struct. Biol.* **1999**, *9* (6), 670–676.
160. A. S. Pandey; T. V. Harris; L. J. Giles; J. W. Peters; R. K. Szilagyi, *J. Am. Chem. Soc.* **2008**, *130* (13), 4533–4540.
161. B. Lemon; J. Peters, *Biochemistry* **1999**, *38* (40), 12969–12973.
162. Y. Nicolet; A. L. de Lacey; X. Vernede; V. M. Fernanadez; E. C. Hatchikian; J. C. Fontecilla-Camps, *J. Am. Chem. Soc.* **2001**, *123*, 1596–1601.
163. M. C. Posewitz; P. W. King; S. L. Smolinski; L. Zhang; M. Seibert; M. L. Ghirardi, *J. Biol. Chem.* **2004**, *279* (24), 25711–25720.
164. A. Böck; P. W. King; M. Blokesch; M. C. Posewitz, Maturation of Hydrogenases. In *Advances in Microbial Physiology*; R. K. Poole, Ed.; Elsevier, Academic Press: New York, 2006; Vol. 51, pp 1–57.
165. P. W. King; M. C. Posewitz; M. L. Ghirardi; M. Seibert, *J. Bacteriol.* **2006**, *188* (6), 2163–2172.
166. J. Rubach; X. Brazzolotto; J. Gaillard; M. Fontecave, *FEBS Lett.* **2005**, *579* (22), 5055–5060.
167. X. Brazzoloto; J. K. Rubach; J. Gaillard; S. Gambarelli; M. Atta; M. Fontecave, *J. Biol. Chem.* **2006**, *281* (2), 769–774.
168. Y. Nicolet; J. Rubach; M. Posewitz; P. Amara; C. Mathevon; M. Atta; M. Fontecave; J. Fontecilla-Camps, *J. Biol. Chem.* **2008**, *283* (27), 18861–18872.
169. S. E. McGlynn; S. S. Ruebush; A. Naumov; L. E. Nagy; A. Dubini; P. W. King; J. B. Broderick; M. C. Posewitz; J. W. Peters, *J. Biol. Inorg. Chem.* **2007**, *12* (4), 443–447.
170. J. Peters; R. Szilagyi; A. Naumov; T. Douglas, *FEBS Lett.* **2006**, *580* (2), 363–367.
171. S. E. McGlynn; E. M. Shepard; M. A. Winslow; A. Naumov; K. S. Duschene; M. C. Posewitz; W. E. Broderick; J. B. Broderick; J. W. Peters, *FEBS Lett.* **2008**, *582* (15), 2183–2187.
172. E. Pilet; Y. Nicolet; T. Douki; J. Fontecilla-Camps; M. Fontecave, *FEBS Lett.* **2009**, *583* (3), 506–511.
173. A. Noma; Y. Kirino; Y. Ikeuchi; T. Suzuki, *EMBO J.* **2006**, *25*, 2142–2154.
174. S. Goto-Ito; R. Ishii; T. Ito; R. Shibata; E. Fusatomi; S. Sekine; Y. Bessho; S. Yokoyama, *Acta Crystallogr. D Biol. Crystallogr.* **2007**, *63* (10), 1059–1068.
175. Y. Suzuki; A. Noma; T. Suzuki; M. Senda; T. Senda; R. Ishitani; O. Nureki, *J. Mol. Biol.* **2007**, *372* (5), 1204–1214.
176. H. Hernandez; F. Pierrel; E. Elleingand; R. Garcia-Serres; B. Huynh; M. Johnson; M. Fontecave; M. Atta, *Biochemistry* **2007**, *46* (17), 5140–5147.
177. F. Pierrel; T. Douki; M. Fontecave; M. Atta, *J. Biol. Chem.* **2004**, *279* (46), 47555–47563.
178. S. Hanson; M. Best; C. Wong, *Angew. Chem. Int. Ed. Engl.* **2004**, *43* (43), 5736–5763.
179. B. Schmidt; T. Selmer; A. Ingendoh; K. von Figura, *Cell* **1995**, *82* (2), 271–278.
180. O. Berteau; A. Guillot; A. Benjdia; S. Rabot, *J. Biol. Chem.* **2006**, *281* (32), 22464–22470.
181. I. Boltes; H. Czapinska; A. Kahnert; R. von Bulow; T. Dierks; B. Schmidt; K. von Figura; M. Kertesz; I. Uson, *Structure* **2001**, *9* (6), 483–491.
182. G. Lukatela; N. Krauss; K. Theis; T. Selmer; V. Gieselmann; K. von Figura; W. Saenger, *Biochemistry* **1998**, *37* (11), 3654–3664.
183. T. Dierks; A. Dickmanns; A. Preusser-Kunze; B. Schmidt; M. Mariappan; K. von Figura; R. Ficner; M. Rudolph, *Cell* **2005**, *121* (4), 541–552.
184. Q. Fang; J. Peng; T. Dierks, *J. Biol. Chem.* **2004**, *279* (15), 14570–14578.
185. T. Grove; K. Lee; J.St. Clair; C. Krebs; S. Booker, *Biochemistry* **2008**, *47* (28), 7523–7538.
186. A. Benjdia; S. Subramanian; J. Leprince; H. Vaudry; M. Johnson; O. Berteau, *J. Biol. Chem.* **2008**, *283* (26), 17815–17826.
187. S. Magnet; J. Blanchard, *Chem. Rev.* **2005**, *105* (2), 477–498.
188. K. Yokoyama; M. Numakura; F. Kudo; D. Ohmori; T. Eguchi, *J. Am. Chem. Soc.* **2007**, *129* (49), 15147–15155.
189. K. Yokoyama; D. Ohmori; F. Kudo; T. Eguchi, *Biochemistry* **2008**, *47* (34), 8950–8960.
190. P. Szu; X. He; L. Zhao; H. Liu, *Angew. Chem. Int. Ed. Engl.* **2005**, *44* (41), 6742–6746.
191. P. A. Frey; A. D. Hegeman; F. J. Ruzicka, *Crit. Rev. Biochem. Mol. Biol.* **2008**, *43* (1), 63–88.
192. R. Woodyer; G. Li; H. Zhao; W. van der Donk, *Chem. Commun. (Camb.)* **2007**, *4*, 359–361.
193. H. Seto; S. Imai; T. Tsuruoka; A. Satoh; M. Kojima; S. Inouye; T. Sasaki; N. Otake, *J. Antibiot. (Tokyo)* **1982**, *52* (10), 1719–1721.

194. S. Okumura; T. Deguchi; H. Marumo, *J. Antibiot. (Tokyo)* **1981**, *34* (10), 1360–1362.
195. T. Kuzuyama; T. Hidaka; K. Kamigiri; S. Imai; H. Seto, *J. Antibiot. (Tokyo)* **1992**, *45* (11), 1812–1814.
196. Y. Nicolet; C. L. Drennan, *Nucleic Acids Res.* **2004**, *32* (13), 4015–4025.
197. D. T. Logan; J. Andersson; B.-M. Sjöberg; P. Nordlund, *Science* **1999**, *283*, 1499–1504.
198. P. A. Frey, *Annu. Rev. Biochem.* **2001**, *70*, 121–148.
199. A. F. V. Wagner; S. Schultz; J. Bomke; T. Pils; W. D. Lehmann; J. Knappe, *Biochem. Biophys. Res. Commun.* **2001**, *285*, 456–462.
200. B. W. Lepore; F. J. Ruzicka; P. A. Frey; D. Ringe, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (39), 13819–13824.
201. P. Hänzelmann; H. Schindelin, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (18), 6829–6834.
202. G. Layer; A. Pierik; M. Trost; S. Rigby; H. Leech; K. Grage; D. Breckau; I. Astner; L. Jansch; P. Heathcote; M. Warren; D. Heinz; D. Jahn, *J. Biol. Chem.* **2006**, *281*, 15727–15734.

Biographical Sketches

Eric M. Shepard was born in 1977. He received a BS degree in chemistry from Rocky Mountain College. He studied copper- and TPQ-containing amine oxidases under Dr. David M. Dooley at Montana State University, where he was supported by an NSF IGERT fellowship on complex biological systems. He received his Ph.D. degree in biochemistry from MSU and is currently a postdoctoral research associate under Dr. Joan B. Broderick. His research interests are metal cluster assembly in the [FeFe] hydrogenase system and [Fe–S] cluster spectroscopy and reactivity.

Joan B. Broderick was born in 1965. She received a Bachelor of Science in Chemistry from Washington State University and a Ph.D. from Northwestern University, where she was a National Science Foundation Graduate Fellow. She was an American Cancer Society postdoctoral fellow at MIT before joining the faculty at Amherst College as assistant professor in 1993. She moved to Michigan State University in 1998 and to Montana State University in 2005, where she is currently Professor of Chemistry and Biochemistry. Her research interests are in mechanistic bioinorganic chemistry, with a particular focus on enzymes utilizing iron–sulfur clusters to catalyze radical reactions.

8.18   Detection of Novel Enzyme Intermediates

Cristina M. Furdui, Wake Forest University School of Medicine, Winston-Salem, NC, USA

Karen S. Anderson, Yale University School of Medicine, New Haven, CT, USA

8.18.1   Introduction

This review focuses on the established and newly emerging strategies for identifying and characterizing enzyme intermediates. Several examples of enzymes have been chosen as they all perform unique; novel chemistries involving enzyme intermediates some of which have proven to be exciting pharmaceutical molecular targets.

Key insights into protein structure–function studies and structure-based drug design can be obtained by an understanding of how enzyme catalysis occurs at the enzyme active site. This, in turn, requires knowledge of the structure(s) of enzyme transition states along the reaction pathway in the conversion of substrate(s) into product(s). Steady-state kinetic studies are informative in providing an initial characterization of the enzyme. The application of a steady-state analysis using alternate substrates and isotope effects has also provided detailed mechanistic information. A complementary strategy that allows one to directly examine events occurring at the enzyme active site involves the use of rapid transient kinetic techniques.[1] In order to

examine these catalytic processes directly, one must look on a millisecond timescale since enzymes often turn over many molecules of substrate per second. This approach involves the rapid mixing of substrates and enzymes and monitoring the conversion of substrates into products (as well as potential enzyme intermediates) using either radiolabeled substrates or by exploiting spectroscopic properties of the substrate such as absorbance or fluorescence.

In this review, we discuss (1) the criteria for establishing enzyme intermediates, (2) methodologies and experimental design for optimal intermediate detection, (3) relevant examples demonstrating the use of this strategy to detect enzyme intermediates in several enzyme systems, and (4) new directions/approaches for detecting and characterizing enzyme intermediates. A number of excellent comprehensive reviews have been published.[1–6] The purpose of this review is not to give an exhaustive literature summary but rather to give the reader a broad overview of how one goes about detecting and characterizing enzyme intermediates. The basic concepts will be illustrated by primarily using enzyme systems that have been studied in the author's laboratory. In addition, there are a number of elegant examples included in the reference lists for applications of this strategy for the discovery of novel enzyme intermediates from a number of laboratories that the reader is encouraged to explore.

8.18.2   Criteria for Establishing Enzyme Intermediates

One of the keys to defining an enzyme catalytic reaction pathway is the identification of enzyme reaction intermediates. The criteria (Scheme 1) to establish an enzymatic reaction pathway with a postulated intermediate may be defined by addressing the following questions: (1) Can the intermediate be isolated and its structure determined directly or if it is unstable can analysis of breakdown products support the postulated structure? (2) Is the chemical rationale of the reaction intermediate based upon chemical precedent and reasonable thermodynamics? and (3) Is the intermediate 'kinetically competent', in other words, is it formed and broken down at the enzyme active site on a timescale that is consistent with the disappearance of substrate and the formation of product?

The first step in the isolation of an enzyme intermediate is to design a very sensitive experiment to detect, by an appropriate analytical method, the presence of a novel chemical species that does not correspond to either substrate(s) or product(s). In this type of approach, radiolabeled substrate is helpful to provide a sensitive means of detection. For instance, a new radiolabeled peak may be observed by HPLC/radioactivity detection in monitoring the reaction time course for the conversion of substrate into product. A covalent enzyme intermediate may be detected using protein gel analysis and by looking for a radiolabeled band comigrating at a molecular weight corresponding with the enzyme. It is often a common problem that the intermediate species is chemically labile such that the chemical quench employed to stop the enzymatic reaction can also cause breakdown of the intermediate. If the intermediate is unstable, the analysis of breakdown products can also support the postulated structure. Once evidence suggesting an enzyme intermediate is obtained, further characterization entails the isolation and structure proof by conventional spectroscopic methods such as NMR or mass spectrometry (MS). The use of isotopically labeled substrates may aid in the characterization and structural elucidation of the intermediate.

The second step in postulating that a putative intermediate may be involved in a reaction pathway is that there should be a plausible mechanism based upon a chemical rationale for the conversion of substrate into

Detection of enzyme intermediates

$$E + S \rightleftharpoons E \bullet S \rightleftharpoons E \bullet I \rightleftharpoons E \bullet P \rightleftharpoons E + P$$

- Isolation/structure proof
- Chemical precedence
- Kinetic competence

Scheme 1   Criteria for establishing enzyme intermediates.

product. This is especially important if the structure of the actual intermediate is inferred by analysis of the breakdown products. In addition, the proposed catalytic mechanism should also be reasonable in terms of thermodynamics and maintenance of microscopic reversibility.

A third criterion for establishing that a chemical species is indeed a true enzyme intermediate involved in the catalytic reaction pathway requires the species to be 'kinetically competent'. The kinetic competence is confirmed through the analysis of reaction kinetics to determine the rates of formation and decay of the putative intermediate species. Accordingly, the intermediate should be formed and broken down on a timescale that is consistent with the disappearance of substrate and formation of product. This is an important issue since there are a number of cases in which supposed enzyme intermediates have been shown to actually be enzymatic side products formed on a timescale that is much slower than catalysis. An excellent example of this concept has previously been described in the study of the enzyme 5-enolpyruvoyl shikimate-3-phosphate (EPSP) synthase in which a putative intermediate turned out to be an enzymatic side product.[7] Steady-state kinetic approaches have also been employed to assess the kinetic competence of a presumed intermediate species. This may involve preparing the chemical species through independent synthesis and then determining whether it can be converted into a product when added back to the enzyme. A potential problem with this analysis is that the reaction kinetics may be limited by the binding of the intermediate to the enzyme. One would expect that a reaction intermediate formed at the enzyme active site would bind very tightly and therefore be unlikely to dissociate and rebind. Therefore, this assessment would be based upon evaluating a step that is generally considered not to be on the reaction pathway, and as such, may not provide a definitive conclusion regarding kinetic competence.

8.18.3   Methodologies and Experimental Design for Optimal Intermediate Detection of Enzyme Intermediates

In this section we describe several methodologies that have been frequently employed to detect and characterize a variety of enzyme intermediates. In addition to the selection of the appropriate methodology, another key component for successful intermediate detection is the experimental design. The value of using single enzyme turnover experiments as a means of optimizing the likelihood of intermediate detection is discussed below.

8.18.3.1   Rapid Chemical Quench Methodology

One of the most common and useful techniques employed in a transient kinetic approach to directly examine the formation of an enzyme intermediate is the utilization of rapid chemical quench methodology.[8] Rapid chemical quench techniques involve the use of the enzyme in quantities stoichiometric with the substrate. Although a radiolabeled substrate is not essential, it is very often employed and offers a high degree of sensitivity. A chemical quenching agent is used to stop the enzyme reaction at a fixed time point allowing one to examine the extent to which chemical catalysis has occurred and isolate any stable intermediates, including any covalent enzyme intermediates that may have been formed.

Rapid mixing techniques including rapid chemical quench methods have been used to study enzymatic reactions for over 40 years. A schematic illustrating a rapid chemical quench experiment is shown in Figure 1.[1,4,9] In this type of experiment, the enzyme is loaded into one sample loop and the radiolabeled substrate is loaded into the other sample loop. A precisely controlled motor drives the buffer syringes pushing the substrate and the enzyme through the reaction loop where enzyme catalysis occurs and then the reaction is terminated by denaturing the enzyme with a chemical quenching agent such as strong base or acid, or neat organic solvent. In the selection of the most appropriate quenching agent, one must make sure that substrate, product (and putative intermediate) are stable under the quenching conditions employed and concurrently that the enzymatic reaction is very rapidly terminated. The apparatus shown below is designed to allow the use of very small sample volumes (<15 µl) and allows one to examine reaction times as short as 3 ms.

The sample is then collected and analyzed to quantify the conversion of substrate into product by an appropriate analytical method based upon the enzyme chemistry. These analytical methods may include
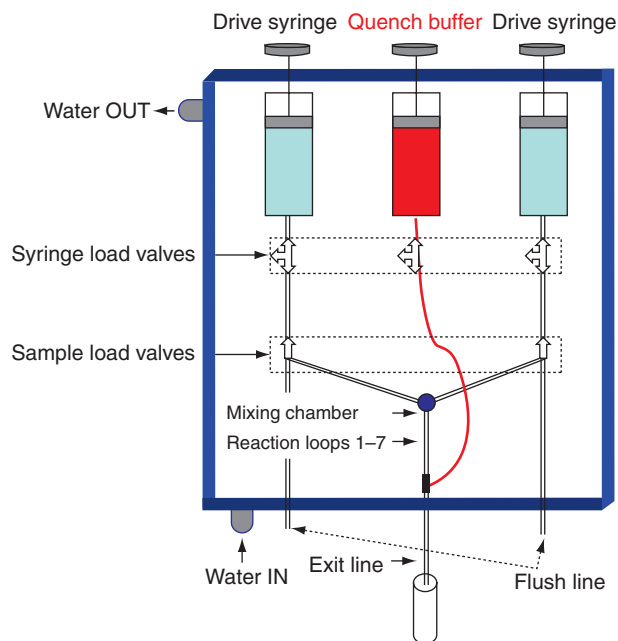
Figure 1 Schematic of a rapid chemical quench apparatus.[1,3]

HPLC, protein or DNA gel analysis in combination with radiolabeled detection, or spectroscopic methods such as NMR and MS. Upon identifying a novel species, the three criteria described above are used to provide convincing evidence that this is a true reaction intermediate.

As discussed further in the following sections, there are other variations of rapid mixing/quench methods in which the enzymatic reaction is terminated by freezing the reaction mixture with liquid isopropane.[10,11] The frozen sample is then analyzed by electron paramagnetic resonance (EPR), solid-state NMR, or other spectroscopic techniques such as resonance Raman spectroscopy that can accommodate a solid sample.[11,12] Perhaps the major limitation for implementation of this methodology is the sensitivity of the spectroscopic method and the requirement for large amounts of enzyme.[10]

Using a rapid chemical quench methodology in a transient kinetic approach, the most definitive experiment to examine the reactions at the enzyme active site is termed a single turnover experiment. This type of experiment involves the conversion of the substrate into the product in a single enzyme turnover with enzyme in excess over the limiting substrate that contains the radiolabel. These experiments permit one to directly examine the events that occur at the active site of the enzyme, and since the reaction proceeds with 100% conversion of the substrate into the product, the sensitivity for detection of any (noncovalently and covalently) enzyme-bound intermediates is optimal. In order to observe the buildup of any transient intermediates, the most important kinetic requirement, in terms of experimental design, is to use a high enough concentration of the enzyme to assure that the binding of the substrate is faster than the substrate turnover. Under conditions in which the enzyme is in excess, the rate of substrate binding is governed by the pseudo-first-order rate constant defined by the product k_{on}[enzyme]. Accordingly, proper design of the experiment depends upon an estimate of the rate of substrate binding, a knowledge of k_{cat} as well as the rate of substrate dissociation, k_{off}. The current equipment available for rapid chemical quench experiments allows one to examine the enzymatic reaction on a timescale of milliseconds that is usually required for a single enzyme turnover. A common mistake in looking for an intermediate is a failure to appreciate that there is a defined time–concentration dependence that should be considered. For instance, if the k_{cat} for an enzyme was $60\,s^{-1}$, the half-life for that reaction would be approximately 12 ms and one would therefore monitor the time course on a very short timescale. Important questions to ask are: (1) what concentrations of substrates would be required to saturate the enzyme with the intermediate? (2) what is the optimal reaction time course? and (3) according to the kinetic reaction pathway, is

the intermediate more likely to accumulate in the forward or reverse direction? These questions can be answered by a direct analysis of reaction kinetics using single turnover experiments.

The following issues may impact the experimental design of a single turnover experiment. For optimal conditions, it is preferable to use an enzyme concentration in which substrate saturation is attained. Accordingly, weakly binding substrates ($\sim$1 mmol l^{-1}) can limit the ability to achieve saturated pseudo-first-order reaction conditions. This obstacle can be overcome by conducting the experiment at very high enzyme concentrations ($>$1 mmol l^{-1}); however, potential technical problems can arise if the protein precipitates or aggregates under these conditions. The k_{cat} or enzyme turnover rate sets a limit for the slowest step in the pathway and allows an assessment of whether the reaction would be within the time domain of the rapid mixing methods currently available. With the rapid mixing apparatus described in **Figure 1**, reliable time points can be collected for as short as 3 ms. It should be noted that enzyme systems having K_d values in the millimolar range and $k_{cat} > 500 \, s^{-1}$ are less amenable to rapid chemical quench methodology. It is also important to keep in mind that the reaction kinetics dictate the amount of intermediate that may build up during a single enzyme turnover. For instance, in **Scheme 2**, the net rate of intermediate formation ($k_2 + k_{-2}$) should be greater than the net rate of intermediate conversion into product ($k_3 + k_{-3}$) to allow an appreciable amount of the intermediate to buildup and allow detection for examining the reaction in the forward direction.

The limits of sensitivity for detection are also dependent upon the nature of the radioisotope in the substrate as well as the analytical method employed to separate and detect substrate, intermediate, and products. In our experience, radiolabels are particularly effective with a detection limit of 4–5% of the total radiolabeled species comprising intermediate for less energetic isotopes such as ^{14}C or ^{3}H, and less than 1% detection limit for more energetic ^{32}P isotopes. If the reaction is reversible and depending on the kinetic pathway, there is also the option of looking for the intermediate in the reverse direction by starting with the product. Enzyme reactions that contain an irreversible step(s) are much more challenging since a fewer number of the options are available.

It is often the case that one is seeking to identify a novel chemical species for which the solution stability is uncertain. This is an important thing to keep in mind in the selection of the best quenching agent. Often a number of different conditions are tested to maintain the stability of substrates and products while anticipating the stability of putative but unknown intermediate species.

8.18.3.2 Stopped-Flow Absorbance/Fluorescence Methodologies

A large number of enzymes involved in the biosynthesis or degradation of natural compounds in all life systems rely on the use of cofactors and coenzymes to facilitate catalysis. Over the years, different rapid kinetics methods have been developed to take advantage of the spectroscopic and spectrophotometric properties of these elements. Since 1934, when the stopped flow was first introduced by Roughton, stopped-flow absorbance and fluorescence methods have been widely applied for the monitoring of rapid enzyme kinetics and detection of transient catalytic intermediates. Both absorbance and fluorescence methods take advantage of the intrinsic optical properties of substrates, intermediates, products, and enzymes which are changing along the reaction coordinate as substrates get converted into intermediates and products. Parallel changes in the absorbance or fluorescence properties of cofactors or coenzymes have often been monitored to assess the formation and decay of different kinetics species. A summary of the most relevant absorption and fluorescence parameters for some of the most common enzyme cofactors and coenzymes is presented here.

The most versatile of the coenzymes is perhaps pyridoxal phosphate (PLP). The PLP containing enzymes catalyze a wide variety of reactions such as racemization, transamination, β- and α-decarboxylation, and interconversion of side chains. The first step of all these reactions is the transition between an internal aldimine intermediate to an external aldimine intermediate, which involves the condensation of PLP with an external amino acid substrate to form a Schiff base. The internal aldimine intermediate can then either undergo α-decarboxylation to convert the amino acid substrate into amines and aldehydes, or lose the α-hydrogen

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E \cdot S \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E \cdot I \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E \cdot P \underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} E + P$$

Scheme 2 Typical reaction pathway containing a putative enzyme intermediate.

and catalyze the other reactions mentioned above. In the case of cystathionine beta synthase (CBS), a (PLP)-dependent enzyme, which catalyzes the condensation of serine and homocysteine to give cystathionine, the heme-free enzyme exhibits an absorption maximum at 412 nm corresponding to a protonated internal aldimine.[13] Addition of serine substrate results in the formation of the external aldimine which shifts the λ_{max} from 412 to 420 nm. A broad shoulder between 450 and 500 nm was observed as well which was assigned as the aminoacrylate intermediate (see Banerjee and Zou[14] for a review on CBS).

Another cofactor-dependent class of enzymes is that of cytochrome P-450 enzymes. Cytochrome P-450 enzymes catalyze a variety of oxidation and some reduction reactions, collectively involving thousands of substrates. In fact, the ability of P-450 enzymes are involved in the metabolism of more than 50% of drugs, a process dependent on P-450 enzymes to catalyze a wide range of reactions, such as carbon hydroxylation, heteroatom oxygenation, dealkylation, and epoxidation.[15–18] The reducing equivalents required for these reactions are provided by NADPH-cytochrome P-450 reductase (CPR), an FMN and FAD-dependent enzyme. The kinetics of NAD+ or NADP+-dependent enzymes is often followed either by monitoring the change in absorbance at 340 nm associated with the reduction of NAD+/NADP+ by monitoring the change in fluorescence associated with this process (NADH λ_{ex} 340 nm; λ_{em} 460 nm).[19–22] Stopped-flow absorbance kinetic studies have been applied to monitor transient flavin reduction steps in the mechanism of CPR.[23] The electrons are transferred from CPR to the heme iron in cytochrome P-450 enzymes reducing it to ferrous state, which then can bind oxygen and form the key ferrous-dioxygen intermediate.[24] Similar to cytochrome P-450, other enzymes also rely on metal cofactors to sustain catalytic activity. The most commonly encountered transition metal ions in enzymes are cobalt, iron, copper, and molybdenum. Changes in the spectrophotometric or paramagnetic properties of these elements during enzyme catalysis were crucial to the identification of transient intermediates in the mechanism of a large number of metalloenzymes.[25–29]

Even in the absence of fluorescent or chromogenic substrates, coenzymes, or cofactors, it is possible to determine kinetic parameters associated with substrate binding or conformational steps by monitoring changes in the enzyme intrinsic fluorescence properties which occur during these events. The major fluorophore in proteins is tryptophan that absorbs at 275–295 nm and emits at 330–340 nm. The amplitude of the signal will depend on the extent of conformational change induced by substrate binding or conformational changes associated with the progression of the reaction as well as the relative location and number of tryptophan residues.[30–33]

8.18.3.2.1   Future microfluidic stopped-flow absorbance/fluorescence methods

As microfluidic technologies continue to improve, we and others have started to investigate possible applications of microfluidic systems for the monitoring of rapid enzyme kinetics. More recent studies have reported on the development of a microfluidic chip equipped with fluorescence readout and capable of reaching milliseconds kinetics, while using nanoliter volumes of reagents.[34–36] The concept of microfluidic reactors was also introduced over the past several years and, in this case, the reaction progress is being monitored by spectrophotometry or, in some cases, directly by MS.[37–42] However, the current methods do not allow the independent monitoring of kinetic events at the protein active site, or the identification of transient intermediates, which could be used as templates for new inhibitors as described for KDO8P synthase system.

8.18.3.3   Spectroscopic Techniques for Detection of Enzyme Intermediates

Coupling of rapid chemical quench or freeze–quench methods with EPR, MS, or NMR has been a very useful strategy for detecting reaction intermediates that are paramagnetic such as transition metal complexes in metalloenzymes and radical species[25,43–46] or to determine the chemical structure of transient reaction intermediates or reaction products.[47–52]

Crystallography has also resulted in the identification of several unique intermediates such as the penta-covalent phosphorus intermediate of β-phosphoglucomutase reaction and the thiamin intermediates in the thiamin diphosphate and flavin-dependent enzyme, pyruvate oxidase.[53,54]

Critical intermediates in the metabolism of abused drugs were identified using stopped-flow fluorescence polarization immunoassay (SF-FPIA).[55,56] Other analytical methods such as stopped-flow FT–IR[57–61] and mass spectrometry[57–61] have also been applied to monitor rapid kinetics of protein folding and conformational change associated with catalysis.[62–66]

8.18.3.4   Mass Spectrometry as a Tool for Detection of Enzyme Intermediates

Rapid chemical quench has been enormously powerful and helped to open the field of transient enzyme catalysis, as described above; however, an alternative strategy is required for the detection and characterization of labile intermediates that avoids the need for chemical quenching. The coupling of rapid mixing techniques with online MS detection not only offers an option for the detection of labile intermediates but also has many general applications in the study of enzyme catalysis to more precisely define the key chemical events occurring at the active site.

This, in fact, was the case for detection of a hemiketal phosphate intermediate formed in the KDO8P synthase reaction pathway described in Section 8.18.5. Other potential limitations of rapid chemical quench methodology include the need for radiolabeled substrates and the need for subsequent analytical techniques, for example, NMR, to elucidate structural information on enzyme intermediates. An earlier article focusing on 'New Concepts in Bioorganic Chemistry' has highlighted the potential of using high-resolution MS coupled with electrospray ionization (ESI) to examine rapid enzyme reactions.[67]

8.18.3.4.1   Recent advances in mass spectrometry to study biomolecules

A powerful, emerging tool in the study of proteins, proteomics, and, more recently, protein complexes is MS. MS has evolved rapidly in the last 15 years and has revolutionized the detection and quantitation of biomolecules.[68–71] Recently, MS has been applied to the study of rapid transient enzyme kinetics. This technique offers advantages to the problems encountered with standard chemical quenching. First, it utilizes a novel, nonchemical enzyme quenching approach by rapidly desolvating the enzyme under vacuum to terminate enzyme catalysis. Second, the charged molecular ions generated during the soft ESI provides spectral information simultaneously from both low-molecular-weight compounds as well as covalent and noncovalent protein intermediate species, depending upon the mass range scanned. This approach thus offers the unique ability to concurrently monitor both enzyme and enzyme complexes along with small molecule substrates, intermediates, and products. The utility of this powerful analytical method for the study of labile intermediates and rapid enzyme kinetics is only now being exploited and more fully appreciated.

8.18.3.4.2   Application of mass spectrometry to monitor reaction kinetics

Previous investigations, using either continuous flow or chemical quenching followed by direct injection, have demonstrated the successful application of both ESI–MS and MALDI to monitor chemical reactions and enzymatic reactions on longer timescales (>0.1 s).[72–81] Similar strategies have been used to monitor protein folding.[82–84] Earlier reports of continuous flow methods were limited to the examination of relatively slow enzymatic reactions on the timescale of many seconds to minutes. These methods, are not suitable for the detection of short-lived chemical species with half-lives shorter than 100 ms time range since these transient species would be converted into products during the time required for slower continuous flow injection.

During the past several years, we have pioneered the feasibility of detecting transient enzyme intermediates by directly interfacing a rapid mixing apparatus with ESI–MS using a rapid mixing, pulsed-flow technique to examine enzymatic reactions on a very short millisecond timescale. This method circumvents the requirement for chemical quenching and provides the additional advantage of rapid, online detection using the high-resolution, accurate, and sensitive technique of MS. Earlier ESI–MS studies were carried out to demonstrate the detection of an enzyme intermediate in the EPSP synthase reaction.[85] We have made advances in the ESI and have improved significantly the time domain for continuous flow ESI–MS. This method is proving to be robust and reproducible and is capable of detecting both noncovalent and covalent enzyme intermediates on a timescale as short as 6–7 ms.[86] Other studies have demonstrated complementary ESI–MS approaches as a reliable tool for examining enzyme reaction kinetics, protein folding, and ligand binding studies.[87–94]

The principle working hypotheses of the rapid mixing, time-resolved ESI–MS technique are illustrated in **Figure 2**.

As the figure indicates, the concept behind these studies is simple: A computer-controlled stepping motor or syringe pump allows the precise control of the speed at which the solutions of enzyme and substrate are mixed.

Experimental design for time-resolved ESI–TOF-MS
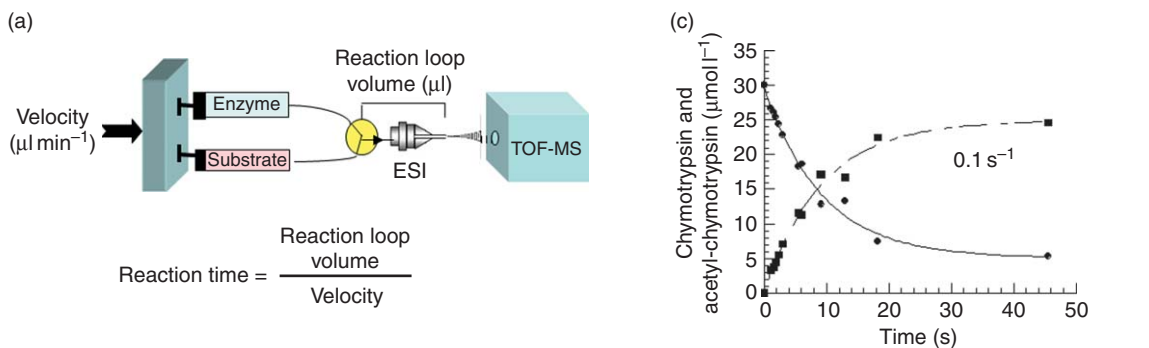
Figure 2   Time-resolved ESI–TOF-MS. (a) Diagram with the experimental design. (b) Chymotrypsin reaction with *p*-nitrophenyl acetate under single-turnover conditions monitored by time-resolved ESI–TOF-MS. The decay of chymotrypsin and formation of the acetyl-chymotrypsin intermediate is observed over a time course of 45 s. (c) Kinetic analysis of chymotrypsin decay and acetyl-chymotrypsin formation. For both traces the rate constant was 0.1 s^{-1}.

The enzyme and substrates come together in a mixing tee that has a short fused silica capillary column on the outlet. The capillary leads directly to the needle of the electrospray unit in the mass spectrometer. There is a high voltage on the needle of the electrospray unit that disperses the solution coming through the capillary into small particles for nebulization. At this point, electroconstriction and rapid desolvation occurs. Earlier studies have indicated that the enzyme reaction is terminated during the rapid desolvation.[80,85] The ions pass into the ion trap or time-of-flight (TOF) detector where they are observed in a rapid scanning mode. The extent to which the enzyme reaction proceeds is a function of both the length and diameter of the capillary column as well as the speed at which the solutions are mixed. The reaction time can be varied by driving the solutions together more slowly or more quickly through the capillary before it is terminated during the electrospray process. The speed at which the two solutions are mixed must be fast enough to avoid problems with laminar flow and maintain conditions of turbulent flow. A short time domain (6–500 ms) is attained with a short length capillary and a longer time domain (60 ms–5 s) is achieved with a longer capillary. We have demonstrated good protein quantitation results with the time-resolved ESI–TOF-MS technique by examining the classic enzymatic chymotrypsin reaction: the formation of a covalent acetyl enzyme intermediate using *p*-nitrophenyl acetate as a substrate (**Figure 2**, panels (**b**) and (**c**)). A side-by-side comparison between the ESI–MS technique and stopped-flow absorbance reveals very similar reaction kinetics. The quantitation of the relative peak intensities for the chymotrypsin decay and acetyl-chymotrypsin formation using ESI–TOF-MS afforded a rate constant of 0.1 s^{-1}. This measurement was in excellent agreement with rate constant (0.1 s^{-1}) determined, independently, by measuring the formation of the *p*-nitrophenol ($\epsilon_{400} = 5.34$ mmol^{-1} l cm^{-1}) using stopped-flow absorbance, illustrating the ability of ESI–TOF-MS to accurately determine kinetic parameters.[86,95] Other studies using ESI–MS with quadrupole detection have made similar observations.[66]

8.18.4 Selected Examples of Enzymological Studies Involving Detection and Characterization of Novel Enzyme Intermediates

In this section, we discuss the (1) application of the concepts of rapid chemical quench studies to detect enzyme intermediates, (2) biological or medical interest of studying these enzymes, and (3) the underlying kinetic concepts behind these studies. The examples chosen all perform unique, novel chemistries, and have proven to be exciting pharmaceutical targets both for antibiotics and anticancer drugs as well as herbicides.

8.18.4.1 PEP-Utilizing Enzymes that Catalyze C–O Bond Cleavage of Phosphoenolpyruvate

Phosphoenolpyruvate (PEP) is a highly functionalized, chemically versatile molecule that is used in several key biochemical steps in cellular energy metabolism and biosynthesis.[96] Although most enzymatic reactions utilizing PEP as a substrate involve cleavage of the high-energy P–O bond ($\Delta G^{\circ\prime} = -14.8\ k_{cal}\,mol^{-1}$), two types of reactions have been shown to involve the unusual cleavage of the C–O bond of PEP:[97] (1) formation of an enol ether linkage through transfer of the enolpyruvoyl moiety in PEP to a cosubstrate alcohol (**Scheme 3**, pathway 1) and (2) formation of the net aldol condensation product through coupling of C-3 of PEP with a cosubstrate aldehyde (**Scheme 3**, pathway 2).

It is interesting to note that, currently, only four enzymes found in nature catalyze the unusual C–O bond cleavage of PEP. Two of the enzymes fall into the type-1 category, while the other two enzymes fall into the type-2 category. The type-1 reaction pathway in **Scheme 3** is represented by EPSP synthase, an enzyme in the shikimate pathway involved in aromatic acid biosynthesis[98] and UDP-GlcNAc enolpyruvoyl transferase (*Mur Z*; more recently termed *Mur A*), an enzyme involved in peptidoglycan biosynthesis.[99] These enzymes catalyze the transfer of enolpyruvoyl moiety from PEP to their respective cosubstrate alcohols and are the targets of commercially important inhibitors; Mur A is targeted by the antibiotic fosfomicin[100] whereas EPSP synthase is the site of action for the herbicide glyphosate.[101] In this section, we briefly describe how a rapid transient kinetic approach was used to provide detailed insight into the nature of enzyme reaction mechanisms including the detection and characterization of enzyme intermediates for both EPSP synthase and UDP-GlcNAc enolpyruvoyl transferase.

There are at least two known enzymes that can carry out the type-2 reaction illustrated in **Scheme 3**. The first enzyme, 3-deoxy-D-*manno*-2-octulosonate-8-phosphate (KDO8P) synthase, catalyzes the formation of KDO8P, an 8-carbon sugar, from the 5-carbon sugar, D-arabinose 5-phosphate (A5P), and PEP. This is an important enzymatic reaction that controls the carbon flow in the biosynthetic formation of another 8-carbon sugar 3-deoxy-D-*manno*-2-octulosonate (KDO), an important constituent of the lipopolysaccharide of most Gram-negative bacteria, and therefore this enzyme is a potential molecular target for new antibiotics.[102] The second enzyme, 3-deoxy-D-arabino-2-heptulosonate acid-7-phosphate (DAHP) synthase, such as EPSP synthase, is found in the shikimate pathway for the biosynthesis of essential aromatic amino acids[97] may also represent a molecular target for the design of novel antibacterial therapeutics. Recently, a third enzyme, *N*-acetyl-neuraminic acid (NeuAc) synthase has been found in the polysaccharide cell wall biosynthetic pathway of *Escherichia coli* that may also fall into this category. NeuAc synthase also catalyzes the net aldol condensation of a 6-carbon sugar, *N*-acetyl-mannosamine with PEP, to form the 9-carbon sugar, *N*-acetyl-neuraminic acid with the release of inorganic phosphate.[103] For a number of years
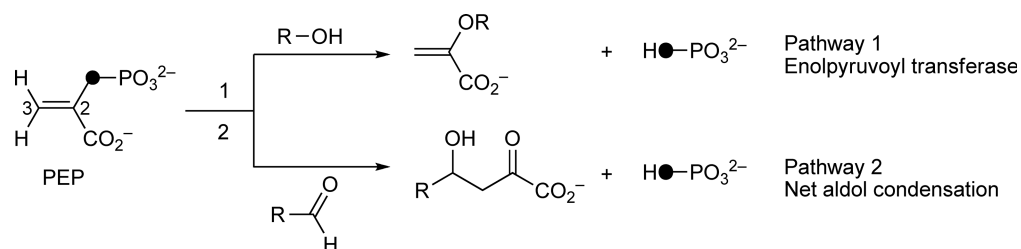
**Scheme 3** Two types of enzymes catalyzing C–O bond cleavage of PEP.

there has been an interest in studying these enzymes for their unique mechanistic features as well as their potential importance as molecular targets for therapeutic intervention. A detailed discussion of the mechanism and characterization of novel enzyme intermediates for each of these unique PEP-utilizing enzymes has recently been reviewed;[104] therefore, they will only be briefly mentioned here.

8.18.4.1.1   EPSP synthase: A tetrahedral ketal phosphate enzyme intermediate

Much earlier studies on EPSP synthase with radiolabeled substrates suggested the existence of an enzyme intermediate.[99,105–108] There were two plausible mechanisms as illustrated in **Scheme 4**: (1) a tetrahedral intermediate formed by protonation of the C-3 carbon of PEP, and subsequent attack of the 5′-OH of shikimate-3-phosphate (S3P) to form an intermediate that was not covalently bound to the enzyme as illustrated in the scheme (pathway a), and a covalent intermediate involving a protonated form of PEP that becomes covalently attached to the enzyme through an active site nucleophile during catalysis (pathway b).

A series of rapid chemical quench experiments under single enzyme turnover conditions using radiolabeled S3P or PEP revealed that the tetrahedral ketal phosphate enzyme intermediate was formed as a new peak upon HPLC analysis with detection of the radiolabel.[109] The time course revealed that the formation of the tetrahedral intermediate species paralleled the disappearance of PEP substrate and formation of the EPSP product thus establishing that it was a kinetically competent species.[7,110] Isolation of the tetrahedral ketal phosphate intermediate using ^{13}C-2 PEP and S3P as substrates coupled with rapid chemical quench was carried out in conjunction with ^{1}H-, ^{13}C-, and ^{31}P- NMR to provide a definitive structure proof.[111] Thus with these studies we have satisfied the criteria for a true reaction intermediate in terms of a chemically plausible mechanism, structure proof, and kinetic competence. Additional studies support the mechanism for EPSP synthase described (**Scheme 4**, pathway a) including observation of the intermediate bound to the enzyme at internal equilibrium using solution NMR and ^{13}C-2 PEP[112] as well as using rapid freeze–quench/solid-state NMR studies.[12]

Concurrent with these studies, it was found that the EPSP ketal species, shown below, could be isolated after long-term incubations with the enzyme and the substrates.[113] When this compound was originally isolated, it was proposed to be a possible enzyme intermediate. However, a detailed kinetic analysis revealed that the rate of formation of the EPSP ketal was a million-fold slower than catalysis and therefore did not meet the criteria of a kinetically competent species.
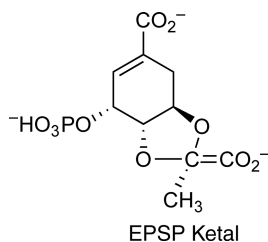
EPSP Ketal

Furthermore, it is difficult to incorporate the EPSP ketal as an intermediate in a chemically plausible reaction mechanism for converting S3P and PEP into EPSP and P_i. In the experiments described above using solution NMR and ^{13}C-2 PEP to observe the tetrahedral intermediate on the enzyme at internal equilibrium with substrates and products, the EPSP ketal was formed over longer times as a dead-end breakdown product of tetrahedral ketal phosphate intermediate.[7,112] The formation of EPSP ketal most likely occurs through trapping of a protonated enol form of PEP using the 4-OH group of S3P as a nucleophile. It is believed that this side product is also observed in a solid-state NMR experiment.[114]

In summary, through the use of rapid chemical quench techniques, multiple studies demonstrated the formation of a single tetrahedral intermediate in the reaction pathway of EPSP synthase (**Scheme 4**, pathway a) which is formed by an attack of the 5-OH group of shikimate-3-phosphate on C-2 of PEP.[31,109,111] A complete kinetic and thermodynamic description of this enzyme reaction pathway could be demonstrated, including the isolation and structural elucidation of a tetrahedral enzyme intermediate as originally proposed by Sprinson.[106] This work established the catalytic mechanism and definitively showed that no covalent enzyme–PEP adduct is formed on the reaction pathway. Subsequent work using rapid mixing pulsed-flow ESI–MS studies[85,110] and solution phase NMR[115] provides additional support for the catalytic pathway in **Scheme 4**, pathway a.
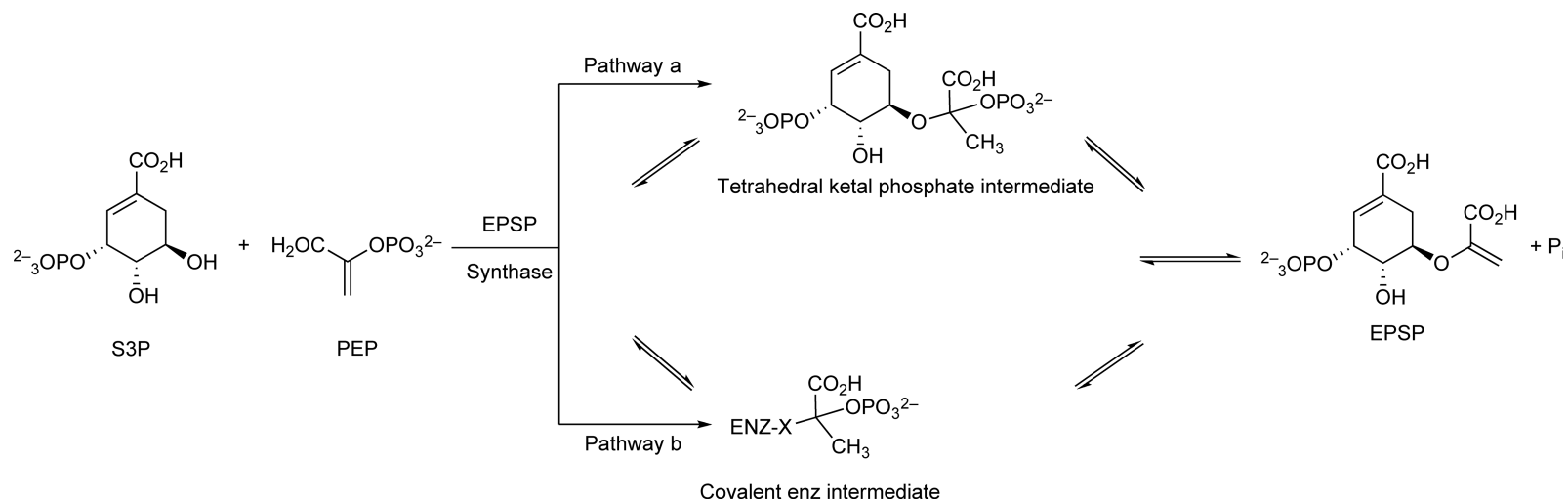
Scheme 4 Two mechanistic pathways proposed for EPSP synthase.

8.18.4.1.2   *Mur Z (Mur A): Tetrahedral and covalent enzyme intermediates*

The overall catalytic reaction for UDP-GlcNAc enolpyruvoyl transferase (Mur Z; also called Mur A) is very similar to EPSP synthase in that the enolpyruvoyl group of PEP is transferred to the 3′-OH group of the amino sugar substrate, UDP-GlcNAc. We have also used a rapid transient kinetic approach to examine the mechanism of UDP-GlcNAc enolpyruvoyl transferase. Early studies with Mur Z suggested an enolpyruvoyl[116] or phospholactoyl[105] covalently bound enzyme intermediate as assessed by tight binding of radiolabeled PEP to the enzyme. Later studies with the *E. coli* Mur Z as well as work with *Enterobacter cloacae* Mur Z both suggested a covalent enzyme intermediate.[100] A series of experiments using ^{32}P-labeled PEP to examine the reaction kinetics under single turnover conditions revealed that there were actually two kinetic approaches as illustrated in **Figure 3**.

The time course indicated two intermediates I_1 (the covalent phospholactoyl) and I_2 (corresponding to a ketal phosphate intermediate) as illustrated in **Scheme 5**. The formation and decay of both I_1 and I_2 paralleled the disappearance of the substrate PEP and the appearance of product P_i establishing that each intermediate represents a kinetically competent species in the reaction. The identity of the soluble intermediate was established by isolation of I_2 prepared using [^{13}C]-2-PEP as a substrate and characterizing the structure using ^{1}H-, ^{13}C-, and ^{31}P-NMR.[117] This compound was found to be a phospholactoyl-UDPG-GlcNAc species analogous to the tetrahedral ketal phosphate intermediate observed in the EPSP synthase. An examination of time course for I_1 and I_2 indicated that each is formed very rapidly and they decay in parallel suggesting an apparent rapid equilibrium between the soluble and enzyme-bound intermediates.

Substrate trapping experiments confirmed that the covalent phospholactoyl-Mur Z intermediate (I_1) was capable of giving rise to the phospholactoyl-UDPG-GlcNAc (I_2) and ultimately the enolpyruvoyl-UDPG-GlcNAc product. Although there is no direct evidence that one intermediate might precede the other intermediate; however, studies conducted by Kim *et al.*[118] suggest that the pathway may indeed be a branched pathway as illustrated in **Scheme 5** rather than one intermediate being converted into the second intermediate.

In summary, these studies have led to the isolation and characterization of *two* kinetically competent intermediates: a covalent phospholactoyl–enzyme adduct and a phospholactoyl-UDP-GlcNAc tetrahedral intermediate.[117,119] Further work is required to definitively establish whether the two intermediates are formed along a sequential or branched pathway as shown in **Scheme 5**, pathways a and b, respectively.

The formation of a 'covalent' phospholactoyl–enzyme adduct is a major distinction between the mechanisms of catalysis by UDP-GlcNAc enolpyruvoyl transferase (MurZ) and EPSP synthase. Nonetheless, the structural and functional homologies suggest at least some common mechanistic features, as indicated by the isolation of similar tetrahedral intermediates in both enzyme reactions.
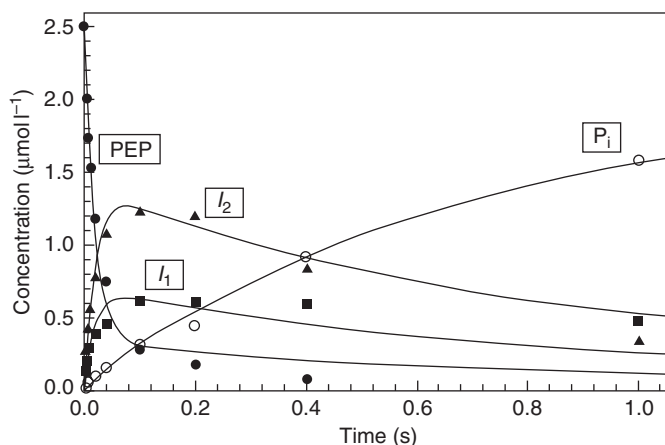
Figure 3   Single turnover experiment for Mur Z-competent intermediates.
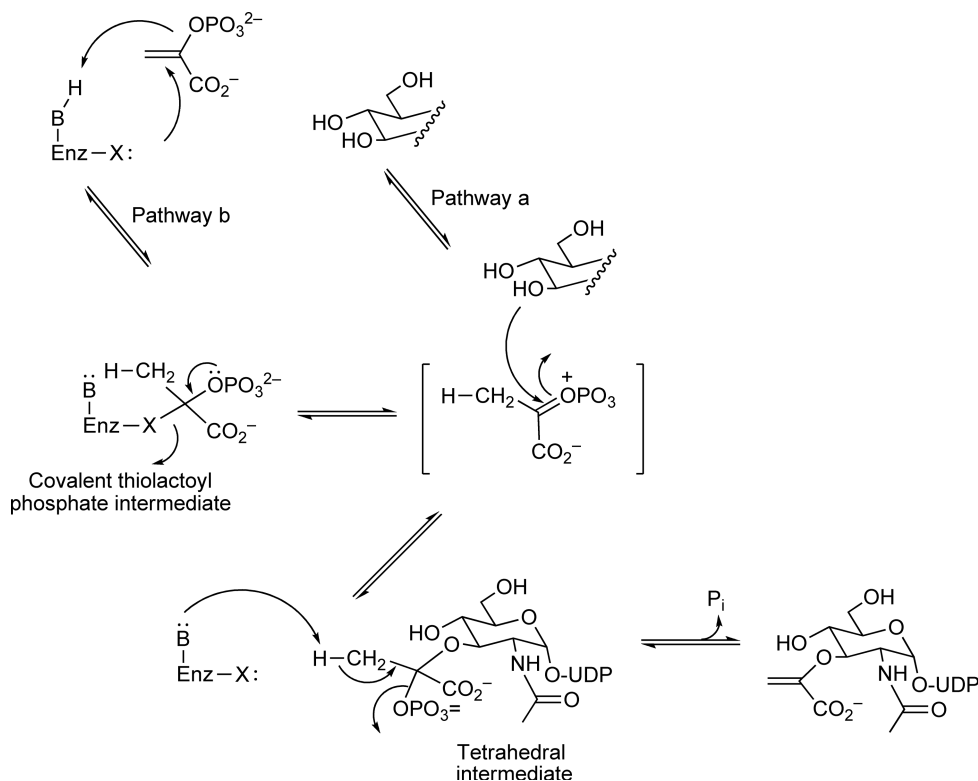
Scheme 5 Mur Z reaction pathway.

8.18.4.1.3 KDO8P synthase: A cyclic ketal phosphate or acyclic hemiketal phosphate enzyme intermediate

While the mechanisms of the C–O bond cleavage for the two 'type-1' enolpyruvoyl transferase enzymes (**Scheme 3**, pathway 1) have been unambiguously characterized, the mechanisms of KDO8P synthase, DAHP synthase, and potentially NeuAc synthase that represent the second distinct class (**Scheme 3**, pathway 2) of enzymatic reactions involving C–O bond cleavage of PEP and the net the aldol condensation continue to be uncertain. Although earlier[120] and more recent[121,122] studies have established exclusively that the DAHP synthase and KDO8P synthase reactions proceed with C–O bond cleavage of PEP, the elementary steps of this transformation until recently were unclear.

The KDO8P synthase enzymatic reaction has received increasing attention in the recent years and there is now a substantial amount of structural information available.[123–125] Unlike the two enolpyruvoyl transferase enzymes, earlier studies of KDO8P synthase suggested that the reaction is largely irreversible based upon the failure of [^{32}P]-phosphate to exchange into PEP or of ^{18}O to scramble from the bridge to the nonbridge positions of [^{18}O]PEP in the presence of A5P.[120,122] Furthermore, the anomeric oxygen of the KDO8P product has been shown to originate from the bulk solvent as evidenced by the incorporation of [^{18}O] at the C-2 position of KDO8P when the reaction is carried out in the presence of [H$_2$ ^{18}O]. Additional studies have indicated that the enzyme acts upon the acyclic form of A5P, and have demonstrated an ordered sequence of substrate binding (PEP followed by A5P) and product release (P$_i$ prior to KDO8P).[126] By using stereospecifically labeled 3-deuterio and 3-fluoro analogs of PEP as alternate substrates of KDO8P synthase, it has been shown that the condensation step is stereospecific, involving the attachment of the *si* face of PEP to the *re* face of the carbonyl of A5P.[121,127] Based on these mechanistic data, in combination with the results accumulated through the synthesis and examination of various analogs of A5P,[126,128] analogs of PEP,[121,129,130] and of the product KDO8P,[126,129,131] as mechanistic probes, two distinct, chemically feasible reaction pathways have been proposed for the reaction of KDO8P synthase. These pathways, as illustrated in
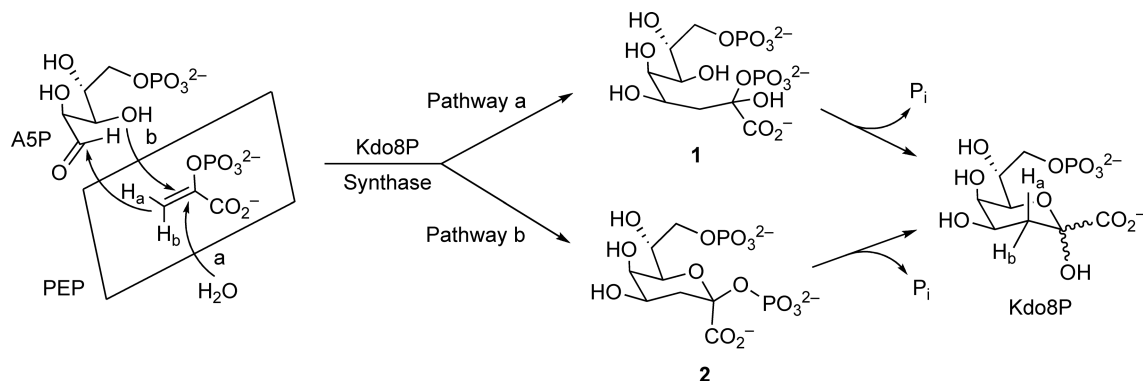
Scheme 6   Potential mechanisms for KDO8P synthase.

Scheme 6, involve the formation of either the acyclic hemiketal phosphate intermediate **1** (pathway a),[120] or the cyclic ketal phosphate intermediate **2** (pathway b).[122,132] An alternate mechanism would involve covalent enzyme catalysis in which an enzyme nucleophile participates in the condensation between PEP and A5P. Similar types of mechanisms can be written for both DAHP synthase and NeuAc synthase.

The proposed cyclic ketal phosphate intermediate **2** in **Scheme 6** has very similar functionality to the tetrahedral enzyme intermediates isolated and characterized in the EPSP synthase and UDP-GlcNAc enol-pyruvoyl transferase reactions. Although the tetrahedral ketal phosphate intermediate for EPSP synthase is quite labile at neutral or acidic pH, it is surprisingly stable at basic pH>12 ($t_{1/2}$ =48 h).[7] Based upon these results, we might predict that intermediate **2** should be detectable by rapid chemical quench techniques if isolated under basic conditions even if it was formed only transiently at the enzyme active site. Although the mechanistic data described above suggest the catalytic pathway outlined in **Scheme 6**, there was no direct information in support of either intermediate.

A number of single turnover experiments, using ^{14}C- and ^{32}P-labeled substrates and products to examine both the forward and reverse reactions, were conducted to identify enzyme intermediates[48] such as the cyclic ketal phosphate, **2**, or covalent enzyme adducts. In addition, the cyclic ketal phosphate intermediate was synthesized and evaluated as both a substrate and an inhibitor of KDO8P synthase.[48] We found no evidence for the cyclic ketal phosphate as an intermediate or substrate for the enzyme. The cyclic ketal phosphate, **2**, did, however, serve as a competitive inhibitor of the enzyme with respect to PEP with a K_i of 35 μmol l^{-1}. Similarly, analysis of single turnover experiments, using SDS/PAGE/RAD analysis provided no evidence of radiolabel associated with the band of enzyme (~30 kDa) indicating that no covalently bound enzyme intermediates are formed during catalysis. Thus, taken together, these experiments provide evidence that rule out both a covalent enzyme intermediate and a cyclic ketal phosphate intermediate and suggest instead the acyclic hemiketal phosphate, **1**, as a possible intermediate for the KDO8P synthase reaction.[48] While there is very little information in the literature on the stability of a hemiketal phosphate, one might expect that a compound containing this functionality would be highly chemical labile upon quenching with either acid or base thus precluding detection using rapid chemical quench methodology. Thus, a major limitation of this approach is that intermediates that are very labile chemically would decompose upon quenching the enzymatic reaction. This problem encountered with trying to detect a labile hemiketal phosphate intermediate in the KDO8P synthase reaction prompted us to develop a new approach for observing intermediates that are chemically labile. This strategy is discussed in Section 8.18.5.

8.18.4.2   A Covalent Phosphoryl Cysteine Enzyme Intermediate in a Tyrosine Phosphatase

Protein tyrosine phosphatases (PTPases) are an important class of enzymes for the regulation of signal transduction pathways that control a number of cellular processes including cell growth and differentiation.[133] The phosphorylation of tyrosyl protein residues has been shown to be one of the key events involved in intracellular signaling for selective gene activation. The PTPases have been shown to work in concert with

their counterparts, the protein tyrosine kinases, in switching signaling pathways off and on by controlling the lifetime of phosphorylated tyrosyl groups in a dynamic and reversible manner.[134] PTPases contain signature motifs including an absolutely required cysteine residue that has been suggested to play an essential role in catalysis. Based on a series of site-directed mutagenesis studies on the conserved cysteine, it has been suggested that the active-site cysteine serves as a catalytic nucleophile by formation of an unusual covalent S-phosphocysteinyl enzyme intermediate.[135,136] In the section, we describe studies using rapid chemical quench methodology and the criteria for defining intermediates to establish that the covalent thiophosphoryl enzyme intermediate as a *true* intermediate along the catalytic reaction pathway of leukocyte antigen-related (LAR) PTPase, a representative tyrosine phosphatase, important in insulin signaling and a potential target for new therapeutics.[133]

The strategy was to first define the reaction kinetics for the formation and decay of a covalent enzyme intermediate. This information would establish the kinetic competence of the intermediate as well as enable the design of optimal conditions for the isolation and structure proof of the intermediate species. As we indicated earlier, it is important to define the time–concentration dependence for the formation of an intermediate in order to make the most successful attempt at isolation and characterization. We conducted a single turnover experiment with a purified, highly active catalytic fragment of human LAR PTPase (LAR-D1), using as a substrate, a ^{32}P-labeled monophosphotyrosyl peptide with high affinity for the enzyme. The monophosphotyrosyl dodecapeptide, TRDIpYETDFFRK, is an analog of residues 1142–1153 of the autophosphorylation site of the insulin receptor in which tyrosyl groups at positions 9 and 10 are replaced with phenylalanines to reduce regiospecificity issues in the pY peptide synthesis and enzymatic dephosphorylation. A single turnover time course with the enzyme in excess over the substrate, demonstrated the formation and decay of a covalent phosphoryl intermediate as illustrated in **Figure 4(a)**. The intermediate formed in 3 ms and decayed to form products (dephosphorylated tyrosine substrate and inorganic phosphate, P$_i$) over the next 50 ms as assessed by SDS/PAGE analysis. This analysis allows the phosphorylated intermediate (MW 40 000) to be readily distinguished from phosphorylated tyrosyl peptide substrate (MW 1400) and inorganic phosphate (MW 164). The rates of formation $(1200 \, \mathrm{s}^{-1})$ and decay $(80 \, \mathrm{s}^{-1})$ for the intermediate paralleled the disappearance of the substrate and the formation of the product and were found to be substantially faster than the steady-state turnover rate $(24 \, \mathrm{s}^{-1})$ indicating that product release is most likely the rate-limiting step. Moreover, an analysis of the reaction kinetics established that the intermediate is kinetically competent and a true intermediate along the catalytic reaction pathway.

The next step was to identify the active-site residue, Enz-X-PO$_3$$^{2-}$, which becomes phosphorylated during catalysis. This was accomplished by preparing quantities of the LAR-D1 phosphoryl intermediate sufficient for
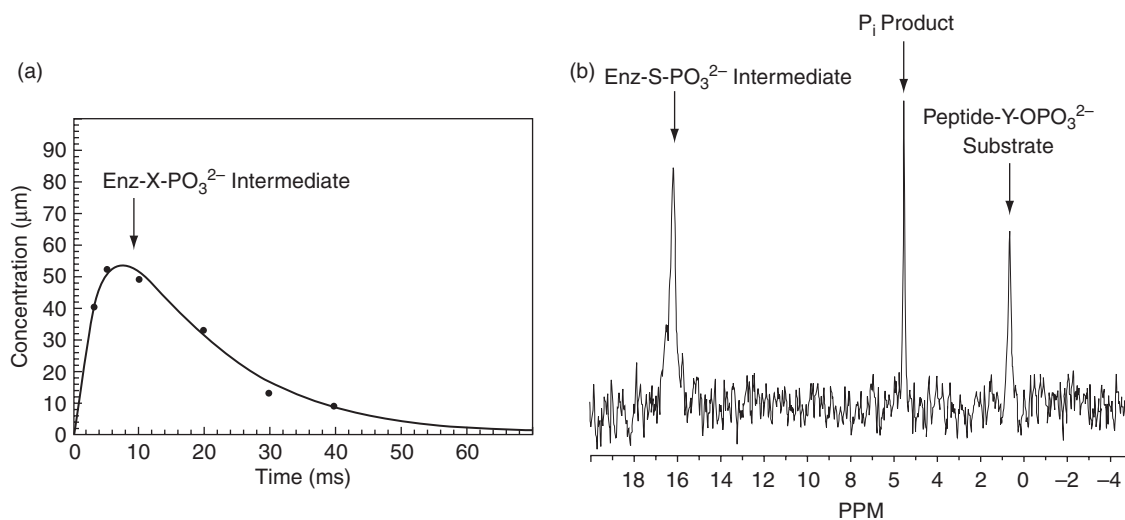
Figure 4   Observation of covalent Enz int in LAR tyrosine phosphatase by (a) PAGE analysis and (b) NMR.

^{31}P-NMR analysis to establish the identity of the active-site residue. This was accomplished by reacting the TRDIpYETDFFRK peptide (0.6 mmol l^{-1}) with LAR catalytic fragment D-1 (0.66 mmol l^{-1}) for 10 ms followed by quenching with 0.2 M NaOH in D$_2$O (final concentrations). A small amount of radiolabeled peptide substrate was included to verify the reaction by SDS/PAGE analysis.

The ^{31}P-NMR spectrum obtained after quenching is shown in **Figure 4(b)**. Three major phosphate signals were observed. The two upfield signals correspond to the substrate (0.6 ppm) and the product, inorganic phosphate, P$_i$ (5.5 ppm). A third signal was observed downfield at 16.1 ppm, suggesting a phosphoryl cysteine intermediate. This chemical shift was consistent with previously reported phosphoryl cysteine intermediate species.[137,138] An authentic phosphocysteine peptide, HpCSAGVGRTG, corresponding to residues 1521–1530 of LAR and containing the active site Cys1522, was prepared and served as a ^{31}P-NMR standard, yielding a ^{31}P resonance at 16.1 ppm and validating the Enz-X-PO$_3{}^{2-}$ as an Enz-S-PO$_3{}^{2-}$. The spectral characterization coupled with our rapid quench kinetics provided definitive identification of the covalent phosphorylcysteine intermediate in the LAR PTPase reaction pathway.

8.18.4.3   Tryptophan Synthase: Probing Substrate Channeling in a Bifunctional Enzyme and Searching for Indole as an Intermediate

In the enzymes described above, important clues to the mechanistic pathway were provided by the isolation and the characterization of both low-molecular-weight and covalently bound enzyme intermediates. Our next example, represents a unique case in which, it was actually the *lack* of observation of an intermediate that provided key features of the reaction kinetics that would enable us to understand how two active sites of a bifunctional protein communicate in a concerted manner to regulate the formation and utilization of an enzyme intermediate.
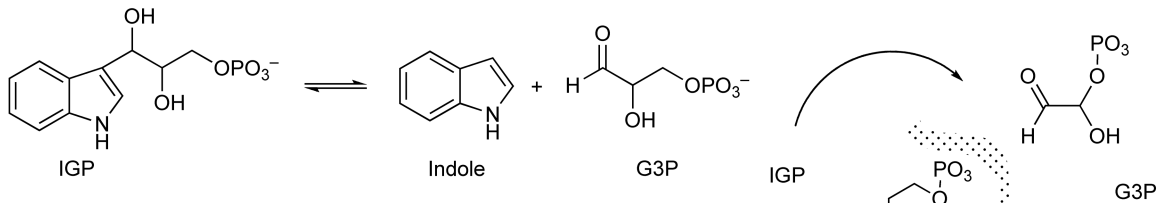
Tryptophan synthase, an $\alpha_2\beta_2$ tetrameric enzyme complex, catalyzes the final two steps in the biosynthesis of tryptophan (for reviews see Anderson,[6] and Miles,[139–141]). The α subunit catalyzes the cleavage of indole 3-glycerol phosphate (IGP) to indole and glyceraldehyde 3-phosphate (G3P) (α reaction), while the β subunit catalyzes the condensation of indole with serine in a reaction mediated by PLP (β reaction). The physiologically important reaction is termed the $\alpha\beta$ reaction and involves the conversion of IGP and serine to tryptophan and water (**Figure 5**).[142,143] The solution of the three-dimensional crystal structure of the enzyme from *Salmonella typhimurium* provides physical evidence for a hydrophobic tunnel 25 Å in length that connects the active sites of the α and β subunits.[144,145] It is suggested that indole is a metabolic intermediate that is transferred from the active site of the α subunit to the active site of the β subunit through the connecting tunnel as illustrated in **Figure 5**.

This direct transfer of the indole intermediate is suggested to occur by a process known as substrate channeling in which the product of the α reaction, indole is directly passed over a distance of 25 Å to the β active site, in a process that does not allow indole to come into contact with the bulk solvent. Indole has not been observed as an intermediate in the $\alpha\beta$ physiological reaction, but it is chemically reasonable to suppose that it is involved in this conversion.
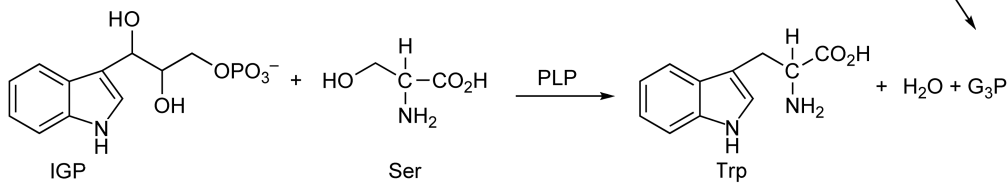
8.18.4.3.1   *Transient kinetic studies to understand substrate channeling*

A series of single turnover experiments examining the α and β reactions as well as the $\alpha\beta$ reaction, allowed us to define the reaction kinetics when each site was working independently as well as in the physiological reaction ($\alpha\beta$) where the putative indole formed at the α active site was required to travel to the β site where it would react with a PLP as the activated form of serine to produce tryptophan.[30,146,147] A reaction time course examining a single turnover experiment for $\alpha\beta$ reaction is shown in **Figure 6**(**a**). As noted in this figure, we see the disappearance of radiolabeled IGP and the formation of tryptophan but *no* indole could be observed within the detection limits of radiolabeled substrate/products. This experiment allowed us to set a lower limit on the required rate of reaction of indole to form tryptophan (>1000 s^{-1}) either channeled from the α site to the β site, or from the solution, to account for the fact observation that no indole could be detected in the $\alpha\beta$ reaction. In separate experiments, the rate of reaction of radiolabeled IGP to form indole (α reaction), in the absence of serine at the β site, was only (0.16 s^{-1}), and the rate of reaction of radiolabeled indole to form tryptophan (β reaction),

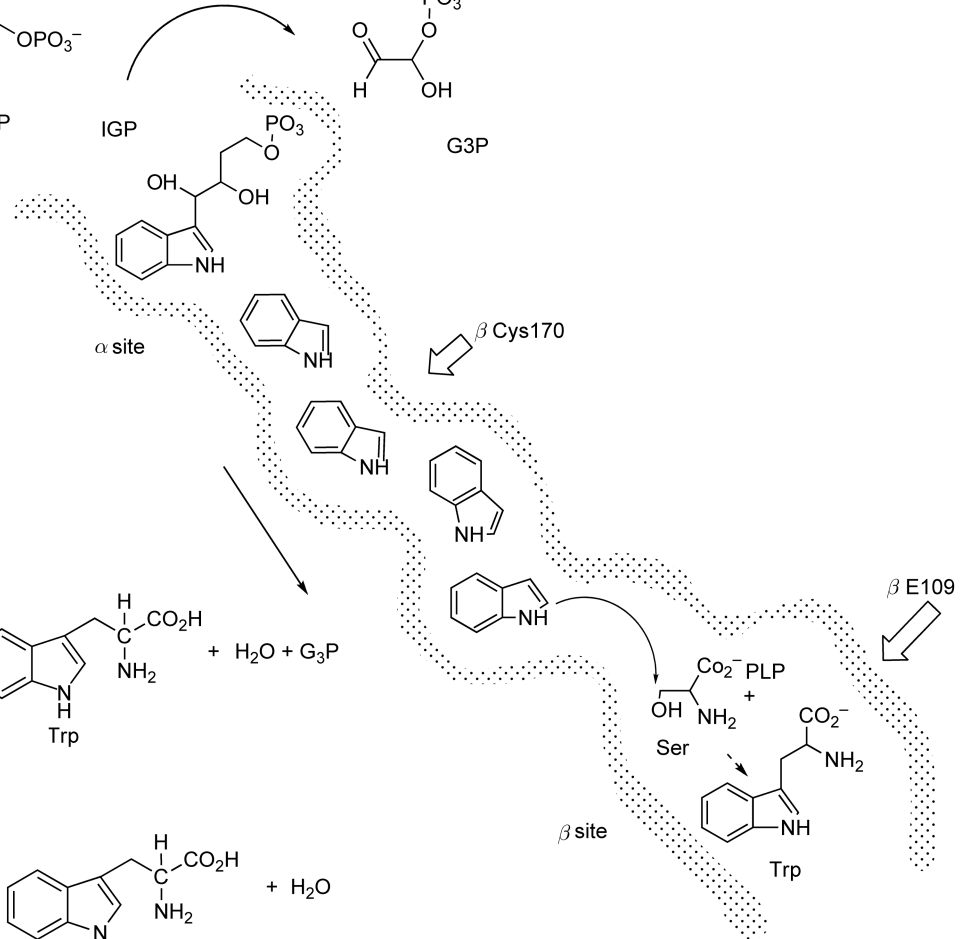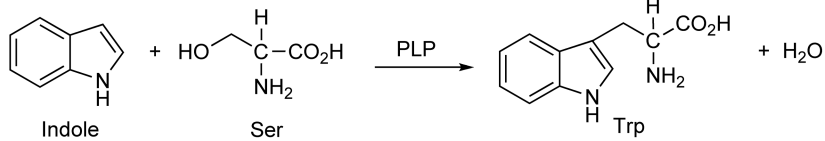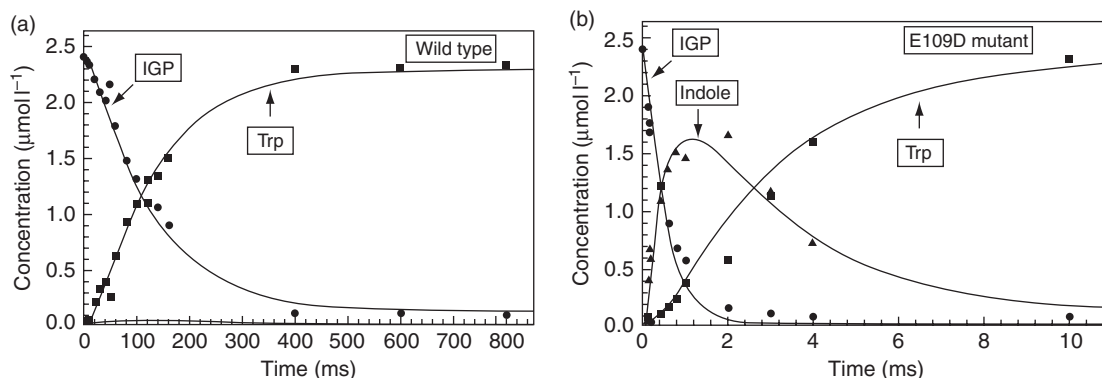**Figure 5** Diagram illustrating the hydrophobic tunnel and catalysis at the α and β active sites.

Figure 6 Single turnover for wild type and channel-impaired E109D mutant Trp synthase.

from solution was determined to be much slower ($\sim$40 s^{-1}). A combination of these three single turnover experiments provides definitive evidence that indole must be channeled in the physiological $\alpha\beta$ reaction.

Efficient channeling depends upon effective communication between the α and β sites to attenuate the formation of indole in the presence of serine.[30,148] An in-depth transient kinetic analysis of wild-type tryptophan synthase using both rapid chemical quench and stopped-flow methods, as described above, lead to the kinetic model to understand the intersubunit communication between the α and β subunits.[30] The catalytic reaction occurring at the α site in which the substrate, IGP, is converted into indole, in the absence of serine, involves binding of IGP to enzyme followed by a rate-limiting conformational change to IGP-E*. If serine is not present at the β site, the rate of this conformational change is relatively slow at 0.16 s^{-1}. The catalytic reaction that occurs at the β site involves the activation of serine. Serine binds to the enzyme and reacts with PLP to form a Schiff base external aldimine species (E-Ser or Ex-Ald).[149,150] This species is then deprotonated at the α-carbon of serine to form a quininoid species that is subsequently dehydrated to form an activated aminoacrylate derivative (E$\sim$AA). If indole is added (as in the β reaction), it rapidly condenses with the aminoacrylate to form tryptophan (Trp). In the physiological $\alpha\beta$ reaction, the presence of serine at the β site activates the α site to cleave IGP to indole. The formation of the E$\sim$AA species induces a protein conformational change to E*$\sim$AA. This activation enhances the rate of E-IGP to E*-IGP such that it is no longer rate limiting (24 s^{-1}). Once an indole is formed it is rapidly channeled to the β site where it reacts with the E*$\sim$AA to form Trp.

The salient features of this model involve the binding of the β subunit ligand serine and its activation through a PLP-dependent reaction to form a reactive enzyme-bound aminoacrylate (PLP$\sim$AA) species that in turn triggers a conformational change that promotes the cleavage of IGP to indole at the α subunit.[30] When a molecule of indole is formed it diffuses rapidly through the hydrophobic tunnel to the β subunit and reacts with the PLP$\sim$AA, also very rapidly, to form the product tryptophan. This intersubunit communication keeps the α and β reactions in phase such that the intermediate indole does not accumulate.

Three features of the reaction kinetics are essential to ensure that indole is channeled efficiently: (1) the reaction of serine at the β site modulates the formation of indole at the α site such that indole is not produced until serine has reacted to form E*$\sim$AA; (2) the rate of reaction of indole and E*$\sim$AA is fast and largely irreversible; and (3) the rate of indole diffusion from the α site to the β site is very fast (>1000 s^{-1}).[30] This mechanism accounts for the fact that indole does not accumulate during a single turnover of conversion of IGP into tryptophan (the $\alpha\beta$ reaction). This model makes several predictions, which have been tested by kinetic and structural analysis of mutants and alternate substrates[30,151] using single enzyme turnover experiments.

The model for α–β intersubunit communication indicates that it is the formation of the aminoacrylate species that leads to activation of the α reaction. When both serine and IGP are added simultaneously to the enzyme in a single enzyme turnover experiment, there is a lag in the cleavage of IGP that is a function of the reaction of serine to form the aminoacrylate species. Accordingly, amino acids other than serine that can undergo dehydration to form the aminoacrylate such as cysteine should serve as alternate substrates but should lead to a longer lag for the α subunit activation as determined by transient kinetic analysis. Cysteine does

indeed serve as an alternate substrate but a substantially slower activation process was observed.[30,152] Nonetheless, the α and β reactions remain coupled, with no observed accumulation of indole, indicating that efficient channeling an still occurs. Amino acids such as glycine and homoserine that cannot undergo dehydration to form an aminoacrylate do not promote activation.[30] Additional evidence for the importance of E*~AA is provided by the preparation and analysis of a mutant form of tryptophan synthase (β Lys87 Thr) in which the lysine involved in the covalent Schiff base formation with PLP has been replaced by a threonine residue. The β Lys87 Thr mutant $\alpha_2\beta_2$ complex contains active α subunits and β subunits that retain the ability to bind PLP but are completely inactive enzymatically. This mutant can form stable Schiff base intermediates with L-serine and other amino acids that can be converted into the corresponding aminoacrylate derivative by addition of ammonia, which partially replaces the deleted ϵ-amino group of β Lys87. Rapid kinetic analysis shows that the 'chemically rescued' mutant exhibits a sixfold increase in the rate of cleavage of IGP to indole relative to the L-serine external aldimine derivative of the mutant thus establishing the importance of the aminoacrylate in activation of the α reaction.[153]

Based on our working model and transient kinetic analysis, we might predict that if we slow the rate of reaction of indole with the E*~AA species in the β reaction, the efficiency will be lost and indole may build up in a single turnover of the $\alpha\beta$ reaction. A residue in the β subunit, Glu109, has been suggested to play a role in activating indole toward nucleophilic attack on the aminoacrylate intermediate and is supported by X-ray crystallographic studies.[30,154] By mutating β Glu109 residue to Asp, the rate of the β reaction is decreased by a factor of 300.[30] The rapid chemical quench analysis of the βE109D mutant confirmed that the buildup of indole is observable if the rate of catalysis at the β site is slowed. This observation is also supported by rapid-scanning stopped-flow spectroscopic analysis of this mutant.[155] The rate of catalysis at the βE109D subunit is $3\,s^{-1}$ compared with $>1000\,s^{-1}$ in wild-type tryptophan synthase. The slower reaction at the β site allows a substantial amount of indole to be detected in a single turnover reaction as illustrated in **Figure 6**(**b**). This is the first time indole has been observed as an actual reaction intermediate in the physiological $\alpha\beta$ reaction of tryptophan synthase.[30]

The importance of rapid diffusion of indole from the α active site through the channel to the β active site was tested by creating a mutation that blocks the tunnel. According to this model, if the hydrophobic tunnel is blocked so as to interfere with the passage of the indole from the α site to the β, indole might be observed in a single enzyme turnover experiment. This hypothesis was tested by mutating one of the hydrophobic residues, β Cys170 (see **Figure 5**), that define the tunnel to a bulkier aromatic residue (Phe or Trp). The kinetic analysis reveals that with both the βC170F and βC170W mutant enzymes, we are able to observe indole in a single enzyme turnover experiment. These studies indicate that the passage of indole from the α to the β subunit is impeded such that it builds up during a single turnover experiment.[151]

8.18.5 Novel Approaches for Detecting and Characterizing Enzyme Intermediates

8.18.5.1 Time-Resolved ESI–TOF–MS to Detect an Intermediate in the KDO8P Synthase Reaction

Our previous transient kinetic studies using rapid chemical quench suggested that the half-life for the KDO8P synthase catalytic reaction was approximately 10 ms.[48] As mentioned above, our strategy for the detection of the hypothesized hemiketal phosphate intermediate (see **Scheme 6**), involved the design of a novel rapid-mixing technique with high-resolution ESI–TOF–MS that would allow real-time monitoring of chemical catalysis for enzyme reaction times as short as 6–7 ms. The desired short time resolution was achieved by a combination of custom-designed electrospray probe, high flow rates, and minimal length, narrow-bore fused silica capillary tubing. This study focused on the simultaneous detection of substrates and products as well as the putative hemiketal phosphate intermediate that are bound to the enzyme as noncovalent complexes. This type of hemiketal phosphate species, while inferred in both chemical and enzymatic reaction, had not been directly observed.[156,157]

The catalytic reaction of KDO8PS with its natural substrates, PEP and A5P, was examined under single enzyme turnover conditions by rapidly mixing the E·PEP solution with a limiting amount of the second
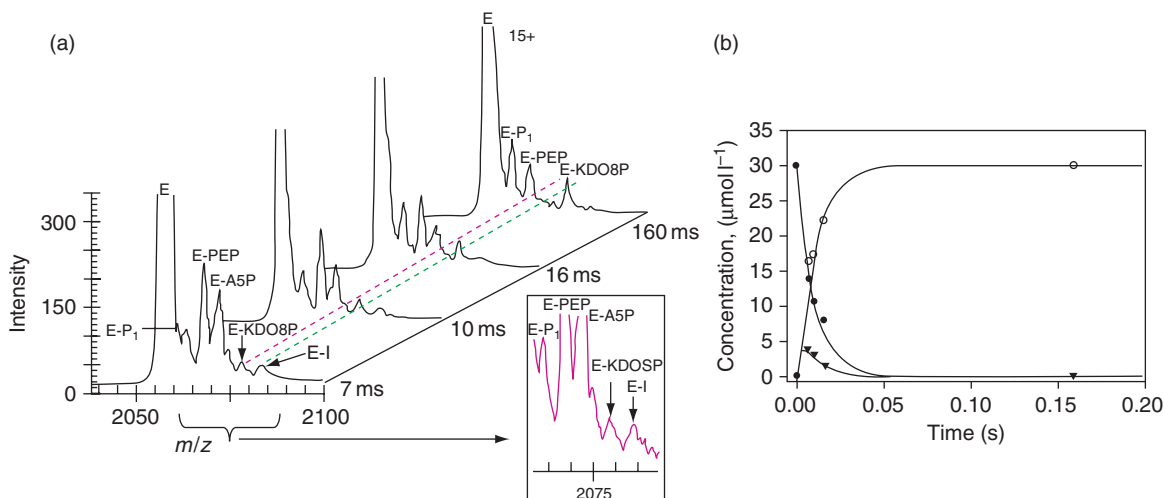
Figure 7   Monitoring catalysis for KDO8P synthase using ESI–TOF-MS. (a) Detection of an enzyme intermediate.
(b) Reaction kinetics using time-resolved ESI–TOF-MS illustrating that the formation and decay of the enzyme intermediate parallel the disappearance of substrate and formation of the product.

substrate, A5P. The reaction was monitored over several time ranges (7–160 ms) as shown in **Figure 7(a)** and quantitation of each of these species based upon the relative intensity provides an estimation of the reaction kinetics for substrate conversion into the product and a kinetically competent enzyme intermediate as illustrated in **Figure 7(b)**. The complete experimental details for this study are described in Li *et al.*[86,95]

This work demonstrates the feasibility of the time-resolved ESI–TOF–MS for the detection of low abundance, short-lived, and chemically unstable enzyme intermediates and opens up the potential of utilizing MS for performing real-time mechanistic enzymology.

8.18.5.2   Monitoring real-time phosphorylation kinetics with time-resolved ESI–TOF–MS

An increasing number of pathological states such as diabetes, obesity, aging, and age-related diseases (e.g., atherogenesis, hypertension, cancer, cardiovascular, and renal disease) have been linked to aberrations in signaling pathways. While there is a wealth of information on the cellular and molecular biology of many of the signaling pathways, the focus has been primarily on identifying signaling partners and delineating plausible downstream pathways. There is a surprising lack of information on the reaction kinetics of the critical early phosphorylation steps that occur in the second and millisecond time domains. In addition, conventional methods monitor the overall kinase phosphorylation and not the phosphorylation of each S/T/Y site. A series of *in vitro* studies published by our laboratory based on a combination of rapid kinetics and proteomics methodologies have shown that in the case of fibroblast growth factor receptor (FGFR1), the six tyrosines in the intracellular kinase domain are phosphorylated in a sequential fashion and that there are discrete sets of defined mono, di, and multiphosphorylated species that could act as molecular switches to recruit downstream partners in an orchestrated manner.[158] We have also applied the time-resolved ESI–TOF-MS method to monitor the increase in the phosphorylation level of FGFR1 and the phosphorylation of FRS2α-derived peptide substrate (**Figure 8**)

8.18.6   Conclusions

Over the past few years, more powerful analytical techniques have become available to study biological molecules and biological processes. These techniques allow greater sensitivity, higher accuracy, quantification, and more rigorous information regarding molecular structure and function. There is a clear trend to understand
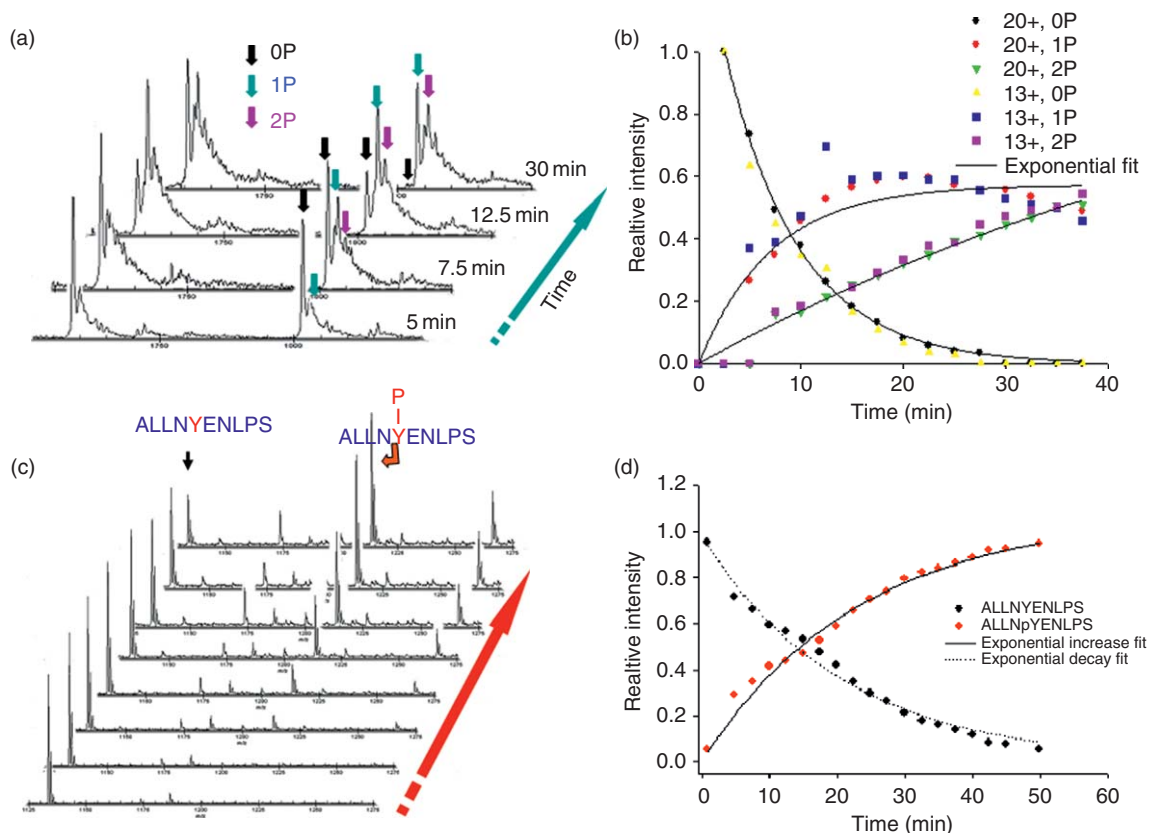
Figure 8 Analysis of phosphorylation kinetics by time-resolved ESI–TOF-MS. (a) FGFR1 autophosphorylation. A selection of MS spectra collected by time-resolved ESI–TOF–MS showing the formation of mono and diphosphorylated FGFR1 species at increasing reaction times. 0P, 1P, and 2P represent the FGFR1 in unphosphorylated, mono, and diphosphoryl states. (b) The relative intensities of unphosphorylated and phosphorylated species were plotted as function of time and the reaction rates were determined for each phosphorylation event (0P→1P: 0.14 min^{-1} and 1P→2P: 0.013 min^{-1}). 20+ and 13+ indicate data for two charge states. (c) A selection of MS spectra collected by time-resolved ESI–TOF-MS showing the phosphorylation of FRS2α-derived peptide by FGFR1. (d) The relative intensities of unphosphorylated and phosphorylated FRS2α ALLNYENLPS peptide were plotted as function of time.

biological molecules and biological processes in greater detail. Enzyme catalysis is at the heart of understanding the relationship between protein structure and function and is especially relevant in this postgenomic era to aid in unraveling protein function.[71] The paradigm for understanding enzyme catalysis involves a complete description of the reaction pathway at the enzyme active site and includes the structural elucidation and kinetic formation of multiple enzyme intermediates. Improved techniques to study enzyme catalysis will further our understanding and ultimately our ability to modulate enzyme reactions. The benefits to drug design, enzyme design, and mechanistic enzymology are clear as we begin to apply these novel methods to target enzymes for developing antibiotics as well as enzymes involved in protein signaling and cellular metabolism.

Acknowledgments

Abbreviations

A5P	D-arabinose 5-phosphate
DAHP	3-deoxy-D-*arabino*-2-heptulosonate acid-7-phosphate
EPSP	5-enolpyruvoyl shikimate-3-phosphate
ESI–MS	electrospray ionization–mass spectrometry
KDO8P	3-deoxy-D-*manno*-2-octulosonate acid-8-phosphate
NeuAc	*N*-acetyl-neuraminic acid
PEP	phosphoenolpyruvate
S3P	shikimate-3-phosphate
TOF	time of flight
UDP-GlcNAc	uridine diphosphate-*N*-acetyl glucosamine enolpyruvoyl transferase

References

1. K. A. Johnson, *Meth. Enzymol.* **1995**, *249*, 38–61.
2. C. A. Fierke; G. G. Hammes, *Meth. Enzymol.* **1995**, *249*, 3–37.
3. K. A. Johnson, *Meth. Enzymol.* **1986**, *134*, 677–705.
4. K. A. Johnson, *The Enzymes* **1992**, *20*, 1–61.
5. K. S. Anderson; K. A. Johnson, *Chem. Rev.* **1990**, *90*, 1131–1149.
6. K. S. Anderson, *Meth. Enzymol.* **1999**, *308*, 111–145.
7. K. Anderson; K. Johnson, *J. Biol. Chem.* **1990**, *265*, 5567–5572.
8. K. S. Anderson, Detection and Characterization of Enzyme Intermediates: Utility of Rapid Chemical Quench Methodology and Single Enzyme Turnover Experiments. In *Kinetic Analysis of Macromolecules: Practical Approach Series*; K. A. Johnson, Ed.; Oxford University Press, Inc.: New York, 2003; Vol. 267, pp 19–44.
9. H. K. Wiskind, In *Rapid Mixing and Sampling Techniques in Biochemistry*; B. Chance; R. H. Eisenhardt; Q. H. Gibson; K. Lonberg-Holm, Eds.; Academic Press: New York, 1964, p 355.
10. A. L. Tsai; V. Berka; R. J. Kulmacz; G. Wu; G. Palmer, *Anal. Biochem.* **1998**, *264*, 165–171.
11. D. P. Ballou, *Fast Methods in Physical Biochemistry and Cell Biology*; Elsevier: Amsterdam, New York, 1983.
12. R. J. Appleyard; W. A. Shuttleworth; J. N. Evans, *Biochemistry* **1994**, *33*, 6812–6821.
13. R. Evande; S. Ojha; R. Banerjee, *Arch. Biochem. Biophys.* **2004**, *427*, 188–196.
14. R. Banerjee; C. G. Zou, *Arch. Biochem. Biophys.* **2005**, *433*, 144–156.
15. C. Veeger, *J. Inorg. Biochem.* **2002**, *91*, 35–45.
16. E. M. Isin; F. P. Guengerich, *J. Biol. Chem.* **2007**, *282*, 6863–6874.
17. T. Spolitak; J. H. Dawson; D. P. Ballou, *J. Biol. Chem.* **2005**, *280*, 20300–20309.
18. T. Spolitak; J. H. Dawson; D. P. Ballou, *J. Inorg. Biochem.* **2006**, *100*, 2034–2044.
19. C. C. Fjeld; W. T. Birdsong; R. H. Goodman, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 9202–9207.
20. C. Furdui; S. W. Ragsdale, *J. Biol. Chem.* **2000**, *275*, 28494–28499.
21. D. K. Shoemark; M. J. Cliff; R. B. Sessions; A. R. Clarke, *FEBS J.* **2007**, *274*, 2738–2748.
22. S. Ueda; M. Oda; S. Imamura; M. Ohnishi, *Eur. J. Biochem.* **2004**, *271*, 1774–1780.
23. A. Gutierrez; L. Y. Lian; C. R. Wolf; N. S. Scrutton; G. C. Roberts, *Biochemistry* **2001**, *40*, 1964–1975.
24. I. G. Denisov; T. M. Makris; S. G. Sligar; I. Schlichting, *Chem. Rev.* **2005**, *105*, 2253–2277.
25. C. Furdui; S. W. Ragsdale, *Biochemistry* **2002**, *41*, 9921–9937.
26. J. T. Jarrett; M. Amaratunga; C. L. Drennan; J. D. Scholten; R. H. Sands; M. L. Ludwig; R. G. Matthews, *Biochemistry* **1996**, *35*, 2464–2475.
27. R. Padmakumar; R. Padmakumar; R. Banerjee, *Biochemistry* **1997**, *36*, 3713–3718.
28. S. Zhao; D. L. Roberts; S. W. Ragsdale, *Biochemistry* **1995**, *34*, 15075–15083.
29. A. W. Tepper; L. Bubacco; G. W. Canters, *J. Biol. Chem.* **2004**, *279*, 13425–13434.
30. K. S. Anderson; E. W. Miles; K. A. Johnson, *J. Biol. Chem.* **1991**, *266*, 8020–8033.
31. K. S. Anderson; J. A. Sikorski; A. Benesi; K. A. Johnson, *Biochemistry* **1988**, *27*, 1604–1610.
32. C. Gondeau; L. Chaloin; A. Varga; B. Roy; P. Lallemand; C. Perigaud; T. Barman; M. Vas; C. Lionne, *Biochemistry* **2008**, *47*, 3462–3473.
33. Y. Zheng; F. Mamdani; D. Toptygin; L. Brand; J. T. Stivers; P. A. Cole, *Biochemistry* **2005**, *44*, 10501–10509.
34. H. Song; D. L. Chen; R. F. Ismagilov, *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 7336–7356.
35. H. Song; R. F. Ismagilov, *J. Am. Chem. Soc.* **2003**, *125*, 14613–14619.
36. H. Song; J. D. Tice; R. F. Ismagilov, *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 768–772.
37. A. R. de Boer; J. M. Alcaide-Hidalgo; J. G. Krabbe; J. Kolkman; C. N. van Emde Boas; W. M. Niessen; H. Lingeman; H. Irth, *Anal. Chem.* **2005**, *77*, 7894–7900.
38. A. R. de Boer; B. Bruyneel; J. G. Krabbe; H. Lingeman; W. M. Niessen; H. Irth, *Lab. Chip.* **2005**, *5*, 1286–1292.
39. A. R. de Boer; T. Letzel; H. Lingeman; H. Irth, *Anal. Bioanal. Chem.* **2005**, *381*, 647–655.
40. A. R. de Boer; T. Letzel; D. A. van Elswijk; H. Lingeman; W. M. Niessen; H. Irth, *Anal. Chem.* **2004**, *76*, 3155–3161.

41. M. B. Kerby; J. Lee; J. Ziperstein; A. Tripathi, *Biotechnol. Prog.* **2006**, *22*, 1416–1425.
42. M. B. Kerby; R. S. Legge; A. Tripathi, *Anal. Chem.* **2006**, *78*, 8273–8280.
43. L. Skipper; W. H. Campbell; J. A. Mertens; D. J. Lowe, *J. Biol. Chem.* **2001**, *276*, 26995–27002.
44. J. H. Kim; R. Hille, *J. Biol. Chem.* **1993**, *268*, 44–51.
45. E. Murakami; U. Deppenmeier; S. W. Ragsdale, *J. Biol. Chem.* **2001**, *276*, 2432–2439.
46. J. M. Bollinger, Jr.; W. H. Tong; N. Ravi; B. H. Huynh; D. E. Edmondson; J. A. Stubbe, *Meth. Enzymol.* **1995**, *258*, 278–303.
47. R. C. Kunz; Y. C. Horng; S. W. Ragsdale, *J. Biol. Chem.* **2006**, *281*, 34663–34676.
48. P. H. Liang; J. Lewis; K. S. Anderson; A. Kohen; F. W. D'Souza; Y. Benenson; T. Baasov, *Biochemistry* **1998**, *37*, 16390–16399.
49. B. E. Eser; E. W. Barr; P. A. Frantom; L. Saleh; J. M. Bollinger, Jr.; C. Krebs; P. F. Fitzpatrick, *J. Am. Chem. Soc.* **2007**, *129*, 11334–11335.
50. D. G. Fujimori; E. W. Barr; M. L. Matthews; G. M. Koch; J. R. Yonce; C. T. Walsh; J. M. Bollinger, Jr.; C. Krebs; P. J. Riggs-Gelasco, *J. Am. Chem. Soc.* **2007**, *129*, 13408–13409.
51. M. R. Seyedsayamdost; J. Xie; C. T. Chan; P. G. Schultz; J. Stubbe, *J. Am. Chem. Soc.* **2007**, *129*, 15060–15071.
52. L. Williams; F. Fan; J. S. Blanchard; F. M. Raushel, *Biochemistry* **2008**, *47*, 4843–4850.
53. S. D. Lahiri; G. Zhang; D. Dunaway-Mariano; K. N. Allen, *Science* **2003**, *299*, 2067–2071.
54. G. Wille; D. Meyer; A. Steinmetz; E. Hinze; R. Golbik; K. Tittmann, *Nat. Chem. Biol.* **2006**, *2*, 324–328.
55. D. Perez-Bendito; A. Gomez-Hens; A. Gaikwad, *Clin. Chem.* **1994**, *40*, 1489–1493.
56. D. Perez-Bendito; A. Gomez-Hens; M. Silva, *J. Pharm. Biomed. Anal.* **1996**, *14*, 917–930.
57. H. Fabian; D. Naumann, *Methods* **2004**, *34*, 28–40.
58. E. Z. Eisenmesser; O. Millet; W. Labeikovsky; D. M. Korzhnev; M. Wolf-Watz; D. A. Bosco; J. J. Skalicky; L. E. Kay; D. Kern, *Nature* **2005**, *438*, 117–121.
59. D. Kern; E. Z. Eisenmesser; M. Wolf-Watz, *Meth. Enzymol.* **2005**, *394*, 507–524.
60. W. Labeikovsky; E. Z. Eisenmesser; D. A. Bosco; D. Kern, *J. Mol. Biol.* **2007**, *367*, 1370–1381.
61. K. Tittmann; H. Neef; R. Golbik; G. Hubner; D. Kern, *Biochemistry* **2005**, *44*, 8697–8700.
62. S. M. Clark; L. Konermann, *Anal. Chem.* **2004**, *76*, 7077–7083.
63. S. G. Codreanu; J. E. Ladner; G. Xiao; N. V. Stourman; D. L. Hachey; G. L. Gilliland; R. N. Armstrong, *Biochemistry* **2002**, *41*, 15161–15172.
64. S. G. Codreanu; L. C. Thompson; D. L. Hachey; H. W. Dirr; R. N. Armstrong, *Biochemistry* **2005**, *44*, 10605–10612.
65. J. Pan; A. C. Rintala-Dempsey; Y. Li; G. S. Shaw; L. Konermann, *Biochemistry* **2006**, *45*, 3005–3013.
66. D. J. Wilson; L. Konermann, *Anal. Chem.* **2004**, *76*, 2537–2543.
67. D. B. Northrop; F. B. Simpson, *Biorg. Med. Chem.* **1997**, *5*, 641–644.
68. E. Gelpi, *J. Chromatogr.* **1995**, *703*, 59–80.
69. J. C. Schwartz; I. Jardine, *Meth. Enzymol.* **1996**, *270*, 552–586.
70. G. Siuzdak, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11290–11297.
71. F. Sobott; C. V. Robinson, *Curr. Opin. Struct. Biol.* **2002**, *12*, 729–734.
72. D. S. Ashton; C. R. Beddell; D. J. Cooper; B. N. Green; R. W. Oliver; K. J. Welham, *FEBS Lett.* **1991**, *292*, 201–204.
73. R. Chavez; B. Bothner; C. Strupp; D. Hilvert; G. Siuzdak, *Proceedings of the American Society for Mass Spectrometry. 44th Conference on Mass Spectrometry and Allied Topics.* Portland, 1996, p 724.
74. F. Y. Hsieh; X. Tong; T. Wachs; B. Ganem; J. Henion, *Anal. Biochem.* **1995**, *229*, 20–25.
75. E. D. Lee; W. Much; J. D. Henion; T. R. Covey, *J. Am. Chem. Soc.* **1989**, *111*, 4600–4604.
76. J. Nairn; T. Krell; J. R. Coggins; A. R. Pitt; L. A. Fothergill-Gilmore; R. Walter; N. C. Price, *FEBS Lett.* **1995**, *359*, 192–194.
77. J. W. Sam; X. Tang; R. S. Magliozzo; J. Peisach, *J. Am. Chem. Soc.* **1995**, *117*, 1012–1018.
78. J. W. Sam; X. Tang; J. Peisach, *J. Am. Chem. Soc.* **1994**, *116*, 5250–5256.
79. C. T. Houston; W. P. Taylor; T. S. Widlanski; J. P. Reilly, *Anal. Chem.* **2000**, *72*, 3311–3319.
80. D. L. Zechel; L. Konermann; S. G. Withers; D. J. Douglas, *Biochemistry* **1998**, *37*, 7664–7669.
81. J. W. Gross; A. D. Hegeman; M. M. Vestling; P. A. Frey, *Biochemistry* **2000**, *39*, 13633–13640.
82. B. M. Kolakowski; L. Konermann, *Anal. Biochem.* **2001**, *292*, 107–114.
83. L. Konermann; B. A. Collings; D. J. Douglas, *Biochemistry* **1997**, *36*, 5554–5559.
84. L. Konermann; F. I. Rosell; A. G. Mauk; D. J. Douglas, *Biochemistry* **1997**, *36*, 6448–6454.
85. A. Paiva; R. F. Tilton; G. C. Crooks; P.-H. Liang; K. S. Anderson, *Biochemistry* **1997**, *36*, 15472–15476.
86. Z. Li; A. Sau; S. Shen; C. Whitehouse; T. Baasov; K. S. Anderson, *J. Am. Chem. Soc.* **2003**, *125*, 9938–9939.
87. B. Bothner; R. Chavez; J. Wei; C. Strupp; Q. Phung; A. Schneemann; G. Siuzdak, *J. Biol. Chem.* **2000**, *275*, 13455–13459.
88. P. V. Attwood; M. A. Geeves, *Anal. Biochem.* **2004**, *334*, 382–389.
89. K. D. Greis, *Mass Spectrom. Rev.* **2007**, *26*, 324–339.
90. Y. H. Liu; L. Konermann, *FEBS Lett.* **2006**, *580*, 5137–5142.
91. M. Loog; B. Ek; N. Oskolkov; A. Narvanen; J. Jarv; P. Ek, *J. Biomol. Screen.* **2005**, *10*, 320–328.
92. K. P. Nichols; H. J. Gardeniers, *Anal. Chem.* **2007**, *79*, 8699–8704.
93. N. Pi; M. B. Hoang; H. Gao; J. D. Mougous; C. R. Bertozzi; J. A. Leary, *Anal. Biochem.* **2005**, *341*, 94–104.
94. N. Pi; J. A. Leary, *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 233–243.
95. Z. Li; A. K. Sau; C. M. Furdui; K. S. Anderson, *Anal. Biochem.* **2005**, *343*, 35–47.
96. C. T. Walsh; T. E. Benson; D. H. Kim; W. J. Lees, *Curr. Biol.* **1996**, *3*, 1074.
97. E. Haslam, *Shikimic Acid: Metabolism and Metabolites*; John Wiley and Sons: New York, 1993.
98. H. C. Steinrucken; N. Amrhein, *Biochem. Biophys. Res. Commun.* **1980**, *94*, 1207.
99. F. M. Kahan; J. S. Kahan, *Ann. N. Y. Acad. Sci.* **1974**, *235*, 364.
100. C. Wanke; N. Amrhein, *Eur. J. Biochem.* **1993**, *218*, 861.
101. H. C. Steinrucken; N. Amrhein, *Eur. J. Biochem.* **1984**, *143*, 351–357.
102. C. Raetz; W. Dowhan, *J. Biol. Chem.* **1990**, *265*, 1235–1238.
103. W. F. Vann; J. J. Tavarez; J. Crowley; E. Vimr; R. P. Silver, *Glycobiology* **1997**, *7*, 697–701.
104. K. S. Anderson, *Arch. Biochem. Biophys.* **2005**, *433*, 47–58.

105. P. Cassidy; F. Kahan, *Biochemistry* **1973**, *12*, 1364–1374.
106. W. Bondinell; P. Vnek; P. Knowles; M. Sprecher; D. B. Sprinson, *J. Biol. Chem.* **1971**, *246*, 6191–6196.
107. D. L. Anton; L. Hedstrom; S. M. Fish; R. H. Abeles, *Biochemistry* **1983**, *22*, 5903–5908.
108. C. E. Grimshaw; S. G. Sogo; J. R. Knowles, *J. Biol. Chem.* **1982**, *257*, 596.
109. K. S. Anderson; J. A. Sikorski; K. A. Johnson, *Biochemistry* **1988**, *27*, 7395–7405.
110. J. Lewis; K. A. Johnson; K. S. Anderson, *Biochemistry* **1998**, *38*, 7372–7379.
111. K. S. Anderson; J. A. Sikorski; K. A. Johnson, *J. Am. Chem. Soc.* **1988**, *110*, 6577.
112. K. S. Anderson; R. D. Sammons; G. C. Leo; J. A. Sikorski; A. J. Benesi; K. A. Johnson, *Biochemistry* **1990**, *29*, 1460–1465.
113. G. C. Leo; J. A. Sikorski; R. D. Sammons, *J. Am. Chem. Soc.* **1990**, *112*, 1653–1654.
114. D. Studelska; L. McDowell; M. Espe; C. Klug; J. Schaefer, *Biochemistry* **1997**, *36*, 15555–15560.
115. D. L. Jakeman; D. J. Mitchell; W. A. Shuttleworth; J. N. S. Evans, *Biochemistry* **1998**, *37*, 12012–12019.
116. R. Zemell; R. Anwar, *J. Biol. Chem.* **1975**, *25*, 4959–4964.
117. J. Marquardt; E. Brown; C. T. Walsh; K. S. Anderson, *J. Am. Chem. Soc.* **1993**, *115*, 10398–10399.
118. D. Kim; W. Lees; T. Haley; C. T. Walsh, *J. Am. Chem. Soc.* **1995**, *117*, 1494–1499.
119. E. Brown; J. Marquardt; J. Lee; C. T. Walsh; K. S. Anderson, *Biochemistry* **1994**, *33*, 10638–10645.
120. L. Hedstrom; R. Abeles, *Biochem. Biophys. Res. Comm.* **1988**, *157*, 816–820.
121. G. Dotson; P. Nanjappan; M. Reily; R. Woodard, *Biochemistry* **1993**, *32*, 12392–12397.
122. G. Dotson; R. K. Dua; J. C. Clemens; E. W. Wooten; R. W. Woodard, *J. Biol. Chem.* **1995**, *23*, 13698–13705.
123. H. S. Duewel; S. Radaev; J. Wang; R. W. Woodard; D. L. Gatti, *J. Biol. Chem.* **2001**, *276*, 8393–8402.
124. J. Wang; H. S. Duewel; R. W. Woodard; D. L. Gatti, *Biochemistry* **2001**, *40*, 15676–15683.
125. O. Asojo; J. Friedman; N. Adir; V. Belakhov; Y. Shoham; T. Baasov, *Biochemistry* **2001**, *40*, 6326–6334.
126. A. Kohen; A. Jakob; T. Baasov, *Eur. J. Biochem.* **1992**, *208*, 443.
127. A. Kohen; R. Berkovich; V. Belakhov; T. Baasov, *Bioorg. Med. Chem. Lett.* **1993**, *3*, 1577.
128. P. Ray; J. Kelsey; E. Bigham; C. Benedict; T. Miller, *ACS Symp. Ser.* **1983**, *231*, 141–170.
129. T. Baasov; A. Kohen, *J. Am. Chem. Soc.* **1995**, *117*, 6165–6174.
130. Y. Benenson; V. Belakhov; T. Baasov, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2901–2906.
131. S. Sheffer-Dee-Noor; V. Belakhov; T. Baasov, *Bioorg. Med. Chem. Lett.* **1993**, *3*, 1583–1588.
132. T. Baasov; S. Sheffer-Dee-Noor; A. Kohen; A. Jakob; V. Belakhov, *Eur. J. Biochem.* **1993**, *217*, 991–999.
133. Z. Y. Zhang, *Curr. Opin. Chem. Biol.* **2001**, *5*, 416–423.
134. T. Hunter, *Cell* **2000**, *100*, 113–127.
135. Z. Y. Zhang; R. L. VanEtten, *J. Biol. Chem.* **1991**, *266*, 1516–1525.
136. K. L. Guan; J. E. Dixon, *J. Biol. Chem.* **1991**, *266*, 17026–17030.
137. H. H. Pas; G. H. Meyer; W. H. Kruizinga; K. S. Tamminga; R. P. van Weeghel; G. T. Robillard, *J. Biol. Chem.* **1991**, *266*, 6690–6692.
138. Y. Y. Wo; M. M. Zhou; P. Stevis; J. P. Davis; Z. Y. Zhang; R. L. Van Etten, *Biochemistry* **1992**, *31*, 1712–1721.
139. E. W. Miles, *Adv. Enzymol.* **1979**, *49*, 127–186.
140. E. W. Miles, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1991**, *64*, 93–172.
141. E. W. Miles, *Subcell. Biochem.* **1995**, *24*, 207–254.
142. W. M. Matchett, *J. Biol. Chem.* **1974**, *249*, 4041–4049.
143. J. A. Demoss, *Biochim. Biophys. Acta* **1962**, *62*, 279–293.
144. C. C. Hyde; S. A. Ahmed; E. A. Padlan; E. W. Miles; D. R. Davies, *J. Biol. Chem.* **1988**, *263*, 17857–17871.
145. C. C. Hyde; E. W. Miles, *Biotechnology* **1990**, *8*, 27–31.
146. A. Lane; K. Kirschner, *Biochemistry* **1991**, *30*, 479–484.
147. P. Brzovic; K. Ngo; M. Dunn, *Biochemistry* **1992**, *31*, 3831–3839.
148. M. F. Dunn; V. Aguilar; P. Brzovic; W. F. Drewe, Jr.; K. F. Houben; C. A. Leja; M. Roy, *Biochemistry* **1990**, *29*, 8598–8607.
149. W. F. Drewe; M. F. Dunn, *Biochemistry* **1986**, *25*, 2494–2501.
150. A. Lane; K. Kirschner, *J. Biochem.* **1983**, *129*, 571–582.
151. K. S. Anderson; A. Kim; H. M. Quillen; E. Sayers; X. Yang; E. Miles, *J. Biol. Chem.* **1995**, *270*, 29936–29944.
152. I. P. Crawford; J. Ito, *Proc. Natl. Acad. Sci. U.S.A.* **1964**, *51*, 980.
153. U. Banik; D. M. Zhu; P. B. Chock; E. W. Miles, *Biochemistry* **1995**, *34*, 12704–12711.
154. S. Rhee; K. D. Parris; C. C. Hyde; S. A. Ahmed; E. W. Miles; D. R. Davies, *Biochemistry* **1997**, *36*, 7664–7680.
155. P. Brzovic; A. Kayastha; E. Miles; M. F. Dunn, *Biochemistry* **1992**, *31*, 1180–1190.
156. A. R. Rendina; J. D. Hermes; W. W. Cleland, *Biochemistry* **1984**, *23*, 5148–5156.
157. J. Haddad; S. Vakulendk; S. Mobashery, *J. Am. Chem. Soc.* **1999**, *121*, 11922–11923.
158. C. M. Furdui; E. Lew; J. Schlessinger; K. S. Anderson, *Mol. Cell.* **2006**, *17*, 1–7.

Biographical Sketches

Cristina M. Furdui completed her undergraduate studies in Romania at Babes-Bolyai University. In 1997, she entered the graduate program at the Department of Biochemistry, University of Nebraska at Lincoln (UNL) where under the supervision of Dr. Stephen W. Ragsdale, she was trained in biochemistry and enzyme kinetics and obtained her Ph.D. in 2002. While at UNL, she received several awards including the Folsom Distinguished Doctoral Dissertation Award (2003). After graduation, Dr. Furdui continued her training in enzyme kinetics, mass spectrometry, and molecular signaling as a postdoctoral associate and later research associate scientist with Dr. Karen S. Anderson (Department of Pharmacology, Yale University). In September 2006, Dr. Furdui was appointed as assistant professor in the Department of Internal Medicine, Section on Molecular Medicine at Wake Forest University School of Medicine. Research in her laboratory is now directed toward applying advanced kinetics and systems biology methodologies to (1) investigate the timing of signaling events in the propagation of receptor tyrosine kinases signaling, (2) quantify the effect of oncogenic mutations and oxidation on the re wiring of these signaling networks under pathogenic conditions, and (3) interface time-resolved mass spectrometry with microfluidics technology to develop new nanokinetics platforms for quantitative monitoring of rapid enzyme kinetics/drug screening assays to further the understanding of potential drug targets at the molecular level.

Karen S. Anderson received her B.S. in microbiology/chemistry from East Tennessee State University. She earned her Ph.D. in organic medicinal chemistry from The Ohio State University under the direction of Professor Duane D. Miller. She then joined Monsanto Co. During her time there, she rose to the rank of Senior Research Specialist and was part of a biorational inhibitor design group that was responsible for research leading to an understanding of the mechanism of action for Monsanto's billion dollar herbicide, Roundup. This research involved a multidisciplinary approach including structural biology and mechanistic

enzymology. Seeking to use this strategy on more medically related problems, she then joined the faculty as an Assistant Professor in the Pharmacology Department at Yale University School of Medicine. Currently a full Professor of Pharmacology at Yale, she has been the recipient of several awards including the Dean's Young Faculty Award, the Hull Award, the Breast Cancer Initiative Award, and ETSU and OSU Alumni Awards. She has served on several national review panels including National Institute of Health and American Cancer Society study sections and has been a reviewer for a number of journals. The primary emphasis of Dr. Anderson's research focuses on developing an understanding of enzymatic reactions and receptor–ligand interactions at a molecular level using a combination of kinetic and structural techniques including rapid transient kinetics, NMR, X-ray crystallography, and mass spectrometry. The ultimate goal of this research is to develop an in-depth understanding of how enzymes function and thereby provide a more effective means of modulating their function and developing novel therapies.

8.19 Bisubstrate Analog Inhibitors

Patrick A. Frantom and John S. Blanchard, Albert Einstein College of Medicine, Bronx, NY, USA

8.19.1 Introduction

Two critical aspects of any effective therapeutic agent are its specificity and affinity for the intended target. The use of bi- or multisubstrate analogs as specific and potent inhibitors for enzymes was proposed to address these two concerns.[1] By making an analog to both substrates, the specificity for the target enzyme should be increased over a single substrate analog. This feature is especially relevant to substrate analogs in metabolic/biosynthetic pathways where the product of one enzyme is a substrate for the next and for analogs of broad-spectrum substrates such as sugar nucleotides or coenzyme A, which might nonspecifically affect entire families of enzymes. Bisubstrate analogs also have the potential for greater affinity toward a target based on the proximity effect of tethering the two substrate analogs together, which should harness the binding energy of each individual substrate into a single molecule.[2] In some cases, the binding energy gained from the covalent attachment of two substrates has been shown to be greater than the product of the individual binding constants.[3]

While developed as possible therapeutics, bisubstrate analogs have found great utility in the dissection and characterization of enzyme structure and mechanism. As discussed below, bisubstrate analogs have been used extensively in structural studies, where the use of natural substrates would result in catalysis, to investigate the architecture of the active site at the Michaelis complex, and to define structural changes at the active site produced by allosteric effectors. In some cases, bisubstrate analogs that are formed during the reaction (a type of mechanism-based inhibitor) can help to support or eliminate proposed chemical mechanisms.

There are a number of classifications for enzyme inhibitors (e.g., bisubstrate analog, suicide substrate, mechanism-based inactivator, transition-state analog, slow-onset, tight binding) all of which have some overlap in their definitions. For the purposes of this chapter, the term bisubstrate analog will refer to compounds where the inhibitory power is derived from the similarity in shape and electrostatics to the natural substrates regardless of whether the final compound is synthesized at the bench or during the catalytic cycle of the enzyme. When dealing with enzymes that act as a 'marriage broker' between two substrates without becoming a part of the reaction covalently, the line between a bisubstrate analog and a transition-state analog can become blurred. Over a decade ago, Bartlett and Marlowe[4] wrote, "the designation 'transition state analogue' in truth is a revelation of the underlying design concept, which led to the inhibitor, or an *a posteriori* rationalization of its tight binding".[4] However, current methods combining the use of isotope effects with advances in computational modeling make it

easier than ever to describe the transition-state structure for an enzymatic reaction and design inhibitors with a specific structure on mind.[5] The analogs described in this chapter were generally developed as bisubstrate analogs, but cases where information on the character of the transition state can be gained are also included.
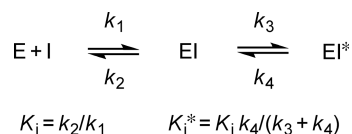
This chapter begins with a brief overview of the practical aspects of performing and analyzing experiments common for bisubstrate analog inhibitor characterization. The experiments described in the overview will be put to practical use in numerous examples from different enzyme classes where bisubstrate analog inhibitors have assisted in the understanding of enzyme structure and function. We have focused on relatively recent examples from the literature and have specifically left out descriptions of more canonical bisubstrate analog inhibitors (e.g., adenosine-p$_5$-adenosine, phosphonoacetyl-L-aspartate) that have been substantially discussed and can be found in most biochemistry textbooks.[1] Finally, an in-depth review of the discovery and elucidation of mechanism of action for several important bisubstrate analog-based therapeutics demonstrates the potential and incentive for continued effort toward design of bisubstrate analog inhibitors.

8.19.2   Design and Analysis of Bisubstrate Analog Inhibitors

The development of most bisubstrate inhibitors follows a simple pattern. A compound is synthesized, chemically or enzymatically, where the natural substrates are linked together with a simple linker group (i.e., acetyl, peptidyl, or methylene spacer) and the inhibitory properties of that molecule are tested. Once moderate inhibition of the enzyme activity is demonstrated, a second generation of analogs is synthesized where each moiety of the analog can be individually or simultaneously altered to change the shape or electrostatics to probe for active site-specific interactions, which increase the inhibitor's affinity for the enzyme. Another common point of variation is the length or chemical identity of the linker group; most enzymes prefer a linker of specific length to mimic the correct spacing of the substrates in the active site. One concern when dealing with the linkage of bisubstrate inhibitors is the prevention of hydrolysis. This problem manifests itself in multiple ways such as the rapid loss of inhibitory power or the inability to survive crystallization conditions in structural studies. Thus, if the proposed experiments require a long-lived inhibitor more hydrolysis-resistant bonds such as C–C or phosphonate moieties should be utilized when possible.

While great success has been demonstrated in *in vitro* inhibition studies using bisubstrate analogs, little of that success has been transferred to *in vivo* studies. The main culprit for the lack of *in vivo* efficacy appears to be lack of uptake by the cell, supported by the observation that charged molecules usually have difficulty penetrating the cell membrane. In order to address this challenge, most second-generation analogs focus on charge neutralization as well as optimization of the inhibitor's potency. Another approach that has been employed effectively to circumvent the problem of cellular uptake is the use of pro-analogs that, once inside the cell, can be processed to the original analog.

A number of bisubstrate analog inhibitors discussed in this chapter are kinetically described as slow-onset inhibitors.[6] The most common model for this type of inhibition is shown in **Scheme 1**. Free enzyme binds the inhibitor to form an EI complex with a dissociation constant of K_i. Then the EI complex undergoes a conformational change to the more stable EI* complex resulting in tighter binding of the inhibitor. $K_i{}^*$ describes the equilibrium between free E and the EI and EI* complexes. In order for slow-onset behavior to be detected, the equilibration between EI and EI* must necessarily be slow relative to the initial formation of the EI complex. The experimental hallmark of slow-onset inhibition is the appearance of nonlinear activity as a function of time in the presence of inhibitor resulting in an initial fast phase followed by a slower steady-state rate. In order to determine if inhibition is reversible and not due to covalent labeling and inactivation of the enzyme, preincubation of the enzyme with the inhibitor followed by rapid dilution should display the opposite

$$E + I \underset{k_2}{\overset{k_1}{\rightleftharpoons}} EI \underset{k_4}{\overset{k_3}{\rightleftharpoons}} EI^*$$

$$K_i = k_2/k_1 \qquad K_i{}^* = K_i\, k_4/(k_3 + k_4)$$

Scheme 1

kinetic result (i.e., an initial slow phase followed by an increase to the uninhibited steady-state rate). Inhibitors that display slow-onset kinetics are favored due to their unique pharmacokinetics. A competitive inhibitor for an enzyme in a metabolic pathway can eventually be overcome by the buildup of the upstream metabolic product (i.e., the substrate for the inhibited enzyme). However, the presence of elevated substrate levels will not affect the equilibrium between the EI and EI* complex, thus not altering the potency of the inhibitor. This also results in longer residency times for the inhibitor on the target.

8.19.3 Gcn5-Related *N*-Acetyltransferases

Bisubstrate analogs have been extensively utilized in the Gcn5-related *N*-acetyltransferase (GNAT) family due to their conserved sequential kinetic mechanisms and the chemical functionality of the acetyl CoA substrate. This family of enzymes has become very significant in the past decade with members participating in mechanisms of gene expression through histone modification[7] and playing a role in antibiotic resistance mechanisms.[8] In addition to probing the structure/function of the enzymes, a method for determining the kinetic mechanism with the aid of a bisubstrate inhibitor has also been established.[9]

8.19.3.1 Aminoglycoside *N*-Acetyltransferases

Aminoglycosides are broad-spectrum antibiotics, which bind to the 16S rRNA at the tRNA acceptor A site.[8] When aminoglycosides are bound, the ability of the acceptor site to distinguish between correct and incorrect tRNAs is diminished leading to misincorporation of amino acids into polypeptide chains, ultimately resulting in cell death.[10] The canonical structure of an aminoglycoside is shown in **Figure 1**. Aminoglycosides have a central aminocyclitol (usually 2-deoxystreptamine) linked to one or more amino sugars by pseudoglycosidic bonds and are biosynthesized by large gene clusters.[11,12] In clinical settings, the primary mechanism of resistance to aminoglycoside antibiotics is modification of the drug by various enzymes. One positive outcome of this mechanism of antibiotic resistance is that while expression of these modifying enzymes renders the bacteria insensitive to specific aminoglycosides, the actual target in the bacteria is susceptible to inhibition by the drug. Thus if inhibitors specific to the drug-modifying enzymes can be found, the potency of the original antibiotic can be restored.
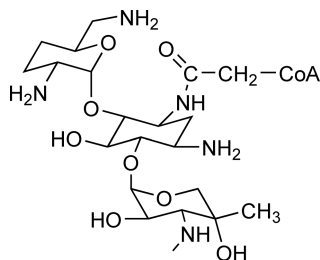
Aminoglycoside *N*-acetyltransferases catalyze the acetyl-CoA-dependent acetylation of one of the four amino groups of typical aminoglycosides. One of the earliest reports of a bisubstrate analog inhibitor for an aminoglycoside acetyltransferase (AAC) was reported in 1979 for gentamycin acetyltransferase.[13] The inhibitor was enzymatically synthesized by initially using chloro-acetyl-CoA as a substrate to generate chloro-acetylated gentamycin, followed by thiol-capture to create CoA-Ac-gentamicin (**Figure 2**). The inhibitor had a K_i of 0.5 nmol l^{-1} and displayed slow-onset inhibition. Unfortunately, the compound did not have an *in vivo* effect most probably due to its inability to cross the cell membrane, a common problem of bisubstrate analogs.

More recently, the Auclair group has reported a more extensive investigation of bisubstrate analog inhibitors toward aminoglycoside 6′-*N*-acetyltransferase (AAC(6′)-Ii) from *Enterococcus faecium*. AAC(6′)-Ii is chromosomally encoded in *E. faecium* and has been shown to use an ordered kinetic mechanism with acetyl-CoA binding
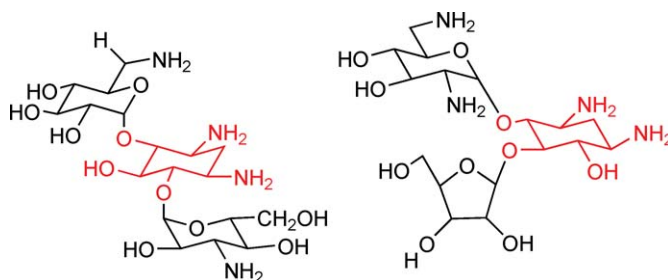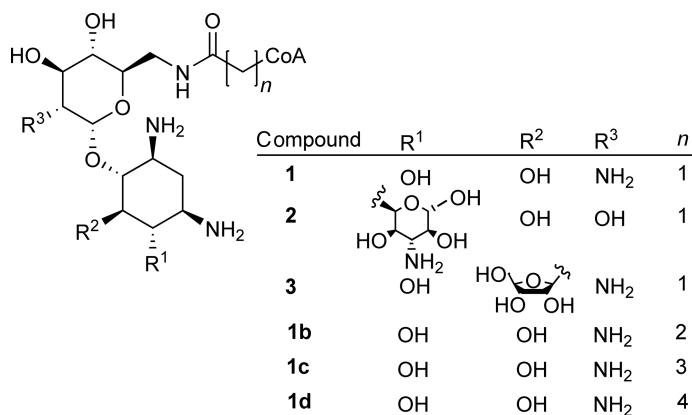

Figure 1 Chemical structure of the common 4,6- and 4,5-substituted deoxystreptamine aminoglycosides ribostamycin and kanamycin A.
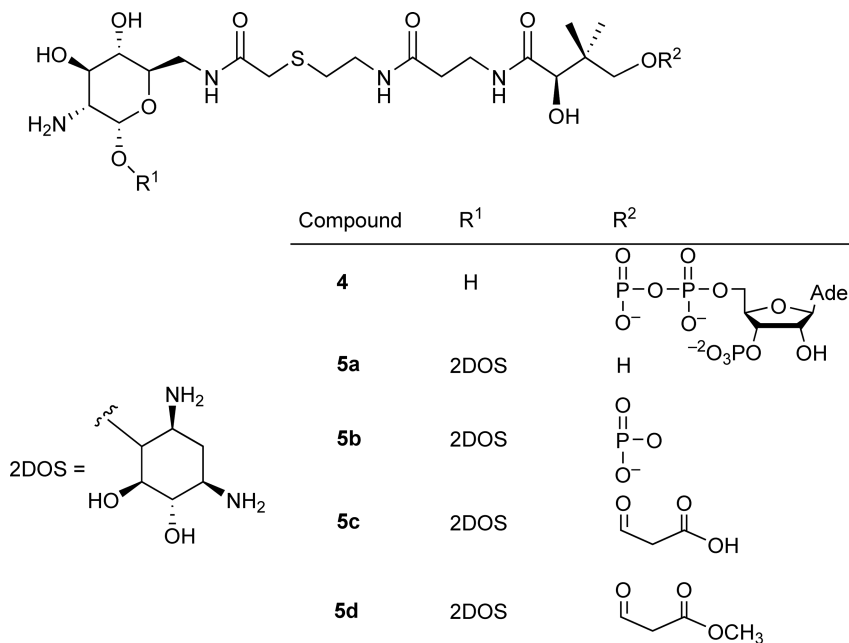
Figure 2   Chemical structure of the bisubstrate analog inhibitor of gentamycin acetyltransferase.

Figure 3   First-generation bisubstrate analog inhibitors of aminoglycoside acetyltransferases.

first, followed by the aminoglycoside.[14] Initially, three potential bisubstrate inhibitors (**Figure 3**) were chemically synthesized by linking CoA to the acetylated 6′ amino group of neamine (**1**), kanamycin A (**2**), and ribostamycin (**3**) (all substrates for AAC(6′)-Ii) with a methylene linker.[15] Each of these bisubstrate analog inhibitors exhibited K_i values of approximately $100 \, \text{nmol} \, \text{l}^{-1}$ as competitive inhibitors versus acetyl-CoA as expected for an ordered mechanism where acetyl-CoA binds first. The spacing between the aminoglycoside and CoA moieties was examined by increasing the methylene linker region to 2–4 carbons (compounds **1b–d**). The most potent inhibition was seen by compound **1b** ($K_i = 43 \, \text{nmol} \, \text{l}^{-1}$ versus acetyl-CoA). All of the bisubstrate analogs were also able to cocrystallize with AAC(6′)-Ii, which allowed for the determination of the three-dimensional structure of the enzyme with substrate analogs bound; previous experiments to do this had not been successful. Once again, though, the best *in vitro* inhibitor had no activity *in vivo*, presumably due to the negative charge from the phosphates of CoA.

   A second generation of truncated bisubstrate analogs based on compound **1** was prepared to probe the structure–activity relationships (**Figure 4**).[16] Loss of the 2-deoxystreptamine ring (2DOS) from the aminoglycoside moiety resulted in an inhibitor with a K_i value of $3.4 \, \mu\text{mol} \, \text{l}^{-1}$ versus acetyl-CoA (**4**). In fact, the second ring structure could be replaced with a number of positively charged compounds such as arginine and piperazine-based moieties without a loss of affinity for the enzyme. Removing the ADP moiety of the inhibitor, the site of the proposed cell-impermeable phosphates, ablated the inhibition (**5a**). Addition of a single phosphate at the end of the pantetheinyl arm resulted in full restoration of the inhibition (**5b**). In order to promote better cell permeability, the terminal phosphate was replaced by a malonyl group (**5c**) without significantly affecting the inhibition constant. A final substitution of a methoxy group at the terminal oxygen removed the negative charge from the inhibitor while retaining a K_i value of $11 \, \mu\text{mol} \, \text{l}^{-1}$ versus acetyl-CoA (**5d**). Although the inhibition constant for compound **5d** is more than 2 orders of magnitude greater than the original compound **1** *in vitro*, **5d** showed a synergistic inhibitory effect with kanamycin A and decreased the MIC for the antibiotic *in vivo*.[16]

Figure 4 Second-generation bisubstrate analog inhibitors of aminoglycoside acetyltransferases.

At the time the experiments with AAC(6′)-Ii were being performed, Blanchard and coworkers reported the characterization of an aminoglycoside 6′-N-acetyltransferase (AAC(6′)-Iy) isolated from *Salmonella enterica*. AAC(6′)-Iy is chromosomally encoded and shares only 14% sequence identity with AAC(6′)-Ii. AAC(6′)-Iy was shown to utilize a random kinetic mechanism as opposed to the ordered mechanism proposed for the -Ii enzyme.[17] In order to test the similarity in the two AAC(6′) enzymes, the first generation of inhibitors toward the -Ii enzyme were tested against the -Iy enzyme.[18] Surprisingly, the inhibitors were only micromolar inhibitors of the -Iy enzyme versus acetyl-CoA. In addition, instead of giving the predicted competitive inhibition Lineweaver–Burk plots, the bisubstrate analogs were shown to be noncompetitive versus acetyl-CoA and uncompetitive versus the aminoglycoside tobramycin. The inhibition constants versus acetyl-CoA were virtually independent of linker length or aminoglycoside side chain. However, there was a decrease in the K_i value versus the aminoglycoside as the linker increased in length. An explanation that is consistent with these observations is that the bisubstrate inhibitor can also bind to one of the enzyme–product complexes. In this case, it is known that the CoA product binds to the enzyme very tightly such that CoA release is rate limiting,[17] suggesting that the free enzyme may not be available to bind the bisubstrate inhibitor. Instead, the aminoglycoside portion binds to the Enz–CoA complex, explaining the loss of affinity when compared with the -Ii enzyme. This can also be seen in the linker length-dependent decrease in the K_i value versus the aminoglycoside as the additional length gives the aminoglycoside moiety more flexibility to bind as 'free aminoglycoside'.

8.19.3.2   Serotonin *N*-Acetyltransferase

Melatonin has been called the 'molecular pacemaker' hormone due to its involvement in such physical processes as aging and sleeping.[19] Considerable interest has been focused on the circadian rhythm of melatonin production. The penultimate step in melatonin biosynthesis is catalyzed by serotonin *N*-acetyltransferase (arylalkylamine *N*-acetyltransferase, AANAT). AANAT catalyzes the acetyl-CoA-dependent acetylation of serotonin to afford *N*-acetylserotonin (**Figure 5**). *N*-Acetylserotonin is then *O*-methylated by hydroxyindole *O*-methyltransferase to produce melatonin. An understanding of AANAT activity is of special importance due to the fact that the expression pattern of the gene displays a similar circadian rhythm that correlates with melatonin levels amounting to changes of up to 100-fold.[20] Thus, a potent inhibitor of AANAT activity could prove very useful in advancing the understanding of numerous sleep and mood disorders.
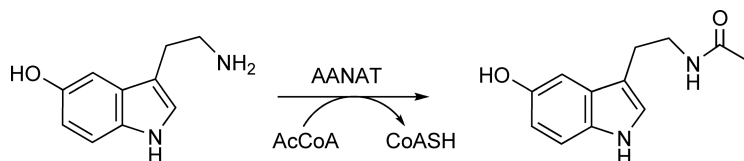
Figure 5 Reaction catalyzed by AANAT.

Compound	R	n
6a	H	0
6b	H	1
6c	H	3
6d	H	4

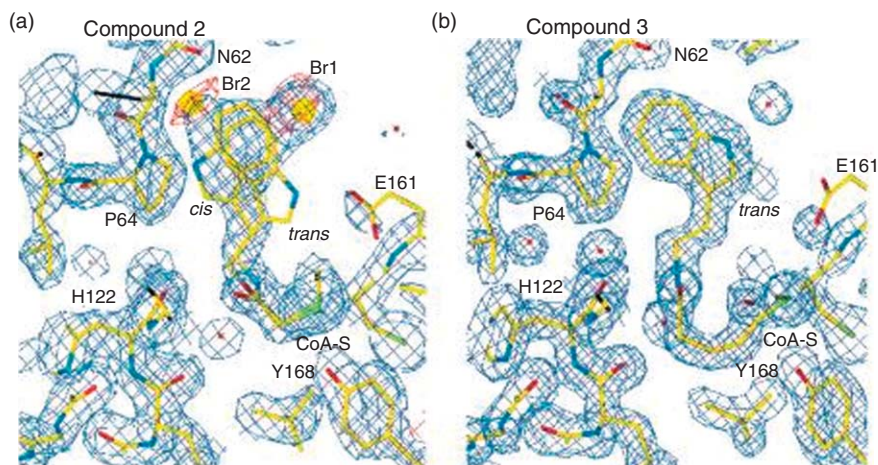| Compound	R
7a	H
7b	O=P−O / O⁻
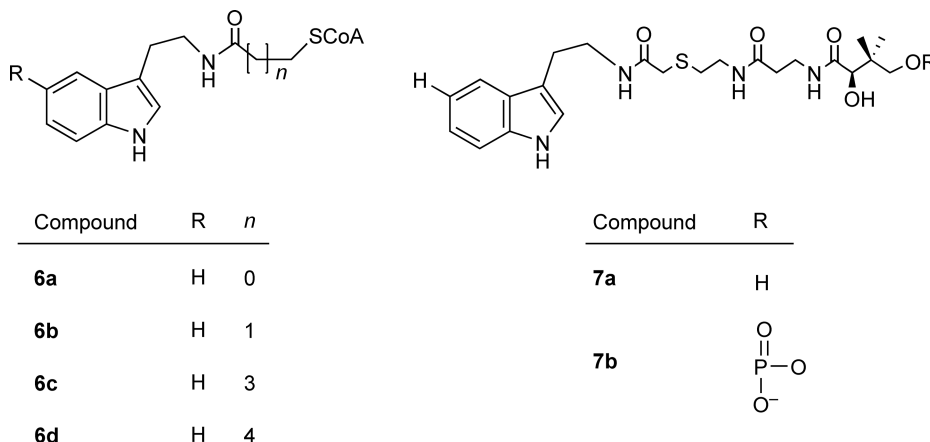
Figure 6 Chemical structures of bisubstrate analog inhibitors of AANAT.

The first reported specific and potent inhibitor of AANAT was a chemically synthesized bisubstrate analog, which linked N-acetyltryptamine with CoA (**6a** in **Figure 6**). As AANAT follows an ordered mechanism with acetyl-CoA binding followed by serotonin (or tryptamine) before chemistry occurs, it was proposed that this bisubstrate analog would be a good inhibitor of the enzyme.[21] Compound **6a** displayed an IC_{50} of 150 nmol l^{-1} against AANAT.[22] Competitive inhibition was observed versus acetyl-CoA ($K_i = 90$ nmol l^{-1}) and noncompetitive inhibition versus tryptamine as expected for an ordered mechanism. As has been discussed, it is difficult for charged CoA moieties to pass through the cell membrane and, therefore, this compound was not subjected to *in vivo* experiments. A series of bisubstrate analogs was chemically synthesized based on the original bisubstrate analog structure (**Figure 6**).[23] Shortening or lengthening the methylene linkers between the carbonyl and sulfur groups beyond the single group resulted in the loss of affinity for the enzyme (**6b–d**). Removal of the AMP or ADP moieties resulted in a much larger loss of inhibitory power (**7a** and **7b**).

In an attempt to avoid problems with the phosphate groups from the CoA moiety, the penultimate intermediate in the synthesis of the bisubstrate inhibitor, N-bromoacetyltryptamine, was tested as a possible 'affinity label' inhibitor or as a substrate for the *in situ* enzymatic synthesis of the bisubstrate inhibitor through an acylation mechanism.[23] N-Bromoacetyltryptamine did act as an inhibitor of the enzyme, but inhibition could be reversed by dialysis suggesting that the inhibition was not due to a covalent adduct. It was shown that the enzyme catalyzed the acylation of N-bromoacetyltryptamine by CoASH to form compound **6a** with a rate enhancement of 3.3×10^4 relative to the uncatalyzed reaction. Ultimately, it was shown that the acylation reaction occurs at the same active site as the acetylation activity. A closer inspection of the kinetics of inhibition by the bisubstrate analog **6a** resulted in the observation of slow-onset inhibition over the first few minutes of the reaction with a K_i^* value of 84 nmol l^{-1}. Owing to its neutrality, N-bromoacetyltryptamine was tested as an inhibitor *in vivo*. The analog precursor was shown to inhibit melatonin production in norepinephrine-stimulated pinealocytes in a concentration-dependent manner and with low cytotoxicity.

The use of bisubstrate analog inhibitors in structural experiments of AANAT provided a very powerful tool to probe the subtle differences in the active site leading to either acetyltransferase or acyltransferase activity. Compound **6a** was shown to have K_i values of 0.048 and 33 µmol l^{-1} for the acetyltransferase and acyltransferase activities, respectively.[23] When this compound is cocrystallized with the enzyme, two forms of the inhibitor

Figure 7 Electron density of AANAT cocrystallized with compound (a) **6a** and (b) **6d**. Reprinted from E. Wolf; J. De Angelis; E. M. Khalil; P. A. Cole; S. K. Burley, *J. Mol. Biol.* **2002**, *317*, 215–224, with permission from Elsevier.

are seen with the tryptamine moiety bound in *cis* and *trans* configurations (**Figure 7(a)**).[24] When compound **6d**, which only inhibits the acetyltransferase activity ($K_i = 6.5 \, \mu\text{mol} \, l^{-1}$ and $>1000 \, \mu\text{mol} \, l^{-1}$ for acetyltransferase and acyltransferase, respectively) is cocrystallized, the tryptamine moiety is only found in the *trans* configuration (**Figure 7(b)**). Although, it is tempting to conclude that this difference in binding accounts for the two activities catalyzed by AANAT the authors suggested that more evidence is required before that conclusion can be drawn. However, the use of alternate bisubstrate inhibitors in this case has certainly provided structural and kinetic evidence supporting the plasticity of the active site environment of AANAT.

8.19.3.3 Histone *N*-Acetyltransferase

The posttranslational modifications of histones have become a paradigm for the mechanisms of epigenetics and gene regulation. These modifications include acetylation, methylation, and phosphorylation at specific residues on histone tails.[25] Acetylation of the ε-amine of specific lysine residues is thought to turn on gene expression by removing the positive charge on the histone tail causing a decrease in affinity for the negatively charged DNA held by the histone proteins. Histone acetylation is catalyzed in an acetyl-CoA-dependent manner by enzymes termed histone acetyltransferases (HATs).[7] Due to their role in the regulation of gene expression, HATs have been intensely studied over the past decade and specific inhibitors would be useful as possible therapeutics for a number of different diseases and cancers.

The enzyme p300/CBP-associated factor (PCAF) belongs to the GNAT family of enzymes and catalyzes the acetylation of Lys14 on histone 3. In 2000, the Cole group reported on a specific bisubstrate analog that is a potent inhibitor of PCAF.[26,27] A peptide corresponding to the first 20 amino acids of H3 was synthesized with *N*-acetyllysine at the 14 position (**Figure 8**). The acetylated lysine residue was then chemically CoA-ylated to form the inhibitor. The inhibitor, H3-CoA-20, was a potent inhibitor of PCAF activity *in vitro* with an IC_{50} of


Compound	R
H3-CoA-20	H
H3-(Me)CoA-20	CH_3
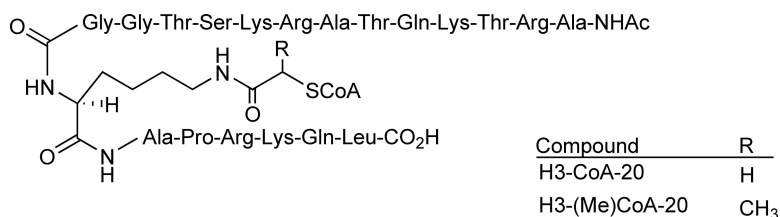
Figure 8 Chemical structure of PCAF bisubstrate analog inhibitor H3-CoA-20.

300 nmol l^{-1}. H3-CoA-20 also displayed nearly 700-fold selectivity toward PCAF inhibition relative to p300, a member of another HAT subfamily. Alternate bisubstrate analogs based on shorter peptides or peptides from histone 3 did not result in inhibition, suggesting that the binding determinants for the peptide substrate are found over the entire substrate and not localized to the target lysine. More in-depth kinetic studies of H3-CoA-20 inhibition demonstrated that the inhibitor acted competitively against acetyl-CoA ($K_i = 28$ nmol l^{-1}) and noncompetitively versus the peptide substrate, consistent with initial velocity studies suggesting an ordered sequential mechanism with acetyl-CoA binding first followed by peptide substrate.[27,28]

A three-dimensional structure of a PCAF homolog (Gcn5 from *Tetrahymena*) bound to a derivative of H3-CoA-20, H3-(Me)CoA-20 (**Figure 8**), where an isopropionyl-linker group was used instead of the original acetyl linker, was solved.[29] In the structure, only seven residues flanking the modified lysine residues were visible. However, bisubstrate analog inhibitors based on smaller peptides display much poorer inhibition of PCAF. When this structure was compared to the structure of the ternary complex (H3 peptide and CoA) the CoA moiety of the inhibitor overlays nicely with the CoA molecule but the peptide residues are in a very different conformation. A model of substrate binding and recognition can be proposed where the initial substrate binding is dependent upon interactions with the full-length H3 peptide, consistent with the results of inhibition kinetics and the complete ordering of the peptide in the ternary complex structure. After the initial binding event, the catalytic event relies mainly on interactions at a much closer range to the modified lysine as seen in the seven-residue span of the inhibitor-bound structure, which would be a much closer approximation to the transition state of the reaction.

As mentioned above, the H3-CoA-20 inhibitor was specific for PCAF relative to p300. However, the simple bisubstrate analog of Lys-CoA (**8a** in **Figure 9**) was able to specifically inhibit p300 400-fold more potently than PCAF.[26] This selectivity was originally proposed to be due to the ping-pong mechanism of PCAF.[30] However, more recent biochemical studies have ruled out the possibility of a ping-pong mechanism and instead favor a Theorell–Chance mechanism.[31,32] In a Theorell–Chance mechanism (**Scheme 2**), there is no stable ternary complex.[33] Instead, the first substrate (acetyl-CoA) binds tightly, followed by a 'hit-and-run' by the second substrate (peptide) resulting in catalysis. Since the peptide is not tightly bound by the enzyme, the
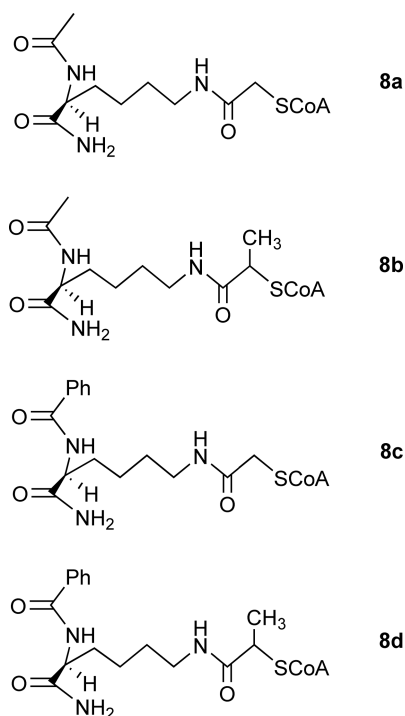
Figure 9   Chemical structures of p300 bisubstrate analog inhibitors.

$$A + Enzyme \rightleftharpoons A-Enzyme \longrightarrow Q-Enzyme \rightleftharpoons Q + Enzyme$$

$$B \qquad P$$
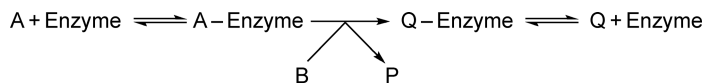
Scheme 2

addition of more peptide sequence to the bisubstrate analog would not be expected to increase potency, consistent with the inhibition results described above.
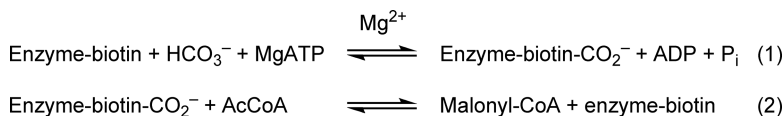
Kinetically, Lys-CoA acts as a slow-onset inhibitor with a K_i^* value of 18.7 nmol l^{-1}.[30] Studies have shown that the entire CoA moiety was required for efficient inhibition of p300,[34] but that CoA by itself was not sufficient.[26] Substitution of an isopropyl linking group (**8b**) was able to increase potency suggesting that limited flexibility is important for binding.[35] The addition of a phenyl group to the carboxyl group of lysine (**8c**) also resulted in increased potency most probably due to accessing addition binding determinants in a nearby hydrophobic pocket. Surprisingly, the combination of the isopropyl linker and the phenyl group (**8d**) resulted in a less potent inhibitor suggesting that the structure–activity relationship may not be straightforward.
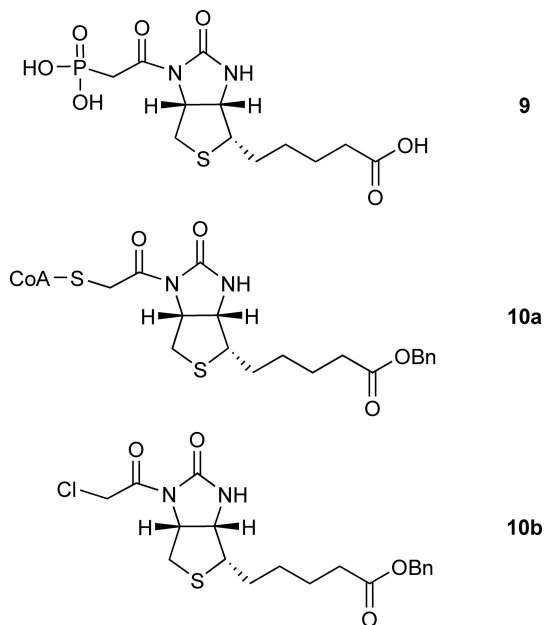
8.19.4 Acetyl-CoA Carboxylase

Acetyl-CoA carboxylase (ACC) catalyzes the first committed step in long-chain fatty acid biosynthesis (see Chapter 7.11).[36] The overall reaction is catalyzed in two sequential reactions (**Scheme 3**). First, the biotin carboxylase domain catalyzes the ATP-dependent carboxylation of biotin (which is attached to a carrier protein) using bicarbonate as a CO_2 donor. In the second reaction, the carboxyl group is transferred from biotin to acetyl-CoA to form malonyl-CoA. In mammals, both reactions are catalyzed by a single protein,[37] but in *Escherichia coli* and other bacteria, the activity is catalyzed by two separate proteins, a biotin carboxylase and a carboxytransferase.[38] Due to its role in fatty acid synthesis, inhibitors of the overall ACC reaction are proposed to be useful as antiobesity drugs in mammals as well as novel antibiotics against bacteria.

The Waldrop group, working with the individual biotin carboxylase and carboxytransferase enzymes from *E. coli*, has developed bisubstrate analogs for both reactions. In the case of biotin carboxylase, the bisubstrate analog was chemically synthesized by linking phosphonoacetate (an analog of the proposed carboxyphosphate intermediate, technically making this compound a reaction intermediate analog) with the 1′-nitrogen of biotin (compound **9** from **Figure 10**).[39] Compound **9** is a competitive inhibitor versus ATP with a K_i of 8.4 mmol l^{-1}, a significant improvement over the Michaelis constant for biotin (134 mmol l^{-1}), making it the first reported biotin-derived inhibitor.[40] The bisubstrate analog for the carboxyltransferase enzyme was synthesized in a similar manner except that CoASH was substituted for phosphonoacetate to form compound **10a**. When the carboxyltransferase is assayed in the reverse direction, the analog is a competitive inhibitor versus malonyl-CoA ($K_i = 23\ \mu$mol l^{-1}) and a noncompetitive inhibitor versus a biotin-like substrate.[41] These results are consistent with an ordered addition of malonyl-CoA followed by biotin prior to chemistry occurring.

By confirming the inhibition of bacterial carboxyltransferase enzyme, the next step was to investigate the use of the compound *in vivo* against the mammalian ACC enzyme. As with the CoA moieties in the Gcn5 acetyltransferase inhibitors, the CoA moiety of the carboxyltransferase inhibitor would not allow efficient cellular uptake. In a strategy similar to that used for serotonin *N*-acetyltransferase, the chloroacetylated biotin precursor (**10b**) was used as an inhibitor that could be CoA-ylated *in vivo* to form **10a**. Treatment of confluent 3T3-L1 preadipocytes for 4 h with 10 μmol l^{-1} compound **10b** resulted in a 79% decrease in ACC activity.[42] More importantly, treatment with **10b** blocked the induction of adipogenic transcription factors in a

$$Enzyme\text{-}biotin + HCO_3^- + MgATP \overset{Mg^{2+}}{\rightleftharpoons} Enzyme\text{-}biotin\text{-}CO_2^- + ADP + P_i \quad (1)$$

$$Enzyme\text{-}biotin\text{-}CO_2^- + AcCoA \rightleftharpoons Malonyl\text{-}CoA + enzyme\text{-}biotin \quad (2)$$

Scheme 3

Figure 10   Chemical structures of acetyl-CoA carboxylase bisubstrate analog inhibitors.
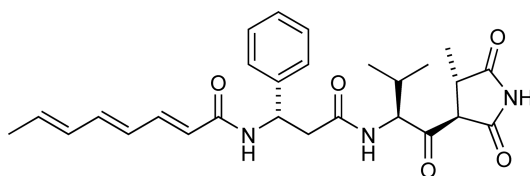

Figure 11   Chemical structure of andrimid.

dose-dependent manner and inhibited lipid accumulation in differentiating 3T3-L1 cells suggesting that inhibition of cytosolic ACC represents a viable target for antiobesity therapeutics.

The natural product andrimid, a peptide-based antibiotic, has been shown to inhibit carboxyltransferase activity (**Figure 11**).[43] Kinetic studies show that andrimid acts as a competitive inhibitor versus malonyl-CoA and a noncompetitive inhibitor versus a biotin analog. As the carboxyltransferase enzyme from *E. coli* has been shown to follow a rapid-equilibrium ordered mechanism with malonyl-CoA binding first followed by biotin,[44] the inhibition results cannot discriminate between andrimid acting as a simple dead-end inhibitor versus malonyl-CoA or as a bisubstrate analog. Additionally, while the structure of the *E. coli* enzyme has been solved,[45] it has not been demonstrated that andrimid interacts with the binding site of both substrates. Thus, while andrimid has some structural similarities to the substrates of the carboxyltransferase reaction, its mode of inhibition remains ambiguous.

8.19.5   Protein Kinases

Protein kinases play an integral role in the mechanism of signal transduction and therefore are intimately connected with a number of disease states (see Chapters 5.21 and 6.15). However, effective therapeutics against protein kinases have been difficult to design due to significant overlap of substrate specificity. Protein kinases utilize ATP as the source of the transferred phosphate making selectivity very difficult with ATP-analog inhibitors. Numerous researchers have attempted to overcome the problems associated with protein kinase inhibitor selectivity by turning to bisubstrate analog inhibitors.[46]

(Arg)$_6$-Ser-CO-(CH$_2$)$_2$-NH-(CH$_2$)$_2$-NH-SO$_2$ 11

NH$_2$-Leu-Arg-Arg-Ala-Gly-Leu-Gly-COOH 12
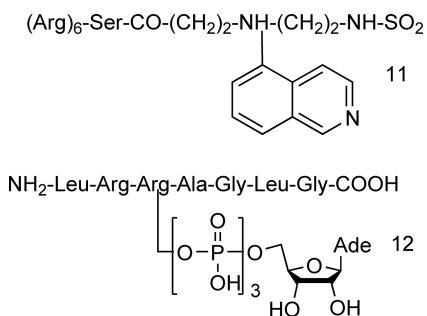
Figure 12 Initial bisubstrate analog inhibitors for protein kinases.

Initial attempts to develop bisubstrate analog inhibitors for protein kinases began in the early 1990s (**Figure 12**). It was found that coupling of isoquinolinesulfonic acid to a hexaarginine peptide (**11**) resulted in bisubstrate inhibition kinetics of protein kinase C (PKC) and cAMP-dependent kinase (PKA) with IC$_{50}$ values in the low nanomolar region.[47] A more traditional bisubstrate analog linking ATP with a peptide substrate through the phosphorylation site (**12**) proved less potent against PKA with a micromolar IC$_{50}$ value.[48]

In 2001, the Cole group reported the design of a potent bisubstrate analog inhibitor (**Figure 13**) for insulin receptor kinase, a tyrosine kinase, which had a K_i value of 370 nmol l^{-1}.[49] The design of this bisubstrate analog is a modification to compound **12** and is based on an understanding of the mechanism of phosphoryl transfer of the tyrosine kinase family. These enzymes are proposed to utilize a dissociative S$_N$1-type mechanism of phosphoryl transfer consistent with significant bond cleavage between the phosphoryl donor and the transferred phosphate

R^1 = AcNH-Lys-Lys-Lys-Leu-Pro-Ala-Thr-Gly-Asp-
R^2 = -Met-Asn-Met-Ser-Pro-Val-Gly-Asp-CO$_2$H
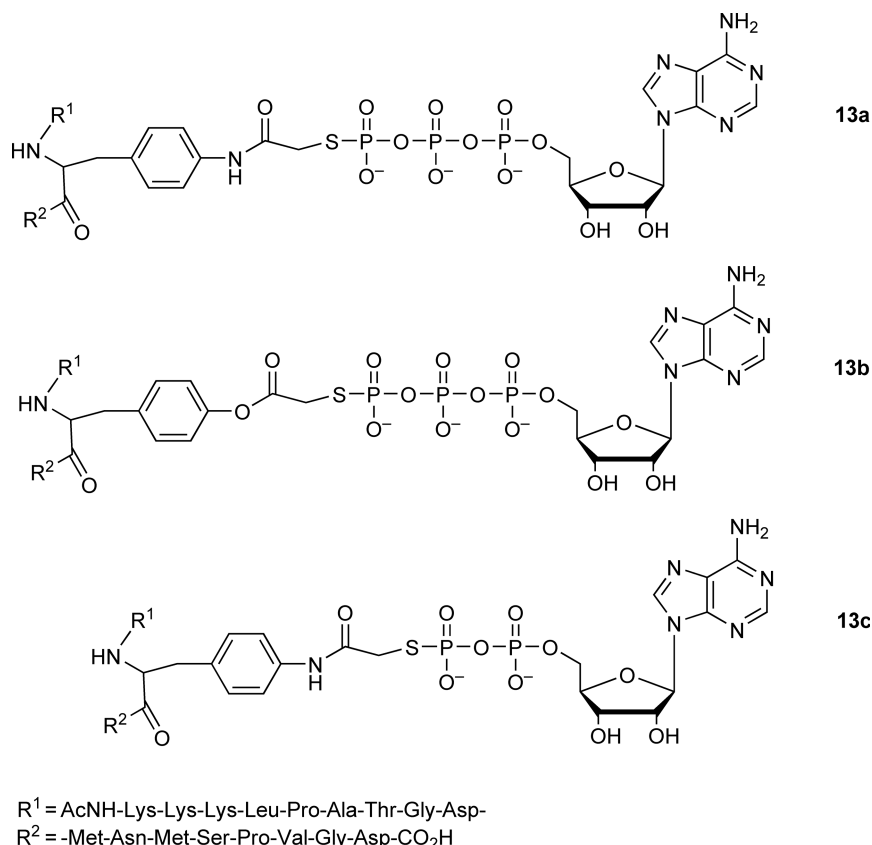
Figure 13 Chemical structures of mechanism-based designed bisubstrate analog inhibitors of protein kinase.
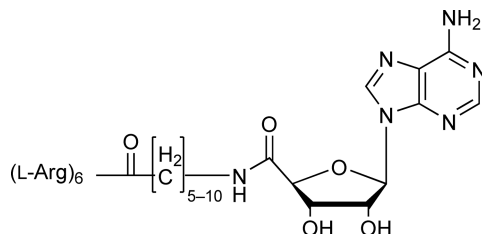
**Figure 14** Chemical structure of adenosine-5′-carboxylic acid-based bisubstrate analogs of protein kinase.

prior to nucleophilic attack by the acceptor peptide.[50,51] In order to mimic this extra space (~ 5 Å), an acetyl group was used as a linker between ATPγS and the peptide substrate (**13a**). In addition, it was known from structural studies that the phenolic hydroxyl of the target tyrosine is in a hydrogen bond with a conserved aspartic acid residue in the active site. Thus, 4-amino-phenylalanine was used in place of the target tyrosine to preserve that active site interaction. The inhibitor was competitive versus both substrates consistent with a random order of addition in the kinetic mechanism. A three-dimensional structure was determined with the inhibitor bound in a bisubstrate fashion. Compound **13a** was also tested against another member of the tyrosine kinase family and was found to be a much less potent inhibitor ($K_i = 40\,\mu\mathrm{mol\,l^{-1}}$), suggesting some selectivity as well. To test the utility of the ATPγS-acetyl linker group as a scaffold for selective protein kinase inhibitor design, a similar study was undertaken by replacing the peptide sequence in compound **13a** for a serine/threonine kinases-specific sequence.[52] The results of this study were less impressive than the insulin receptor kinase studies ($K_i = 4\,\mu\mathrm{mol\,l^{-1}}$), but the bisubstrate analog provided a 30-fold increase in the affinity for PKA relative to compound **12**.

Building on the successful design of compound **13a**, a series of modifications were made to the inhibitor to probe the active site of insulin receptor kinase.[53] Changing the aminophenylalanine back to a tyrosine residue (**13b**) resulted in an 80-fold loss of potency consistent with the loss of the proposed hydrogen bonding interaction. Removal of a phosphate (resulting in ADPγS, **13c**) shortened the distance that the acetyl group had added without changing the chemical aspect of linker group. This removal caused at least a 500-fold decrease in affinity reinforcing the hypothesis that the distance between the donor and nucleophile is paramount to the efficacy of the inhibitor. Surprisingly, changes to the peptide sequence portion of the bisubstrate analog based on their functionality as substrates did not coincide with their utility as inhibitors. The authors suggest using the peptide sequence to produce specificity may not be as straightforward as thought previously.

The use of the triphosphate moiety, however, suggests that while this family of compounds will be useful *in vitro*, *in vivo* efficacy will be limited. A more bioavailable series of bisubstrate inhibitors has been synthesized based on adenosine-5′-carboxylic acid peptidyl derivatives (**Figure 14**). These inhibitors vary from those described above by using only the adenosine portion of ATP, linking with a simple aliphatic moiety, and using a hexa-L-arginine peptide. Similar to results from the Cole lab, researchers found that these analogs had an optimal linker length for inhibition. Inhibitors with linker lengths of 5–10 methylene groups resulted in IC_{50} values of 0.12–0.33 $\mu\mathrm{mol\,l^{-1}}$ versus protein kinases A and C, but >30 $\mu\mathrm{mol\,l^{-1}}$ against protein kinases CK1 and CK2.[54] These inhibitors were also able to effectively cross cellular membranes.[55] A second generation of adenosine-5′-carboxylic acid-based inhibitors incorporated a protease-resistant hexa-peptide of D-arginine that resulted in an IC_{50} value of 8 $\mathrm{nmol\,l^{-1}}$ versus cAMP-dependent protein kinase.[56]

8.19.6 Catechol-*O*-Methyltransferase

Low levels of the neurotransmitter dopamine in the brain have been linked to Parkinson's disease.[57] Oral administration of L-dihydroxyphenylalanine (L-DOPA), the biosynthetic precursor to dopamine, has been successful in treating the symptoms of Parkinson's disease. The key to this success is that L-DOPA is capable of crossing the blood–brain barrier while dopamine is not (**Scheme 4**). Once in the brain, L-DOPA is decarboxylated by the enzyme aromatic amino acid decarboxylase (AAD) to give dopamine. Prior to entering the brain L-DOPA is susceptible to metabolism by several different enzymes. The initial entry into metabolism is also
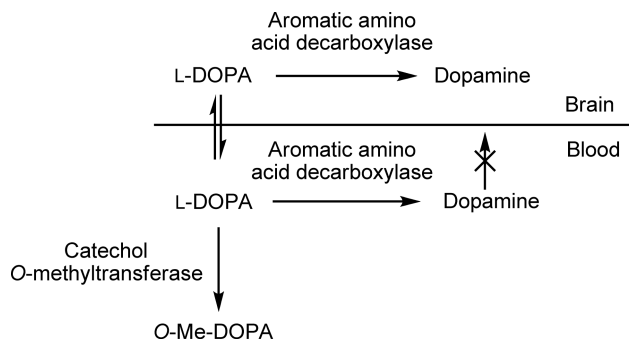
Scheme 4

catalyzed by AAD, resulting in brain-impermeable dopamine. Most Parkinson's disease therapies include an AAD inhibitor to decrease this side reaction. When AAD is inhibited, a second metabolic pathway catalyzed by catechol-*O*-methyltransferase (COMT) acts on L-DOPA. COMT catalyzes the transfer of a methyl group from *S*-adenosylmethionine to the phenolic oxygen group on catechols, again reducing the amount of L-DOPA that can be transported to the brain. Nitro-substituted catechols that act as inhibitors of COMT (such as tolcapone[58] and entacapone[59]) have been included as part of the drug cocktail for the treatment of Parkinson's disease.

Building on this, Diederich and coworkers proposed to link the nitro-substituted catechols to adenosine (to mimic *S*-adenosylmethionine) with hopes of developing more potent bisubstrate inhibitors (**Figure 15**).[60,61] Compound **14**, where a nitro-catechol moiety was linked to adenosine through an aliphatic chain, proved to be
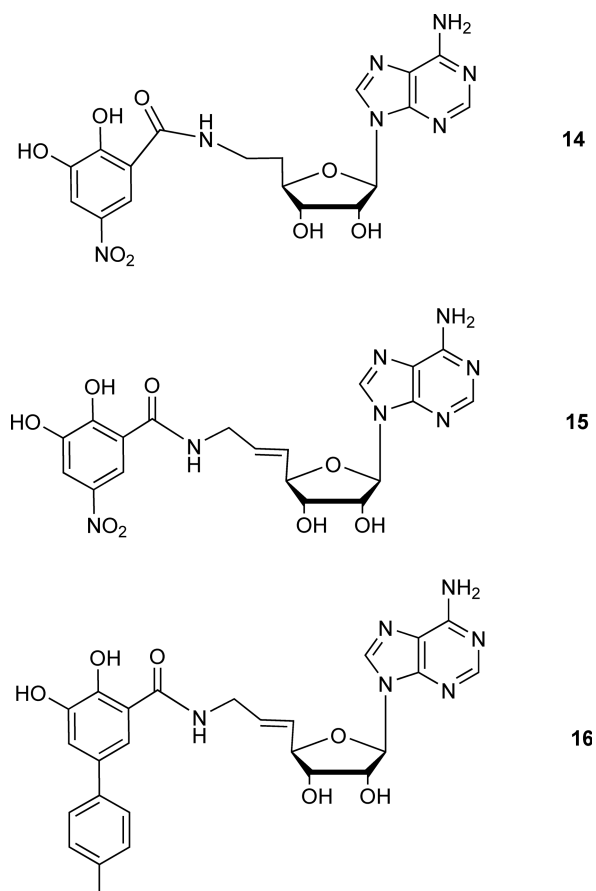
Figure 15 Chemical structure of bisubstrate analog inhibitors of catechol *O*-methyltransferase.

a moderate inhibitor of COMT. Variations on the length of linker domain of **14** showed an optimum linker length of two methylene groups ($IC_{50} = 60$ nmol l^{-1}) with order of magnitude decreases in potency for each additional methylene linker (up to four total). However, the crystal structure had shown that at least three or four methylene spacers were required for the bisubstrate analog to reach both binding sites. The authors proposed that the flexibility of the longer linker domains was the source of the lost potency. Therefore, a hindered *trans* double bond was inserted into a three methylene linkers to create compound **15**. Compound **15** was a potent inhibitor of COMT activity *in vitro* with an IC_{50} value of 9 nmol l^{-1}. Compound **15** was shown to be a competitive inhibitor versus both *S*-adenosylmethionine and benzene-1,2-diol, consistent with bisubstrate inhibitor kinetics. In addition, a cocrystal of COMT with **15** confirmed that it did indeed bind as a bisubstrate analog. Further studies showed that a 3′-deoxy ribose ring could be substituted without diminishing the potency of the inhibitor (low nanomolar IC_{50} value), but the 2′-deoxy substituent shows a dramatic decrease in affinity for the enzyme (micromolar IC_{50} value).[62]

An extremely important aspect of bisubstrate inhibitors is that they usually utilize multiple binding points to achieve their high affinities for enzymes. The current catechol analog COMT inhibitors rely on the nitro-substituent to provide potency. When uncertainty about the cause of occurrences of hepatoxicity in Parkinson's disease patients on the cocktail therapy was directed at the nitro-substituent,[63] the question quickly arose as to whether the nitro-substituent was required for bisubstrate analog potency. It was shown that utilizing a 4-methyl-phenyl group in place of the nitro-substituent (**16**) resulted in an inhibitor with only twofold less potency (IC_{50} value of 23 versus 9 nmol l^{-1}).[64] When the new inhibitor was modeled into the COMT active site, the phenyl substituent was shown to stack with a tryptophan and a proline residue suggesting that further potency could be attained by probing pockets in the active site.

8.19.7   Farnesyltransferase

Posttranslational modifications are a key step in the regulation of many proteins. Farnesyltransferase (FT) catalyzes the post-translational transfer of a farnesyl group from farnesyl pyrophosphate to a cysteine residue near the carboxy terminus of its target protein (see Chapter 5.08).[65] The consensus target sequence is CaaX, where 'a' represents an aliphatic residue and X is either serine or methionine. Interest in FT inhibitors rapidly progressed once it was reported that the first required step in activating the protein product of the oncogene *ras* was farnesylation catalyzed by FT. The farnesylation of the protein Ras ultimately leads to membrane association and activation of the protein.[66,67] Mutations in the *ras* gene have been found in 30% of all human cancers suggesting that inhibitors of its action would be broad-spectrum anticancer agents.[68]

In 1993, Takeshima and coworkers reported the isolation of a series of fermentation products from *Streptomyces* strain OH-4652 that displayed FT inhibition with high nanomolar IC_{50} values.[69] The structure of these compounds, termed pepticinnamins, were shown to be composed of a substituted cinnamaldehyde group linked to a peptide-like structure (**Figure 16**).[70] Subsequently, the total synthesis of pepticinnamin E was
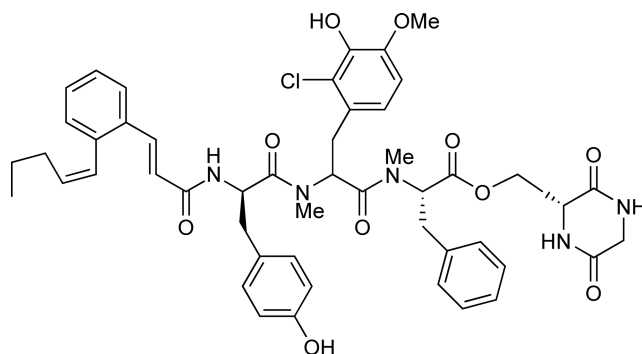

**Figure 16**   Chemical structure of the natural product pepticinnamin E.

reported.[71] Kinetically, pepticinnamin E was a competitive inhibitor versus both the CaaX peptide ($K_i = 30\ \mu mol\ l^{-1}$) and farensyl pyrophosphate ($K_i = 8\ \mu mol\ l^{-1}$), indicating that the natural product was acting as a bisubstrate inhibitor.[71]

In an attempt to improve the inhibition kinetics, a pepticinnamin E library was synthesized to explore the chemical space of the natural product.[72] Three types of modifications were investigated: changes to the N- and C-terminal substituents on the backbone, changes in N-methylation status, and the addition of an N-terminal histidine residue to form an interaction with the active site zinc atom. The inclusion of a histidine residue at the N-terminal peptide position proved favorable and modulation of the N-terminal substituent (i.e., the farnesyl pyrophosphate mimic) proved important for inhibition. Several of the compounds from the library were shown to induce apoptosis in MDCK-f3 tumor cells at concentrations of $100\ \mu mol\ l^{-1}$; however, in the case of only one compound could apoptosis be linked to FT inhibition.

The majority of inhibitors developed for FT are based on mimics of the CaaX consensus sequence. In these inhibitors, the free thiol of the cysteine group is required for inhibitory activity causing an obvious problem for *in vivo* studies. In an attempt to work around this problem, researchers at Bristol-Meyers Squibb proposed to replace the cysteine group with a phenolic substituent to create a sulfhydryl-free inhibitor.[73] Compound **17** (Figure 17), with a meta-hydroxyl phenolic group attached to a -VLS peptide was an inhibitor of FT with an IC_{50} value of $29\ \mu mol\ l^{-1}$. With the precedent that the cysteine residue could be substituted with retention of inhibitory potency, the next step was taken to design a bisubstrate analog by linking a farnesyl group to
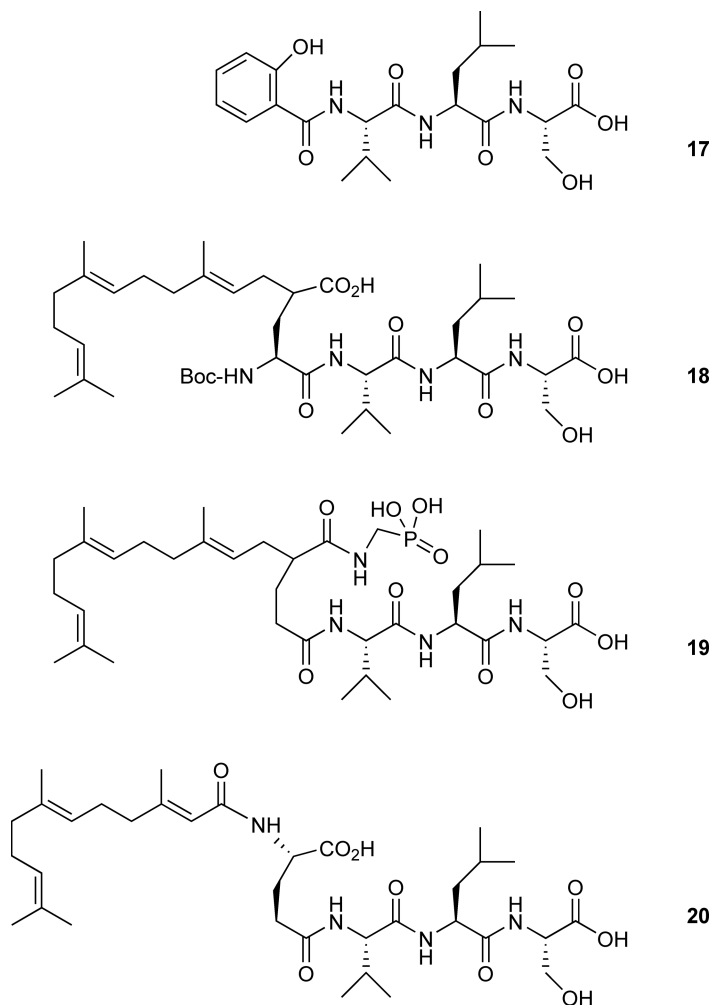
Figure 17 Chemical structure of first-generation bisubstrate analog inhibitors of farnesyl transferase.

Compound	R^1	R^2	R^3
21a	CH_2CHMe_2	CH_2OH	OH
21b	$CHMe_2$	$CHMe_2$	OH
21c	$CHMe_2$	$CHMe_2$	OMe
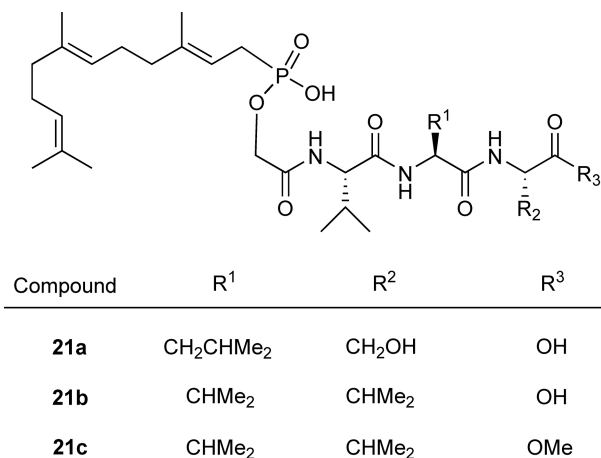
Figure 18   Chemical structures of phosphonic acid-based bisubstrate analog inhibitors of farnesyl transferase.

the -VLS peptide with a carboxylic acid in the linker group to mimic the sulfhydryl group (**Figure 17**).[74] Compound **18** had an IC$_{50}$ value of $20 \, \mu mol \, l^{-1}$, similar to the phenol-based compounds. Removal of the N-terminal amide group resulted in a 20-fold decrease in the IC$_{50}$ value ($0.9 \, \mu mol \, l^{-1}$). The important binding determinant appeared to be the carboxylic acid in the linker group, since the farnesyl group could be replaced by a dodecyl group with only a 10-fold decrease in potency. However, substitution of either a hydrogen or ethyl ester moiety for the carboxylic acid compound abolished inhibition. In an attempt to add the binding affinity of the pyrophosphate group in addition to the farnesyl and peptide groups, the carboxylic acid was extended with an amidomethylphosphonic acid moiety to create compound **19** with an IC$_{50}$ value of $150 \, nmol \, l^{-1}$. Another variation was the addition of a hydrogen bonding amide group to mimic the proposed hydrogen bonding interaction with the oxygen from farnesyl pyrophosphate (**20**). The L-isomer of compound **20** demonstrated an IC$_{50}$ value of $33 \, nmol \, l^{-1}$, 30-fold more potent than the original bisubstrate analog.

   Following the precedents of metalloprotease inhibitors that a phosphonic or phosphinic acid moiety was a preferred sulfhydryl mimic, a second generation of FT bisubstrate analogs was designed (**Figure 18**).[75] The incorporation of a phosphonic acid group in place of the carboxyl group on the farnesyl-X-VLS inhibitor (**21a**) resulted in an inhibitor with and IC$_{50}$ value of $60 \, nmol \, l^{-1}$, comparable to the value for the best carboxylic acid-based compound. An alteration in the peptide sequence to VVM (**21b**) increased the potency by 10-fold. A phosphinic acid version of the same inhibitor did not enhance potency. While the inhibitors as synthesized were not able to produce *in vivo* results due to the charged C-terminus of the peptide region, methyl ester prodrugs (**21c**) of these compounds were found to be viable *in vivo* inhibitors of *ras*-induced cell transformation in NIH3T3 cells.[75,76] Further studies attempted to incorporate a hydroxymate linker region.[77] While this simplified the synthesis of inhibitors, its incorporation did not improve upon the inhibitory properties of the phosphonate-linked compounds. Attempts have been made to create entirely synthetic bisubstrate analogs to avoid possible problems due to oxidation of the isoprenyl groups or hydrolysis of the peptide groups.[78,79] These compounds show *in vivo* antiproliferative activity, but have very high nanomolar IC$_{50}$ values.[80]

8.19.8   Current Bisubstrate Analogs as Therapeutics

While many bisubstrate analog inhibitors have been reported, few are applicable as therapeutics as discussed above. In cases where these hurdles are overcome, the therapeutic impact of such compounds has been nothing short of outstanding. The sections below contain descriptions of one of the most tightly bound noncovalent inhibitors ever reported, the most widely prescribed topical antibiotic, and the most effective medication thus far versus the human disease tuberculosis. It is of note that only the natural product mupirocin does not act as a prodrug suggesting that while we should continue to focus on developing bisubstrate analog inhibitors those that can be administered as prodrugs stand the best chance of being therapeutically useful.

8.19.8.1 Finasteride

The development and characterization of the 3-oxo-4-azasteroid finasteride is, in the authors' opinion, an elegant example of the full drug design process from target discovery to determination of mechanism of action. Finasteride, marketed by Merck as Propecia and Proscar, is an effective therapeutic for benign prostatic hyperplasia (BPH or enlarged prostate) and male-pattern baldness. The discovery that these two maladies could be treated by a single drug stems from a genetic study of a small population of male pseudohermaphrodites in the Dominican Republic.[81] The affected population is born with ambiguous external genitalia but is genetically male. At puberty, the fully functional external male genitalia develop, and the individuals go on to lead normal lives. A loss of function of the enzyme Δ^4-steroid 5α-reductase (5AR) was linked to the autosomal-recessive condition suggesting a role for the enzyme in sexual differentiation. In addition, the affected population does not experience male-pattern baldness or prostate enlargement, implicating 5AR as mediating these more common problems as well. In the early 1990s, the role of 5AR in pseudohermaphrodism was further clarified by the discovery that there are two isozymes of the enzyme.[82] Type 1 5AR was shown to have a slightly basic pH optimum and is fully present and active in pseudohermaphrodites. Type 2 5AR, which was found to be deleted in the DNA of pseudohermaphrodites, has an acidic pH optimum and is proposed to be the primary isozyme in genital tissue. In an attempt to develop drugs aimed at treating BPH and male-pattern baldness, Merck began testing potential inhibitors of 5AR.

The enzyme 5AR catalyzes the NADPH-dependent reduction of the hormone testosterone to dihydrotestosterone (**Figure 19**). Dihydrotestosterone is described as a more potent androgen than testosterone due to its stronger affinity for the androgen receptor. However, both testosterone and dihydrotestosterone are translocated to the nucleus using the same androgen receptor. This suggests that drug candidates need to be specific for 5AR since simple testosterone mimics might also inhibit the androgen receptor interfering with testosterone translocation. 4-Methyl-4-azasteroid was initially characterized as an NADPH-dependent, low nanomolar inhibitor of 5AR from rat prostate and liver and human foreskin fibroblasts, but at high doses it also affects the androgen receptor.[83] 4-Aza-steroid was subsequently shown to be a potent enzyme inhibitor, but the *des*-methyl compound had little affinity for the receptor.[84] Further studies aimed at defining the structure–activity relationship would describe the kinetics and specificity of the 4-aza-steroid-based inhibitor 17β-*N*-*t*-butylcarbamoyl-4-aza-5α-androst-1-en-3-one (**Figure 20**), soon to be labeled finasteride. Finasteride was initially characterized as a reversible competitive inhibitor versus testosterone with a $K_i = 6\,\mathrm{nmol\,l^{-1}}$ for the rat prostatic enzyme.[85] Finasteride was also shown to have a >20 000-fold preference for 5AR versus the androgen receptor.

Clinical trials for finasteride began in 1986 and resulted in almost unbelievable findings.[86] It was found that a single oral dose of 0.5 mg resulted in a 65% decrease in plasma DHT for 5–7 days with no significant changes in serum lipids. Over the course of a 6-month treatment, patients with BPH had an overall decrease in prostate
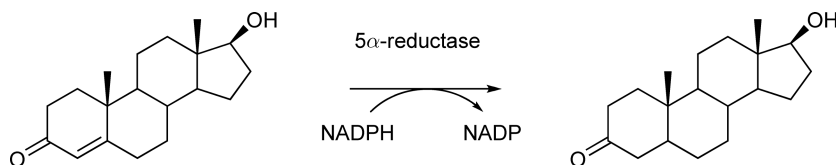

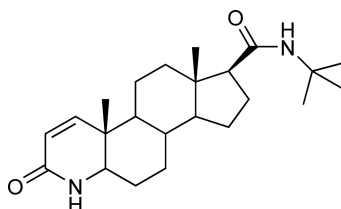**Figure 19** Reaction catalyzed by 5α-reductase.

Figure 20 Chemical structure of finasteride.

size of 28%. While the clinical results were impressive, they were almost highly impressive in light of the previous kinetic studies. Studies on 5AR from human prostate had established that the enzyme follows an ordered kinetic mechanism with NADPH binding first, followed by testosterone.[87] The Michaelis constant for testosterone was determined to be 34 nmol l^{-1}. This value is only a few fold higher than the determined K_i for finasteride (6 nmol l^{-1}) acting as a competitive inhibitor. The data from the clinical trial suggested that a small dose of finasteride can completely inhibit the enzyme, whereas the kinetic data suggested that *in vivo* concentrations of testosterone will effectively compete against the inhibitor.

In 1993, a report, which helped to understand that conundrum of the K_i/K_m ratio to the high efficacy seen *in vivo*, was published. Researchers at Ciba-Geigy in Switzerland demonstrated that finasteride is a slow-onset inhibitor of type 2 5AR.[88] However, the rate of inhibition is very slow such that steady-state rates are not reached until 30 min in the presence of 10 nmol l^{-1} finasteride. For finasteride concentrations greater than 10 nmol l^{-1}, all progress curves displayed a plateau suggesting complete saturation of the EI* complex. Attempts were made to determine a dissociation constant; however, the EI* complex could not be reactivated after 12 h of dialysis suggesting irreversible inhibition by covalent modification. The authors calculated a revised upper limit of 1 nmol l^{-1} for the K_i value of finasteride, which is more consistent with the clinical results.

Finasteride was subsequently shown to be a slow-onset inhibitor of type 1 5AR by researchers at Glaxo Research Institute.[89] Attempts to dialyze and remove the inhibitor from a saturated enzyme–inhibitor complex again failed to display any reactivation of the enzyme even after 3 days. This led to the conclusion that the complex is either irreversible or has a half-life of greater than 10 days. The authors of the second study also demonstrated that the presence of a double bond in ring A of 5AR inhibitors is responsible for the appearance of the slow-onset kinetics. This, combined with the inability to demonstrate reversibility, was cited as further evidence for a covalent modification of the enzyme active site.

A possible chemical mechanism involving a Michael addition at C1 of finasteride by an enzymic nucleophile was proposed. In order to test this mechanism, Prakash and coworkers measured the secondary tritium isotope effect on the EI to EI* transition using [1,2-^{3}H]-finasteride.[90] Large, inverse isotope effects were reported, consistent with a Michael addition to the double bond. Thus, the authors concluded that the mechanism of 5AR inhibition by finasteride was due to a covalent modification of the enzyme (**Figure 21**) although there had been no physical demonstration of such an intermediate.
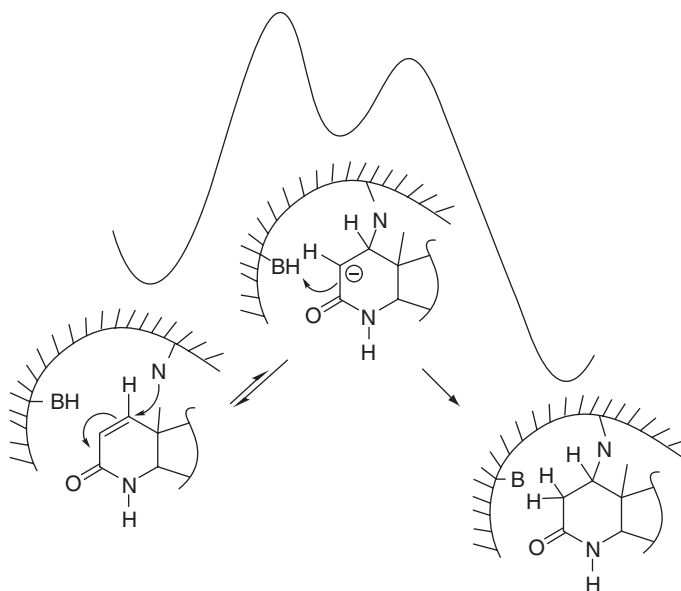
Figure 21 Proposed chemical mechanism of covalent modification of 5α-reductase by finasteride. Reprinted with permission from G. Tian; S.-Y. Chen; K. L. Facchine; S. R. Prakash, *J. Am. Chem. Soc.* **1995**, *117*, 2369–2370. Copyright 1995 American Chemical Society.

Around this time, researchers at Merck reported that finasteride could act as a substrate for type 2 5AR producing dihydrofinasteride with an off-rate of $8 \times 10^{-7}\,s^{-1}$. This rate corresponded to a half-life of approximately 10 days, near the limit that the group from Glaxo would have been able to detect using the dialysis experiments. Researchers at Merck performed a very straightforward experiment to determine if the inhibitor was covalently bound to the enzyme.[91] A solution of ^{3}H-finasteride–enzyme complex was denatured in $7\,mol\,l^{-1}$ guanidine hydrochloride and >98% of the label was released, inconsistent with a covalent modification mechanism of inhibition. Therefore, the Merck group set out to determine the spontaneous dissociation rate by isotope exchange rather than dialysis. The timeframe for this experiment was on the order of weeks (rather than hours or days as in previous experiments) and resulted in k_{off} value of $2.6 \times 10^{-7}\,s^{-1}$. The authors calculated a half-life for the finasteride–enzyme complex of approximately 30 days and a $K_i{}^* \leq 3 \times 10^{-13}$. However, the chemical mechanism of inhibition was unknown.

Further characterization of the chemical outcome of the tritium label revealed that the tritium was only found in the reduced product dihydrofinasteride and not in finasteride. This suggested that finasteride binding was irreversible and all bound finasteride went through catalysis. The possibility that the product dihydrofinasteride was the actual inhibitor was ruled out when it was determined to be a simple, irreversible inhibitor of the enzyme with a $K_i = 1\,nmol\,l^{-1}$. A more vexing problem for the researchers was that while the overall release of tritium from the ^{3}H-finasteride–enzyme complex was >98%, only 30–45% could be recovered as ^{3}H-dihydrofinasteride. This, combined with the inhibition kinetics of dihydrofinasteride, suggested that the potent inhibitor compound had not yet been isolated.

Armed with the knowledge that finasteride was a substrate for the enzyme, researchers revisited the chemical mechanism for 5AR and proposed that a reaction intermediate prior to the formation of dihydrofinasteride could be the inhibitor. The most obvious choice was a covalent adduct between finasteride and NADP$^+$. This type of inhibitor had been observed before in the reverse reaction of lactate dehydrogenase.[92] Repeating the isotope-exchange experiments and isolating the resulting compounds under NADP-friendly conditions, 62% of the label was recovered. Multiple spectroscopic characterizations were all consistent with the bisubstrate analog NADP-dihydrofinasteride (**Figure 22**). Purified NADP-dihydrofinasteride adduct was shown to be a potent inhibitor of type 2 5AR with slow-onset kinetics identical to those determined with finasteride. The bisubstrate analog was competitive with NADP and noncompetitive versus testosterone consistent with bisubstrate inhibition of an ordered bi–bi mechanism. A proposed mechanism for the enzymatic formation of the adduct, consistent with the chemical mechanism for testosterone reduction, is shown in **Figure 23**.

8.19.8.2   Mupirocin/Pseudomonic Acid A

Protein synthesis has been a productive target for the design of antibiotic therapeutics. It is not surprising then that nature has already discovered this family of targets and developed some very potent inhibitors of this process. Mupirocin (pseudomonic acid A) (**Figure 24**) is a natural antibiotic produced by *Pseudomonas fluorescens* isolated almost 40 years ago. In that space of time mupirocin, marketed as Bactroban, has become the most prescribed topical antibiotic.[93] Due to its unique mechanism of action, mupirocin is also used as an antibiotic of last resort versus methicillin-resistant *Staphylococcus aureus* infections of open wounds.[94] However, like most antibiotics, bacteria have also been able to acquire resistance to mupirocin creating problems when this antibiotic is used in clinical settings.[95,96]
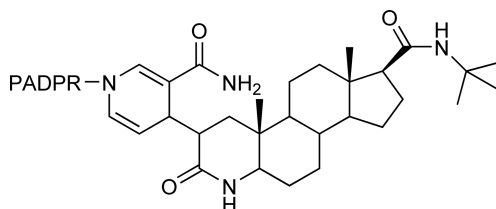
Figure 22   Proposed chemical structure of the NADP–finasteride adduct. PADPR, phosphoadenosine diphosphoribose.
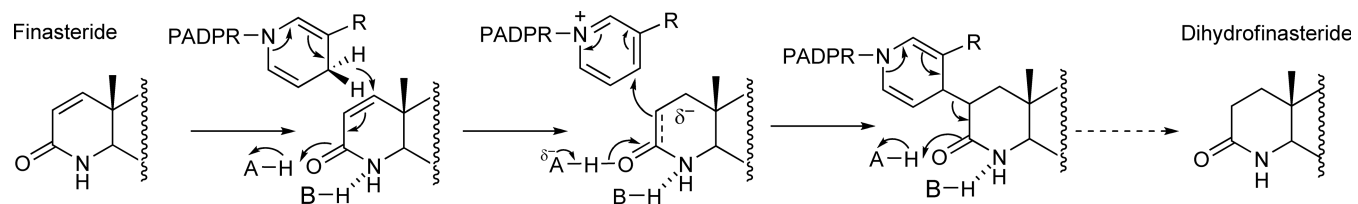
Figure 23 Proposed chemical mechanism for formation of NADP–finasteride bisubstrate analog. Reprinted with permission from H. G. Bull; M. Garcia-Calvo; S. Andersson; W. F. Baginsky; H. K. Chan; D. E. Ellsworth; R. R. Miller; R. A. Stearns; R. K. Bakshi; G. H. Rasmusson; R. L. Tolman; R. W. Myers; J. W. Kozarich; G. S. Harris, *J. Am. Chem. Soc.* **1996**, *118*, 2359–2365. Copyright 1996 American Chemical Society.
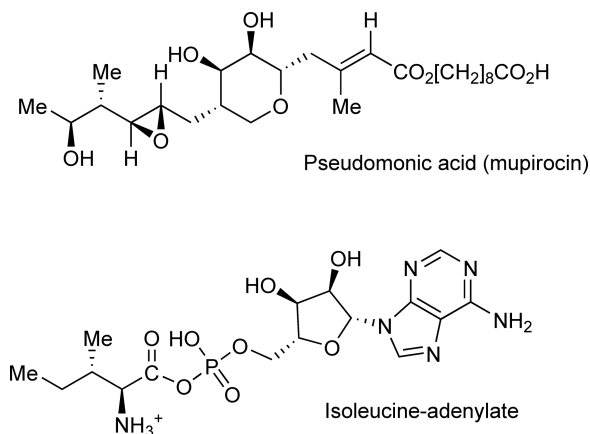
Figure 24 Chemical structures of mupirocin and the isoleucyl-adenylate formed by Ile-RS.

Mupirocin is bacteriostatic at low concentrations versus *S. aureus* and *E. coli* and shows low toxicity.[97] Initial structural characterization was consistent with a 9-hydroxynonanioic acid moiety. Further studies indicated that mupirocin displayed no cross-resistance with other antibiotics suggesting that it worked through a novel mechanism. It was not until 7 years later that the structure and stereochemistry of the compound was elucidated.[98,99] Mupirocin is comprised of two main parts: a C_{17} moiety termed 'monic acid' that consists of a hydrophobic side chain containing an epoxide and an unusual 9-hydroxynonanoic acid joined to the monic acid by an α,β-unsaturated ester linkage. Early isotope feeding studies aimed at understanding the biosynthesis of mupirocin provided evidence for the incorporation of acetate-derived units allowing researchers to propose a biosynthetic pathway.[100] In 2003, characterization of the biosynthetic gene cluster for mupirocin allowed for a more detailed proposal of the biosynthetic pathway by demonstrating the similarity of mupirocin gene clusters to those involved in polyketide synthase and fatty acid synthase type I systems.[101]

Initial studies focused on the effects of mupirocin on various cellular processes. It was demonstrated that treatment of *S. aureus* culture with 0.05–0.5 μg ml^{-1} (near the bacteriostatic MIC values) resulted in decreased RNA and protein synthesis whereas DNA and cell wall production were unaffected.[102] Depletion of an amino acid in the cytoplasm will not only result in the inhibition of protein synthesis, but also activates the stringent control of RNA synthesis. In the presence of the antibiotic chloramphenicol, which releases the stringent control of RNA synthesis, only protein synthesis was affected by mupirocin. This suggested a mode of action that inhibited protein synthesis prior to ribosomal assembly. Soon after, mupirocin was shown to specifically inhibit aminoacylation of isoleucine-tRNA in *E. coli*.[103] From the structural similarities of the epoxide side chain terminus of mupirocin and the aliphatic side chain of isoleucine, it was easily suggested that the mode of action would include a competition for the isoleucine binding site on Ile-tRNA synthetase (Ile-RS). Accordingly, mupirocin is a competitive inhibitor of Ile-tRNA synthetase from *E. coli* versus isoleucine with a K_i of 2–6 nmol l^{-1}, several orders of magnitude below the K_m for isoleucine.[104]

Isoleucyl-tRNA synthetase (Ile-RS) catalyzes the activation of tRNA with the amino acid isoleucine in an ATP-dependent reaction, prior to its utilization by the ribosome during protein synthesis (**Scheme 5**). The enzyme uses a ping-pong kinetic mechanism where ATP and isoleucine react to form isoleucine-AMP with the release of pyrophosphate. In the second step of the reaction, the AMP-activated isoleucine is transferred to the appropriate tRNA with release of AMP. Mupirocin was shown to only interfere with the first step of the reaction. Mupirocin was also characterized as an inhibitor with respect to mammalian Ile-tRNA synthetase.[104] Mupirocin acted as a competitive inhibitor versus Ile, but with a K_i of 20 μmol l^{-1}, about fourfold

$$Ie + ATP \rightleftharpoons Ie - AMP + PP_i$$

$$Ie - AMP + tRNA \rightleftharpoons Ie - tRNA + AMP$$
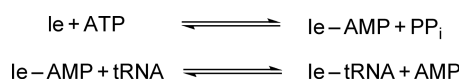
Scheme 5

higher than the K_m for Ile explaining the low toxicity of the antibiotic. Mupirocin was also shown to bind to Ile-tRNA synthetase with a 1:1 stoichiometry; however, there was no evidence for covalent modification of the enzyme as might be expected from the presence of the epoxide and active site cysteines.[104]

In the 1980s and 1990s, organic chemists worked to find derivatives of mupirocin with enhanced efficacy and hydrolytic stability.[105–117] Despite numerous experiments, it proved difficult to improve upon the overall antibiotic power of the original drug, though some improvements have been made toward oral availability. However, when the next flurry of biochemical studies on mupirocin was published (1998–2001) a more in-depth understanding of the drug's mechanism was elucidated. A major breakthrough in this understanding was the description of the three-dimensional structure of Ile-tRNA synthetase from *S. aureus* in the presence of mupirocin and its cognate tRNA substrate (**Figure 25**).[118,119] Mupirocin was found bound to the Rossman domain, a common structural motif for binding dinucleotides, and the site of isoleucine binding. The side chains of Glu554, Asp557, Asn558, and the backbone nitrogen of Gly555 provide interactions with the 6-OH and 7-OH of the monate core. Additionally, the backbone nitrogen of Val588 forms a hydrogen bond with the ester at the beginning of the monate side chain. The epoxide linker group was found pointing into a hydrophobic pocket formed by Pro56 and Pro57. Two motifs that are conserved in class I amino acid tRNA synthetases are also involved in binding of mupirocin. His64 belongs to the conserved 'HIGH' motif and interacts with the monate core ring oxygen. The main chain carbonyls of Met596 and Ser597 are part of the 'KMSKS' motif and form an interaction with the 1′OH of the nonanoic acid moiety of mupirocin. Shortly after the structure of the Ile-RS from *S. aureus* was solved, a second mupirocin bound structure was reported.

Yokoyama and coworkers solved the three-dimensional structure of Ile-RS from *Thermus thermophilus*, an archaebacteria, in the presence of mupirocin and in a second structure, with the nonhydrolyzable substrate analog Ile-AMS (5′-N-[N-(L-isoleucyl)sulfamoyl]adenosine) (compound **22**, **Figure 26**).[120] Mupirocin binds to both the *T. thermophilus* and *S. aureus* enzymes using similar active site interactions. The hydroxyl groups of the pyran hexa-ring mimic the interactions of the 2′ and 3′ hydroxyls of the ribose ring. However, the isoleucyl moiety of mupirocin binds in a different conformation than the isoleucine side chain of Ile-AMS. The authors suggest that this is due to the C13 hydroxyl group that is not represented in isoleucine and the lengthening of the linker group from an epoxy group in mupirocin to a sulfamoyl group of Ile-AMS. Since Ile-RS from *T. thermophilus* is an intermediary in the inhibition spectrum of mupirocin (bacterial > archaebacterial > eukaryotic), the authors hoped that by comparing active site residues of Ile-RS from *S. aureus*, *T. thermophilus*, and *Saccharomyces cerevisiae* they might be able to discern the unique structural components, which lead to the extraordinary selectivity of mupirocin for the bacterial enzyme form. The researchers were able to demonstrate that by making 'eukaryotic-like' substitutions to the *T. thermophilus* enzyme, the K_i could be increased an order of magnitude. However, they were unable to lower the K_i by making '*S. aureus*-like' substitutions, thus leaving the question of selective inhibition by mupirocin unanswered.

In an attempt to address this question, Pope and coworkers performed an in-depth kinetic study of inhibition of Ile-RS from *S. aureus* by both nonhydrolyzable bisubstrate analogs and mupirocin. It is known that nonhydrolyzable analogs of the Ile-adenylate act as potent inhibitors of Ile-tRNA synthetases from all
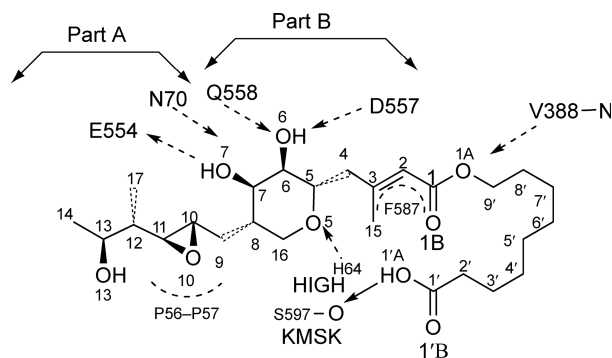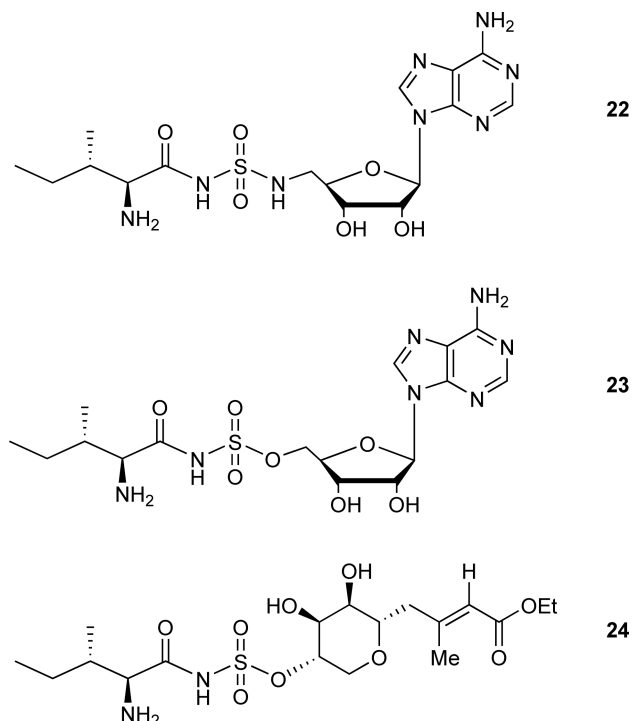

Figure 25   Line drawing of mupirocin bound to Ile-RS.[119]

Figure 26 Chemical structures of nonhydrolyzable bisubstrate analog inhibitors for Ile-RS.

organisms tested; however, mupirocin was a poor inhibitor of Ile-RS from nonbacterial sources. Pope's group was able to demonstrate that mupirocin acts as a slow-onset inhibitor of *S. aureus* Ile-tRNA synthetase whereas nonhydrolyzable analogs act as simple competitive inhibitors.[121] Mupirocin displays a K_i^* value of 23 pmol l^{-1} versus isoleucine.

Biophysical experiments also exhibited a difference in the binding modes of mupirocin and other nonhydrolyzable bisubstrate analogs.[121] Binding of the nonhydrolyzable bisubstrate analog Ile-NHSO$_2$-AMP (**23**, **Figure 26**) to Ile-RS from *S. aureus* resulted in a large increase in the intrinsic fluorescence of the enzyme similar to that seen when the enzyme binds the isoleucyl-adenylate intermediate. However, mupirocin binding caused very little change in fluorescence. It was proposed that this result was due to the differences in the binding of the isoleucyl moiety of each inhibitor. This agreed with the two reported three-dimensional structures, which showed that isoleucyl side chain of mupirocin does not fully occupy the amino acid binding site. This meant that there might be additional binding energy available for suitably designed mupirocin analogs. With the three-dimensional structure and kinetic model at hand, a study to optimize the isoleucyl side chain and linker regions of mupirocin was proposed.

When varying the linker region of mupirocin, the first result was that the presence of the monate core was required for slow-onset inhibition.[122] Removal of the epoxide linker and replacement with a slightly longer linking group to better mimic the distances found in the isoleucyl-adenylate intermediate resulted in increased affinity consistent with filling unoccupied pockets in the amino acid binding site (compound **24**, **Figure 26**). Further lengthening of the linker region compromised the binding affinity. It was also noted that the amino acid carboxylate group was involved in an important interaction with deletion of this group causing a 1000-fold decrease in binding affinity. An examination of the amino acid side chain of the inhibitor demonstrated that L-Ile was indeed the best moiety for inhibition. Interestingly, poor inhibition was observed when the amino acid side chain was substituted with another amino acid and tested as an inhibitor against its cognate tRNA synthetase suggesting that the monate core is specific for the Ile-RS structure. Compound **24** created a large increase in fluorescence of Ile-RS upon binding consistent with full occupation of all active site binding pockets. Attempts to determine a binding constant for the optimized inhibitor using traditional methods were

unsuccessful due to the ultratight binding. Using a melting circular dichroism experiment, the K_d for the compound **24** was estimated to be 10 fmol l^{-1} as compared to 140 pmol l^{-1} for mupirocin.[122] While this study successfully incorporated new binding elements into the inhibitor structure, experiments to look at selectivity at a bacterial/eukaryotic level were not reported.

8.19.8.3   Isoniazid

Isoniazid (isonicotinic acid hydrazide, INH) was reported in 1952 by Middlebrook[123] to be potently effective against *Mycobacterium tuberculosis*, the causative agent of human tuberculosis. This extremely simple synthetic agent (**Figure 27**) exhibited MIC values of 0.03–0.05 µg ml^{-1} against a number of strains of *M. tuberculosis*. While it was originally thought to act as an antimetabolite, or an inactive pyridoxal analog, its mechanism of action remained obscure. Hints concerning its target accumulated in the literature, where its effect on DNA synthesis,[124] the accumulation of C26 fatty acids,[125] the loss of acid-fast staining,[126] and inhibition of mycolic acid biosynthesis[127] were reported. Finally, in 1992, the Cole group at the Institut Pasteur reported that high-level resistance to the drug was due to mutations in *katG*-encoded catalase–peroxidase.[128] Following this report, a number of clinical strains that were similarly resistant to INH were also shown to have mutations in this gene, with most isolates having a single-nucleotide change that generated the S315T mutant form of the protein.[129] This was followed in 1994 by the report that a single step selection for INH resistance in *Mycobacterium smegmatis* selected for mutations that mapped to the promoter region of a two-gene operon encoding the *mabA* and *inhA* genes.[130] The InhA protein was functionally identified as the enoyl-ACP reductase involved in fatty acid elongation in the fatty acid synthase II system.[131] This structural gene mutation, generating a S93A mutant form of the enzyme, had an increased K_m value for the substrate NADH, but no other obvious catalytic diminution.[132] INH was not an inhibitor of the enzyme, making the connection between resistance and *in vitro* activity difficult to understand. However, subsequent reports[133,134] clarified the relationship between the drug, the genetic resistance profile, and the activity against InhA. It is now clear that INH is a prodrug that is activated *in vivo* by the KatG-catalyzed peroxidase reaction to generate a radical species that nonenzymatically reacts with oxidized pyridine nucleotides to generate an 'adduct,' similar to finasteride. This adduct (**Figure 28**) binds to InhA and inhibits the InhA-catalyzed reaction.[135] The three-dimensional structure of 4*S*-isonicotinyl-NAD bound to InhA was reported and revealed that the NAD portion bound in an identical manner to the enzyme compared to NADH alone, and that the isonicotinoyl moiety occupied that adjacent space where the Δ-2-enoyl thioester was bound.[136]
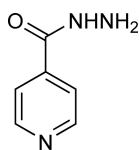
Figure 27   Chemical structure of isoniazid.
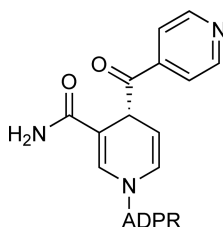
Figure 28   Chemical structure of the isoniazid–NAD adduct.

The kinetics of inhibition of InhA by the adduct was reported by Tonge in 2003,[137] who showed that the adduct exhibited slow-onset inhibition, and that the adduct had nanomolar affinity for the enzyme. The InhA protein was shown to be essential for viability,[138] suggesting that inhibition of this enzyme resulted in the potent bactericidal effect of the drug.[139]

However, the report that the isonicotinoyl-NAD adducts existed as a pair on enantiomers (4R and 4S) as well as pairs of diastereomeric cyclic aminals,[140] suggested that these other adducts might inhibit other pyridine nucleotide-dependent enzymes in *M. tuberculosis*. The early report on a specific effect of INH on DNA synthesis suggested that some enzymes involved in this essential process might be inhibited by a pyridine nucleotide adduct of INH, and dihydrofolate reductase (DHFR) seemed a possibility. This enzyme catalyzes the NADPH-dependent reduction of dihydrofolate and is a validated drug target in other bacteria, as well as mammals. In fact, the isonicotinoyl adduct of NADP was a nanomolar bisubstrate inhibitor of DHFR.[141] The three-dimensional structure of the DHFR–isonicotinoyl–NADP adduct revealed that the enzyme bound the 4R-acyclic stereoisomer exclusively, in contrast to the 4S-isomer of the NAD adduct bound by InhA, consistent with the known stereochemistry of hydride transfer catalyzed by the two enzymes. These data suggested that perhaps there were additional pyridine nucleotide-dependent enzymes that could bind to the isonicotinoyl adducts of NAD and NADP. When affinity chromatography matrices were prepared with either the NAD or NADP adducts and crude soluble extracts of *M. tuberculosis* applied to these columns, a 2000-fold purification resulted. SDS–PAGE of either of the eluates revealed a small number of bands that were excised, subject to trypsin digestion followed by tandem mass spectrometry. About 16 proteins were unambiguously identified and most were known to use either oxidized or reduced pyridine nucleotides as substrates.[142] To date, only one of these has been shown to be inhibited by the isonicotinoyl adduct of NADP; the *E. coli ribD*-encoded bifunctional deaminase–reductase that catalyzes the second step in riboflavin biosynthesis.[143] This protein is composed of two domains, of which the reductase domain is most similar to that of DHFR, and thus its inhibition by the adduct is not surprising.

8.19.9   Future Perspectives

Over the past two decades, the successful use of bisubstrate analogs as potent inhibitors of numerous classes of enzymes has certainly validated the hypothesis that using binding determinants from both substrates in a single molecule is advantageous. Researchers working on multisubstrate enzymes should have no trepidation about proposing the types of experiments described in this chapter. While the current trend may be leaning toward transition-state analogs, there are many details about the structure and function of an enzyme active site that one can probe directly using bisubstrate analogs in a very straightforward manner. The predominant drawback to these types of experiments thus far is the lack of translation to the clinical setting. From the examples cited here, the application of bisubstrate analogs as cell-permeable prodrugs seems promising for improving this situation in the future.

Abbreviations

5AR	Δ^4-steroid 5α-reductase
AAC	aminoglycoside acetyltransferase
AAD	aromatic amino acid decarboxylase
AANAT	arylalkylamine *N*-acetyltransferase
ACC	acetyl-CoA carboxylase
COMT	catechol *O*-methyl transferase
FT	farnesyltransferase
GNAT	Gcn5-related *N*-acetyltransferase
HAT	histone acetyltransferase
Ile-RS	isoleucine tRNA synthetase
INH	isoniazid

L-**DOPA**	L-dihydroxyphenylalanine
PCAF	p300/CBP-associated factor
PKA	cAMP-dependent kinase
PKC	protein kinase C

References

1. A. D. Broom, *J. Med. Chem.* **1989**, *32*, 2–7.
2. M. I. Page; W. P. Jencks, *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1678–1683.
3. R. Wolfenden, *Acc. Chem. Res.* **1972**, *5*, 10–18.
4. P. A. Bartlett; C. K. Marlowe, *Biochemistry* **1983**, *22*, 4618–4624.
5. V. L. Schramm, *J. Biol. Chem.* **2007**, *282*, 28297–28300.
6. J. F. Morrison; C. T. Walsh, *Adv. Enzymol. Relat. Areas Mol. Biol.* **1988**, *61*, 201–301.
7. J. E. Brownell; J. Zhou; T. Ranalli; R. Kobayashi; D. G. Edmondson; S. Y. Roth; C. D. Allis, *Cell* **1996**, *84*, 843–851.
8. S. Magnet; J. S. Blanchard, *Chem. Rev.* **2005**, *105*, 477–498.
9. M. Yu; M. L. Magalhaes; P. F. Cook; J. S. Blanchard, *Biochemistry* **2006**, *45*, 14788–14794.
10. A. P. Carter; W. M. Clemons; D. E. Brodersen; R. J. Morgan-Warren; B. T. Wimberly; V. Ramakrishnan, *Nature* **2000**, *407*, 340–348.
11. C. T. Walsh, *Antibiotics: Actions, Origins, Resistance*; ASM Press: Washington, DC, 2003.
12. N. M. Llewellyn; J. B. Spencer, *Nat. Prod. Rep.* **2006**, *23*, 864–874.
13. J. W. Williams; D. B. Northrop, *J. Antibiot. (Tokyo)* **1979**, *32*, 1147–1154.
14. K. A. Draker; D. B. Northrop; G. D. Wright, *Biochemistry* **2003**, *42*, 6565–6574.
15. F. Gao; X. Yan; O. M. Baettig; A. M. Berghuis; K. Auclair, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 6859–6862.
16. F. Gao; X. Yan; T. Shakya; O. M. Baettig; S. Ait-Mohand-Brunet; A. M. Berghuis; G. D. Wright; K. Auclair, *J. Med. Chem.* **2006**, *49*, 5273–5281.
17. S. Magnet; T. Lambert; P. Courvalin; J. S. Blanchard, *Biochemistry* **2001**, *40*, 3700–3709.
18. M. L. Magalhaes; M. W. Vetting; F. Gao; L. Freiburger; K. Auclair; J. S. Blanchard, *Biochemistry* **2008**, *47*, 579–584.
19. W. Pierpaoli; W. Regelson; C. Colman, *The Melatonin Miracle. Nature's Age-Reversing, Disease-Fighting, Sex-Enhancing Hormone*; Pocket Books: New York, 1995.
20. S. L. Coon; P. H. Roseboom; R. Baler; J. L. Weller; M. A. Namboodiri; E. V. Koonin; D. C. Klein, *Science* **1995**, *270*, 1681–1683.
21. J. De Angelis; J. Gastel; D. C. Klein; P. A. Cole, *J. Biol. Chem.* **1998**, *273*, 3045–3050.
22. E. M. Khalil; P. A. Cole, *J. Am. Chem. Soc.* **1998**, *120*, 6195–6196.
23. E. M. Khalil; J. De Angelis; M. Ishii; P. A. Cole, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12418–12423.
24. E. Wolf; J. De Angelis; E. M. Khalil; P. A. Cole; S. K. Burley, *J. Mol. Biol.* **2002**, *317*, 215–224.
25. T. Jenuwein; C. D. Allis, *Science* **2001**, *293*, 1074–1080.
26. O. D. Lau; T. K. Kundu; R. E. Soccio; S. Ait-Si-Ali; E. M. Khalil; A. Vassilev; A. P. Wolffe; Y. Nakatani; R. G. Roeder; P. A. Cole, *Mol. Cell* **2000**, *5*, 589–595.
27. O. D. Lau; A. D. Courtney; A. Vassilev; L. A. Marzilli; R. J. Cotter; Y. Nakatani; P. A. Cole, *J. Biol. Chem.* **2000**, *275*, 21953–21959.
28. K. G. Tanner; M. R. Langer; J. M. Denu, *Biochemistry* **2000**, *39*, 11961–11969.
29. A. N. Poux; M. Cebrat; C. M. Kim; P. A. Cole; R. Marmorstein, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14065–14070.
30. P. R. Thompson; H. Kurooka; Y. Nakatani; P. A. Cole, *J. Biol. Chem.* **2001**, *276*, 33721–33729.
31. X. Liu; L. Wang; K. Zhao; P. R. Thompson; Y. Hwang; R. Marmorstein; P. A. Cole, *Nature* **2008**, *451*, 846–850.
32. Y. Hwang; P. R. Thompson; L. Wang; L. Jiang; N. L. Kelleher; P. A. Cole, *Angew. Chem. Int. Ed. Engl.* **2007**, *46*, 7621–7624.
33. I. H. Segel, *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady State Enzyme Systems*; Wiley: New York, 1975.
34. M. Cebrat; C. M. Kim; P. R. Thompson; M. Daugherty; P. A. Cole, *Bioorg. Med. Chem.* **2003**, *11*, 3307–3313.
35. V. Sagar; W. Zheng; P. R. Thompson; P. A. Cole, *Bioorg. Med. Chem.* **2004**, *12*, 3383–3390.
36. S. J. Wakil; J. K. Stoops; V. C. Joshi, *Annu. Rev. Biochem.* **1983**, *52*, 537–579.
37. T. Tanabe; K. Wada; T. Okazaki; S. Numa, *Eur. J. Biochem.* **1975**, *57*, 15–24.
38. M. D. Lane; J. Moss; S. E. Polakis, *Curr. Top. Cell. Regul.* **1974**, *8*, 139–195.
39. D. R. Amspacher; C. Z. Blanchard; F. R. Fronczek; M. C. Saraiva; G. L. Waldrop; R. M. Strongin, *Org. Lett.* **1999**, *1*, 99–102.
40. C. Z. Blanchard; D. Amspacher; R. Strongin; G. L. Waldrop, *Biochem. Biophys. Res. Commun.* **1999**, *266*, 466–471.
41. K. L. Levert; G. L. Waldrop, *Biochem. Biophys. Res. Commun.* **2002**, *291*, 1213–1217.
42. K. L. Levert; G. L. Waldrop; J. M. Stephens, *J. Biol. Chem.* **2002**, *277*, 16347–16350.
43. C. Freiberg; N. A. Brunner; G. Schiffer; T. Lampe; J. Pohlmann; M. Brands; M. Raabe; D. Habich; K. Ziegelbauer, *J. Biol. Chem.* **2004**, *279*, 26066–26073.
44. C. Z. Blanchard; G. L. Waldrop, *J. Biol. Chem.* **1998**, *273*, 19140–19145.
45. P. Bilder; S. Lightle; G. Bainbridge; J. Ohren; B. Finzel; F. Sun; S. Holley; L. Al-Kassim; C. Spessard; M. Melnick; M. Newcomer; G. L. Waldrop, *Biochemistry* **2006**, *45*, 1712–17122.
46. K. Parang; P. A. Cole, *Pharmacol. Ther.* **2002**, *93*, 145–157.
47. A. Ricouart; J. C. Gesquiere; A. Tartar; C. Sergheraert, *J. Med. Chem.* **1991**, *34*, 73–78.
48. D. Medzihradszky; S. L. Chen; G. L. Kenyon; B. W. Gibson, *J. Am. Chem. Soc.* **1994**, *116*, 9413–9419.

49. K. Parang; J. H. Till; A. J. Ablooglu; R. A. Kohanski; S. R. Hubbard; P. A. Cole, *Nat. Struct. Biol.* **2001**, *8*, 37–41.
50. K. Kim; P. A. Cole, *J. Am. Chem. Soc.* **1998**, *120*, 6851–6858.
51. A. J. Ablooglu; J. H. Till; K. Kim; K. Parang; P. A. Cole; S. R. Hubbard; R. A. Kohanski, *J. Biol. Chem.* **2000**, *275*, 30394–30398.
52. A. C. Hines; P. A. Cole, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2951–2954.
53. A. C. Hines; K. Parang; R. A. Kohanski; S. R. Hubbard; P. A. Cole, *Bioorg. Chem.* **2005**, *33*, 285–297.
54. M. Loog; A. Uri; G. Raidaru; J. Jarv; P. Ek, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1447–1452.
55. A. Uri; G. Raidaru; J. Subbi; K. Padari; M. Pooga, *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2117–2120.
56. E. Enkvist; D. Lavogina; G. Raidaru; A. Vaasa; I. Viil; M. Lust; K. Viht; A. Uri, *J. Med. Chem.* **2006**, *49*, 7150–7159.
57. B. Ludecke; P. M. Knappskog; P. T. Clayton; R. A. Surtees; J. D. Clelland; S. J. Heales; M. P. Brand; K. Bartholome; T. Flatmark, *Hum. Mol. Genet.* **1996**, *5*, 1023–1028.
58. G. M. Keating; K. A. Lyseng-Williamson, *CNS Drugs* **2005**, *19*, 165–184.
59. E. Nissinen; I. B. Linden; E. Schultz; P. Pohto, *Naunyn Schmiedebergs Arch. Pharmacol.* **1992**, *346*, 262–266.
60. B. Masjost; P. Ballmer; E. Borroni; G. Zurcher; F. K. Winkler; R. Jakob-Roetne; F. Diederich, *Chemistry* **2000**, *6*, 971–982.
61. C. Lerner; B. Masjost; A. Ruf; V. Gramlich; R. Jakob-Roetne; G. Zurcher; E. Borroni; F. Diederich, *Org. Biomol. Chem.* **2003**, *1*, 42–49.
62. R. Paulini; C. Trindler; C. Lerner; L. Brandli; W. B. Schweizer; R. Jakob-Roetne; G. Zurcher; E. Borroni; F. Diederich, *ChemMedChem* **2006**, *1*, 340–357.
63. C. W. Olanow, *Arch. Neurol.* **2000**, *57*, 263–267.
64. R. Paulini; C. Lerner; R. Jakob-Roetne; G. Zurcher; E. Borroni; F. Diederich, *Chembiochem* **2004**, *5*, 1270–1274.
65. F. L. Zhang; P. J. Casey, *Annu. Rev. Biochem.* **1996**, *65*, 241–269.
66. W. R. Schafer; R. Kim; R. Sterne; J. Thorner; S. H. Kim; J. Rine, *Science* **1989**, *245*, 379–385.
67. P. J. Casey; P. A. Solski; C. J. Der; J. E. Buss, *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8323–8327.
68. J. L. Bos, *Cancer Res.* **1989**, *49*, 4682–4689.
69. S. Omura; D. Van der Pyl; J. Inokoshi; Y. Takahashi; H. Takeshima, *J. Antibiot. (Tokyo)* **1993**, *46*, 222–228.
70. K. Shiomi; H. Yang; J. Inokoshi; D. Van der Pyl; A. Nakagawa; H. Takeshima; S. Omura, *J. Antibiot. (Tokyo)* **1993**, *46*, 229–234.
71. K. Hinterding; P. Hagenbuch; J. Retey; H. Waldmann, *Chem. Eur. J.* **1999**, *5*, 227–236.
72. M. Thutewohl; L. Kissau; B. Popkirova; I.-M. Karaguni; T. Nowak; M. Bate; J. Kuhlmann; O. Muller; H. Waldmann, *Bioorg. Med. Chem.* **2003**, *11*, 2617–2626.
73. D. V. Patel; M. M. Patel; S. S. Robinson; E. M. Gordon, *Bioorg. Med. Chem. Lett.* **1994**, *4*, 1883–1888.
74. R. S. Bhide; D. V. Patel; M. M. Patel; S. P. Robinson; L. W. Hunihan; E. M. Gordon, *Bioorg. Med. Chem. Lett.* **1994**, *4*, 2107–2112.
75. D. V. Patel; E. M. Gordon; R. J. Schmidt; H. N. Weller; M. G. Young; R. Zahler; M. Barbacid; J. M. Carboni; J. L. Gullo-Brown; L. Hunihan; C. Ricca; S. Robinson; B. R. Seizinger; A. V. Tuomari; V. Manne, *J. Med. Chem.* **1995**, *38*, 435–442.
76. V. Manne; N. Yan; J. M. Carboni; A. V. Tuomai; C. S. Ricca; J. G. Brown; M. L. Andahazy; R. J. Schmidt; D. Patel; R. Zahler; R. Weinmann; C. J. Der; A. D. Cox; J. T. Hunt; E. M. Gordon; M. Barbacid; B. R. Seizinger, *Oncogene* **1995**, *10*, 1763–1779.
77. D. V. Patel; M. G. Young; S. P. Robinson; L. Hunihan; B. J. Dean; E. M. Gordon, *J. Med. Chem.* **1996**, *39*, 4197–4210.
78. M. Schlitzer; I. Sattler, *Angew. Chem. Int. Ed. Engl.* **1999**, *38*, 2032–2034.
79. M. Schlitzer; M. Bohm; I. Sattler; H. M. Dahse, *Bioorg. Med. Chem.* **2000**, *8*, 1991–2006.
80. M. Schlitzer; M. Bohm; I. Sattler, *Bioorg. Med. Chem.* **2002**, *10*, 615–620.
81. J. Imperato-McGinley; L. Guerrero; T. Gautier; R. E. Peterson, *Science* **1974**, *186*, 1213–1215.
82. S. Andersson; D. M. Berman; E. P. Jenkins; D. W. Russell, *Nature* **1991**, *354*, 159–161.
83. T. Liang; C. E. Heiss; S. Ostrove; G. H. Rasmusson; A. Cheung, *Endocrinology* **1983**, *112*, 1460–1468.
84. T. Liang; C. E. Heiss; A. H. Cheung; G. F. Reynolds; G. H. Rasmusson, *J. Biol. Chem.* **1984**, *259*, 734–739.
85. T. Liang; M. A. Cascieri; A. H. Cheung; G. F. Reynolds; G. H. Rasmusson, *Endocrinology* **1985**, *117*, 571–579.
86. E. Stoner, *J. Steroid Biochem. Mol. Biol.* **1990**, *37*, 375–378.
87. B. Houston; G. D. Chisholm; F. K. Habib, *Steroids* **1987**, *49*, 355–369.
88. B. Faller; D. Farley; H. Nick, *Biochemistry* **1993**, *32*, 5705–5710.
89. G. Tian; J. D. Stuart; M. L. Moss; P. L. Domanico; H. N. Bramson; I. R. Patel; S. H. Kadwell; L. K. Overton; T. A. Kost, *Biochemistry* **1994**, *33*, 2291–2296.
90. G. Tian; S.-Y. Chen; K. L. Facchine; S. R. Prakash, *J. Am. Chem. Soc.* **1995**, *117*, 2369–2370.
91. H. G. Bull; M. Garcia-Calvo; S. Andersson; W. F. Baginsky; H. K. Chan; D. E. Ellsworth; R. R. Miller; R. A. Stearns; R. K. Bakshi; G. H. Rasmusson; R. L. Tolman; R. W. Myers; J. W. Kozarich; G. S. Harris, *J. Am. Chem. Soc.* **1996**, *118*, 2359–2365.
92. J. W. Burgner, II; W. J. Ray, Jr., *Biochemistry* **1984**, *23*, 3636–3648.
93. J. Wuite; B. I. Davies; M. Go; J. Lambers; D. Jackson; G. Mellows, *Lancet* **1983**, *2*, 394.
94. M. W. Casewell; R. L. Hill, *Lancet* **1983**, *2*, 1312.
95. J. G. Hurdle; A. J. O'Neill; I. Chopra, *J. Antimicrob. Chemother.* **2004**, *53*, 102–104.
96. S. Fujimura; Y. Tokue; A. Watanabe, *Antimicrob. Agents Chemother.* **2003**, *47*, 3373–3374.
97. A. T. Fuller; G. Mellows; M. Woolford; G. T. Banks; K. D. Barrow; E. B. Chain, *Nature* **1971**, *234*, 416–417.
98. E. B. Chain; G. Mellows, *J. Chem. Soc. [Perkin 1]* **1977**, (3), 294–309.
99. R. G. Alexander; J. P. Clayton; K. Luk; N. H. Roger; T. J. King, *J. Chem. Soc. Perkin Trans. 1* **1978**, 561–565.
100. T. C. Feline; R. B. Jones; G. Mellows; L. Phillips, *J. Chem. Soc. [Perkin 1]* **1977**, (3), 309–318.
101. A. K. El-Sayed; J. Hothersall; S. M. Cooper; E. Stephens; T. J. Simpson; C. M. Thomas, *Chem. Biol.* **2003**, *10*, 419–430.
102. J. Hughes; G. Mellows, *J. Antibiot. (Tokyo)* **1978**, *31*, 330–335.
103. J. Hughes; G. Mellows, *Biochem. J.* **1978**, *176*, 305–318.
104. J. Hughes; G. Mellows, *Biochem. J.* **1980**, *191*, 209–219.
105. P. Brown; D. J. Best; N. J. P. Broom; R. Cassels; P. J. O'Hanlon; T. J. Mitchell; N. F. Osborne; J. M. Wilson, *J. Med. Chem.* **1997**, *40*, 2563–2570.
106. N. J. P. Broom; R. Cassels; H.-Y. Cheng; J. S. Elder; P. C. T. Hannan; N. Masson; P. J. O'Hanlon; A. Pope; J. M. Wilson, *J. Med. Chem.* **1996**, *39*, 3596–3600.

107. A. Abson; N. J. P. Broom; P. A. Coates; J. S. Elder; A. K. Forrest; P. C. T. Hannan; A. J. Hicks; P. J. O'Hanlon; N. D. Masson; N. D. Pearson; J. E. Pons; J. M. Wilson, *J. Antibiot.* **1996**, *49*, 390–394.
108. P. Brown; D. T. Davies; P. J. O'Hanlon; J. M. Wilson, *J. Med. Chem.* **1996**, *39*, 446–457.
109. N. J. P. Broom; J. S. Elder; P. C. T. Hannan; J. E. Pons; P. J. O'Hanlon; G. Walker; J. Wilson; P. Woodall, *J. Antibiot.* **1995**, *48*, 1336–1344.
110. A. K. Forrest; P. J. O'Hanlon; G. Walker, *J. Chem. Soc. Perkin Trans. 1* **1994**, (18), 2657–2665.
111. A. K. Forrest; P. J. O'Hanlon; G. Walker, *Tetrahedron* **1994**, *50*, 10739–10748.
112. L. L. Klein; C. M. Yeung; P. Kurath; J. C. Mao; P. B. Fernandes; P. A. Lartey; A. G. Pernet, *J. Med. Chem.* **1989**, *32*, 151–160.
113. M. J. Crimmin; P. O'Hanlon; N. H. Rogers; G. Walker, *J. Chem. Soc. Perkin Trans. 1* **1989**, (11), 2047–2057.
114. P. J. O'Hanlon; N. H. Rogers, *Tetrahedron* **1987**, *43*, 2165–2175.
115. M. J. Crimmin; P. J. O'Hanlon; N. H. Rogers, *J. Chem. Soc. Perkin Trans. 1* **1985**, 541–547.
116. M. J. Crimmin; P. J. O'Hanlon; N. H. Rogers, *J. Chem. Soc. Perkin Trans. 1* **1985**, 549–555.
117. S. Coulton; P. J. O'Hanlon; N. H. Rogers, *J. Chem. Soc. Perkin Trans. 1* **1982**, 729–734.
118. L. F. Silvian; J. Wang; T. A. Steitz, *Science* **1999**, *285*, 1074–1077.
119. J. Wang; L. F. Silvian; T. A. Steitz. U.S. Patent 6, 631, 329, 2003.
120. T. Nakama; O. Nureki; S. Yokoyama, *J. Biol. Chem.* **2001**, *276*, 47387–47393.
121. A. J. Pope; K. J. Moore; M. McVey; L. Mensah; N. Benson; N. Osbourne; N. Broom; M. J. Brown; P. O'Hanlon, *J. Biol. Chem.* **1998**, *273*, 31691–31701.
122. M. J. Brown; L. M. Mensah; M. L. Doyle; N. J. Broom; N. Osbourne; A. K. Forrest; C. M. Richardson; P. J. O'Hanlon; A. J. Pope, *Biochemistry* **2000**, *39*, 6003–6011.
123. G. Middlebrook, *Am. Rev. Tuberc.* **1952**, *65*, 765–767.
124. P. R. Gangadharam; F. M. Harold; W. B. Schaefer, *Nature* **1963**, *198*, 712–714.
125. L. A. Davidson; K. Takayama, *Antimicrob. Agents Chemother.* **1979**, *16*, 104–105.
126. K. Takayama; L. Wang; H. L. David, *Antimicrob. Agents Chemother.* **1972**, *2*, 29–35.
127. F. G. Winder; P. B. Collins, *J. Gen. Microbiol.* **1970**, *63*, 41–48.
128. Y. Zhang; B. Heym; B. Allen; D. Young; S. Cole, *Nature* **1992**, *358*, 591–593.
129. B. Heym; P. M. Alzari; N. Honore; S. T. Cole, *Mol. Microbiol.* **1995**, *15*, 235–245.
130. A. Banerjee; E. Dubnau; A. Quemard; V. Balasubramanian; K. S. Um; T. Wilson; D. Collins; G. de Lisle; W. R. Jacobs, Jr., *Science* **1994**, *263*, 227–230.
131. A. Quemard; J. C. Sacchettini; A. Dessen; C. Vilcheze; R. Bittman; W. R. Jacobs, Jr.; J. S. Blanchard, *Biochemistry* **1995**, *34*, 8235–8241.
132. L. A. Basso; R. Zheng; J. M. Musser; W. R. Jacobs, Jr.; J. S. Blanchard, *J. Infect. Dis.* **1998**, *178*, 769–775.
133. K. Johnsson; P. G. Schultz, *J. Am. Chem. Soc.* **1994**, *116*, 7425–7426.
134. R. F. Zabinski; J. S. Blanchard, *J. Am. Chem. Soc.* **1997**, *119*, 2331–2332.
135. A. Dessen; A. Quemard; J. S. Blanchard; W. R. Jacobs, Jr.; J. C. Sacchettini, *Science* **1995**, *267*, 1638–1641.
136. D. A. Rozwarski; G. A. Grant; D. H. Barton; W. R. Jacobs, Jr.; J. C. Sacchettini, *Science* **1998**, *279*, 98–102.
137. R. Rawat; A. Whitty; P. J. Tonge, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13881–13886.
138. C. Vilcheze; H. R. Morbidoni; T. R. Weisbrod; H. Iwamoto; M. Kuo; J. C. Sacchettini; W. R. Jacobs, Jr., *J. Bacteriol.* **2000**, *182*, 4059–4067.
139. C. Vilcheze; F. Wang; M. Arai; M. H. Hazbon; R. Colangeli; L. Kremer; T. R. Weisbrod; D. Alland; J. C. Sacchettini; W. R. Jacobs, Jr., *Nat. Med.* **2006**, *12*, 1027–1029.
140. M. Nguyen; A. Quemard; S. Broussy; J. Bernadou; B. Meunier, *Antimicrob. Agents Chemother.* **2002**, *46*, 2137–2144.
141. A. Argyrou; M. W. Vetting; B. Aladegbami; J. S. Blanchard, *Nat. Struct. Mol. Biol.* **2006**, *13*, 408–413.
142. A. Argyrou; M. W. Vetting; J. S. Blanchard, *J. Am. Chem. Soc.* **2007**, *129*, 9582–9583.
143. M. L. Magalhaes; A. Argyrou; S. M. Cahill; J. S. Blanchard, *Biochemistry* **2008**, *47*, 6499–6507.

Biographical Sketches

Patrick A. Frantom received his BS degree in biochemistry from Louisiana State University. Following graduation, he began graduate studies at the Department of Biochemistry at Texas

A&M University as a student in Dr. Paul Fitzpatrick's laboratory. His graduate research focused on understanding the chemical mechanism of the enzyme tyrosine hydroxylase using kinetic isotope effects. Upon receiving his Ph.D. in biochemistry, he began postdoctoral studies in the laboratory of Dr. John Blanchard at Albert Einstein College of Medicine. His work at Dr. Blanchard's laboratory focused on the characterization of enzymes as potential drug targets against the human pathogen *Mycobacterium tuberculosis*. He was recently appointed assistant professor of chemistry at the University of Alabama, where he will focus on the chemical and regulatory mechanisms of enzymes. His work has been supported by a number of funding agencies including a graduate training grant from the United States National Institutes of Health and a postdoctoral fellowship from the Charles H. Revson Foundation.

John S. Blanchard received his BS in chemistry from Lake Forest College and obtained his Ph.D. from the laboratory of W. W. Cleland at the University of Wisconsin. After a 3-year NIH-sponsored postdoctoral fellowship, he was appointed assistant professor of biochemistry at the Albert Einstein College of Medicine in New York City in 1983. In 1998, he became the Dan Danciger Professor of Biochemistry. His early research interests focused on the determination of kinetic isotope effects exhibited by flavin-containing enzymes. His collaborative studies on the mechanism of action, and resistance, to isoniazid in *Mycobacterium tuberculosis* led to his current interests in antibiotic resistance. His present interests include the structure and function of essential biosynthetic enzymes in *M. tuberculosis*, resistance to aminoglycosides and fluoroquinolones, and proteome-wide identification of acetylated proteins. He is the author of over 140 research papers and 20 reviews and has been awarded seven United States patents. His work has been generously supported by the United States National Institutes of Health for the last 24 years.

8.20 Quantum Chemical Modeling of Enzymatic Reactions – Applications to Epoxide-Transforming Enzymes

Kathrin H. Hopmann and Fahmi Himo, Royal Institute of Technology, Stockholm, Sweden

8.20.1 Introduction

Density functional theory (DFT) is today a very powerful tool in the study of electronic structures of molecules. Advancements in DFT, in particular the development of Becke's 3-parameter functional (B3LYP), together with the nearly exponential growth of computer power, have made it possible to treat ever larger systems at a reasonable level of accuracy. Using the B3LYP with a medium-sized basis set, one can routinely handle systems containing more than 100 atoms today, a development that has opened the door for many applications. One of the fields that quantum chemical methods have had very positive impacts on in recent years is the study of enzymatic reaction mechanisms.

When investigating mechanistic proposals, one has to perform a large number of calculations to find different transition states and intermediates to test the various scenarios. Consequently, the computational scheme used has to be fast and robust enough to allow this. At the same time, the accuracy of the approximations made in the models has to be higher than or comparable to the accuracy of the underlying computational method. One very fruitful approach has been to cut out a relatively small model of enzyme around the active site and treat it at a quite high level of theory. The effects of the parts of the enzyme that are not included in the quantum model are modeled using different approximations.

To account for the polarization effects, the enzyme surrounding to a first approximation can be considered as a homogenous polarizable medium, which can be modeled using some dielectric cavity techniques. In addition, to model the steric effects that the enzyme surrounding imposes on the active site, it has been shown to be very useful to simply fix atoms at the edge of the active site model. The combination of continuum solvation and the coordinate-locking scheme represents a quite simple but yet powerful way to account for the parts of the enzyme that are not included in the model.

It is not *a priori* evident that such a model is able to describe enzymatic reaction mechanisms correctly. However, one can start by noting that the energies involved in chemical reactions, bond breaking and formation, are usually much higher than long-range electrostatic effects. The effect of the catalyst, in this case the active site of the enzyme, is thus to a large extent local, and environmental effects are usually of lower order. Accordingly, in the last decade, many researchers have used this strategy to develop models of enzyme active sites and to study their reaction mechanisms.[1–21] The large number of investigations testifies to the usefulness of such models.

In this chapter, we will provide an overview of the employed methodology. To illustrate the various aspects of the methodology and to give the reader a feeling about the state of the art of the field, three very recent applications will be discussed in detail. All three enzymes are concerned with epoxide-transforming reactions, namely limonene epoxide hydrolase (LEH),[22] soluble epoxide hydrolase (sEH),[23,24] and haloalcohol dehalogenase C (HheC).[25,26] First, however, a brief presentation of DFT and its accuracy will be given.

8.20.2   Density Functional Theory

8.20.2.1   General

Almost all quantum chemical studies of enzyme reactions use DFT today.[1–7] In DFT, the basic quantity is the electron density, in contrast to the wavefunction, in wavefunction-based methods. Modern DFT is founded on the Hohenberg–Kohn theorems from 1964.[27] The first theorem shows that the ground-state energy of a system is uniquely defined by its electron density $\rho(\mathbf{r})$, while the second theorem shows that the density-dependent functional obeys the variational principle. The total energy of the system can hence be expressed as a functional of the electron density, $E[\rho]$, which transforms the problem of finding the $3n$-dimensional wavefunction (n, number of particles) to finding the three-dimensional electron density. However, the exact form of this functional is unknown. In the Kohn–Sham formulation of DFT,[28] the problem reduces to finding the so-called exchange-correlation functional, F_{XC}. Over the years, a whole range of different functionals has been developed. For molecular systems, the most successful functionals are the ones that include also the gradient of the density, $E[\rho, \nabla\rho]$, in the so-called generalized gradient approximation.

8.20.2.2   B3LYP

B3LYP is one of the most popular DFT functionals.[29–34] It is a so-called hybrid functional and is usually expressed in the following form:

$$F_{XC}^{B3LYP} = (1-a)F_X^{Slater} + aF_X^{HF} + bF_X^{B88} + cF_C^{LYP} + (1-c)F_C^{VWN}$$

where F_X^{Slater} refers to the Slater exchange, F_X^{HF} the Hartree–Fock exchange, F_X^{B88} Becke's exchange functional,[31] F_C^{LYP} the correlation functional of Lee, Yang, and Parr,[32] and F_C^{VWN} the correlation functional of Vosko, Wilk, and Nusair.[33] The coefficients are $a = 0.20$, $b = 0.72$, and $c = 0.81$, which were adapted from another hybrid functional, B3PW91.[34] The values of the coefficients were originally determined empirically by a linear least-squares fit to 116 experimentally determined energies.[34]

8.20.2.3   Accuracy of B3LYP

Various benchmark tests have been performed with B3LYP to establish its accuracy with respect to geometries and energies.

The accuracy of B3LYP with regard to structural parameters has been evaluated on 53 molecules from the G2 test set.[35] This set consists of 71 bond lengths, 26 bond angles, and 2 dihedral angles. The mean absolute errors at the B3LYP/6-31G(d) level of theory are 0.013 Å for bond lengths, 0.62° for angles, and 0.35° for dihedral angles.[35] The errors in bond lengths and angles were slightly reduced if a larger basis set was used. At the B3LYP/6-311+G(3df,2p) level of theory, the average errors were 0.008 Å for bond lengths and 0.61° for angles. Increasing the basis set also resulted in an increased error for the dihedral angles of 3.66°. However, as only two dihedral angles were included in the test set, the result should be evaluated with caution.[35] The overall results indicate that B3LYP has good accuracy with respect to geometrical parameters, already with a medium-sized basis set.

The accuracy of B3LYP with respect to various absolute and relative energies has been evaluated. Atomization energies for 41 molecules, 26 diatomics from the G2 test set plus 15 other molecules, yielded an average error of 2.2 kcal mol^{-1} at the B3LYP/6-311+G(3df,2p) level.[35] Atomization energies calculated with B3LYP/6-311+G(3df,2p) for the entire G2 set (55 molecules) showed the same average error, 2.2 kcal mol^{-1}, independent if geometries were optimized with a medium or large basis set.[36]

An extensive evaluation has been done by Curtiss *et al.*, who tested a number of density functionals on the G3/05 test set.[37] This set includes 454 energies, all of which have experimental uncertainties less than ± 1 kcal mol^{-1}. The computed results are based on single-point B3LYP/6-311+G(3df,2p) energies at second-order Møller–Plesset perturbation theory (MP2)/6-31G(d) geometries with scaled (0.89) Hartree–Fock (HF)/6-31G(d) zero-point energies. Table 1 shows that B3LYP achieves high accuracy for certain energies in the G3/05 set such as proton affinities, while it performs less well for enthalpies of formation.

The study of enzymatic reactions involves optimization of transition states and evaluation of reaction barriers. The performance of B3LYP in this respect is thus of particular interest. A number of benchmarks have been performed, mostly on small organic reactions. Kang and Musgrave[38] have investigated a variety of hydrogen atom transfer and nonhydrogen abstraction reactions. Optimizations and energy determinations were performed at the B3LYP/6-311+G(3df,2p) level and include zero-point vibrational effects. The barriers of 29 hydrogen-atom transfer reactions were compared to experimental values, resulting in a mean absolute deviation of 3.3 kcal mol^{-1} for B3LYP.[38] For all except two reactions, B3LYP underestimated the barrier. For 11 nonhydrogen abstraction reactions, comparison to experimental values results in a mean absolute deviation of 4.2 kcal mol^{-1} for B3LYP.[38] For seven of these, the barrier was underestimated, while four reactions exhibited an overestimation of the barrier.

Truhlar and coworkers have studied the barrier heights of 38 hydrogen-transfer reactions as well as the forward and backward barriers for 19 nonhydrogen transfer reactions (including heavy atom transfer, nucleophilic substitution, and association reactions). B3LYP exhibited mean errors of 4.3 and 4.2 kcal mol^{-1}, respectively (B3LYP/MG3S energies were determined on QCISD/MG3 geometries).[39] A systematic underestimation of barriers by B3LYP was observed. For nine pericyclic reactions, the enthalpies of activation at the B3LYP/6-31G level of theory show a mean absolute deviation from predicted $\Delta H^{\ddagger}$ values of 1.7 kcal mol^{-1}.[40]

It can thus be concluded that the average error of B3LYP on reaction barriers for small organic reactions is a few kilocalories per mole. Considering that the parameters of this functional were fitted to atomization energies and ionization potentials and not to barrier heights, B3LYP performs surprisingly well. Unfortunately, for enzymatic reactions, no extensive benchmarks exist. Siegbahn[41] concludes that using B3LYP, the error in

Table 1 Mean Absolute Deviation (kcal mol^{-1}) of B3LYP on the G3/05 set[18]

Energies (# of test set energies)	Mean absolute deviation
Enthalpies of formation (270)	4.63
Ionization energies (105)	3.83
Electron affinities (63)	2.99
Proton affinities (10)	1.39
Hydrogen bond strength (6)	1.19
All (454)	4.11

relative energies of enzymatic reactions is in general about $3\,\text{kcal}\,\text{mol}^{-1}$ for molecules containing first- and second-row atoms. For systems involving transition metals, the error appears to be larger, but rarely more than $5\,\text{kcal}\,\text{mol}^{-1}$.[41]

8.20.3   Modeling Enzyme Active Sites and Reactions

8.20.3.1   Construction of the Model

Knowledge about the enzyme structure is usually a prerequisite to set up a quantum chemical model and investigate the reaction mechanism. However, there are cases where the energetic feasibility of reaction mechanisms can be evaluated by studying individual steps without information about the structures. One example is the study of pyruvate-formate lyase (PFL), where the calculations were able to support one of the suggested mechanisms before the X-ray crystal structure was solved.[42] Another example is the study of the reaction mechanism of spore-photoproduct lyase (SPL), for which the crystal structure still remains to be solved.[43]

Information about how the substrate binds to the active site is also very useful. This could come from X-ray crystal structures determined in the presence of substrate analogues or inhibitors, or from mutant structures with real substrates. If this information is not available, various possibilities can be considered and assessed based on the calculated barriers of the following steps. It should be stressed here that the active site models used in this kind of investigations cannot be used for docking studies. These need in general much larger models.

Information about the residues involved in the reaction from for example, mutational studies, is also very valuable when setting up the model, as groups that are known to be essential for the reaction must be included in the model for a correct description of the reaction.

Protein X-ray crystal structures only reveal the position of heavy atoms whereas hydrogen atoms have to be added manually to the model. This is a straightforward procedure except in cases where the protonation state of a group is unknown. General considerations about the pK_a value of a residue in aqueous solution can provide a guideline. However, the pK_a value of a residue in the enzyme active site can be affected by the surroundings and might be lower or higher than its corresponding value in aqueous solution. It is thus not always straightforward how a given residue should be modeled. One possibility is to build two models, one with the given group in the neutral state and one where it is ionized, and then to assess the effect on the reaction mechanism. Another possibility is to compute the relative pK_a value of a residue by computing its proton affinity and comparing it to a similar compound with known pK_a value.

Taking into account this information the model can now be set up. First, the residues of interest are extracted from the Protein Data Bank (PDB) file of the X-ray crystal structure. Also, other possibly important elements of the active site are included, such as water molecules, inhibitors or substrate analogues, and metal ions. The extracted residues are often truncated so that in principle only their side chains are included in the final model. This is necessary to reduce the size of the model and the computational cost. For example, tyrosine and phenylalanine are typically modeled by phenol and phenyl, respectively; aspartate and glutamate by acetic acid; asparagine and glutamine by acetamide; and serine by ethanol. However, in some calculations, parts of the backbone are also included, for example, if this is suspected to be involved in important interactions with the substrate.

Figure 1 shows an example of how a model of the active site of the human sEH was built. Important active site residues were extracted from the PDB file (1VJ5)[44] and truncated. The N-cyclohexyl-N'-(iodophenyl) urea inhibitor (CIU) present in the crystal structure was remodeled into the substrate $(1S,2S)$-β-methylstyrene oxide (MSO).[23] The phenyl side chain and the oxygen atom of the inhibitor were kept and the missing atoms were added manually. Two different orientations of the substrate were tested and were found to yield virtually identical barriers for epoxide opening.[23]

Typically, active site models consist of 100–150 atoms today. It is possible to do larger models, but the calculations become very time consuming. A strategy to gain deeper understanding into the reaction mechanism and the roles of the various groups at the active site is to start with a small model and gradually increase it by adding more groups. This allows the identification of the important parts contributing to the reactivity and selectivity. By directly setting up a large model, this task becomes more difficult. Another aspect is that the local
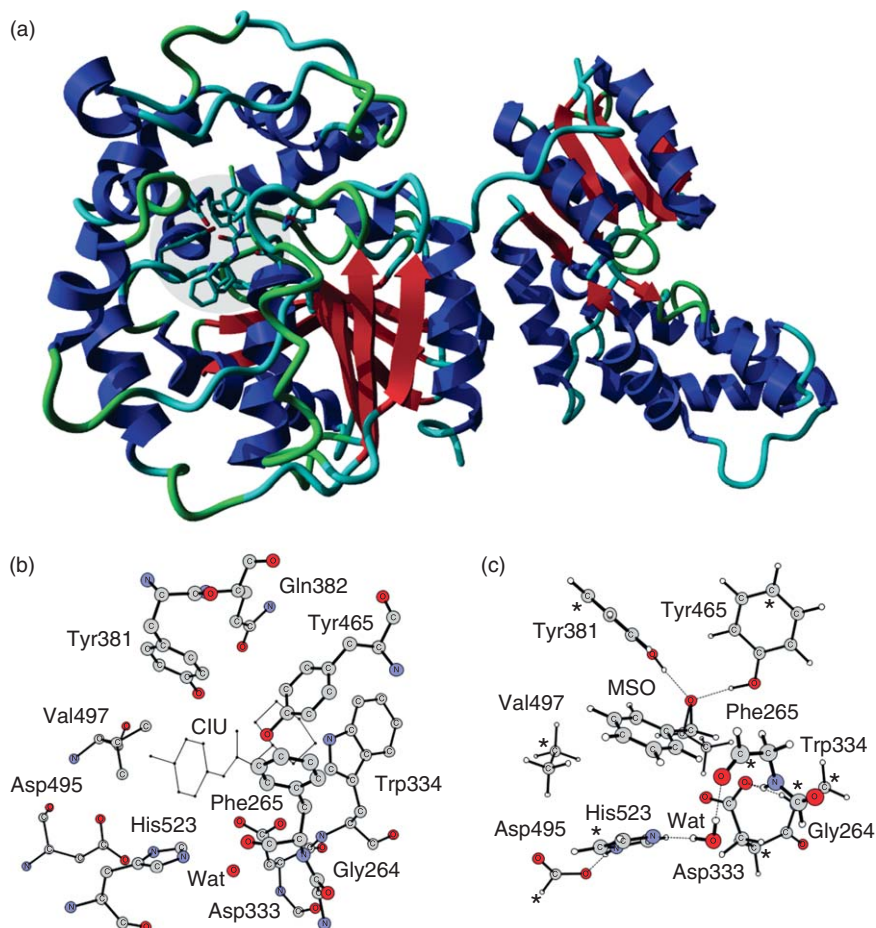
Figure 1 Building an active site model of the human soluble epoxide hydrolase. (a) X-ray crystal structure with the active site highlighted (PDB 1VJ5); (b) Important active site residues, a water molecule, and the CIU inhibitor, are extracted from the PDB file; (c) Final quantum chemical model of the sEH active site. Residues are truncated so that in principle only important side chains and backbone parts were included in the model. The substrate MSO is modeled instead of the inhibitor. Asterisks indicate atoms that were kept fixed to their crystallographically observed positions.

structures of the transition states and intermediates are usually quite similar in the various models. Thus, by spending the effort in finding these stationary points using a rather small model first and then transferring the information to the large model speeds up the process considerably.

Active site models of 100–150 atoms open the door for the study of mutational effects. *In silico* mutation of a residue can be analogous to experimental introduction of mutation, such as replacing one natural residue with another. The mutant quantum chemical model is typically based on the X-ray crystal structure of the wild-type enzyme (unless a mutant structure exists), and it is thus assumed that the overall structure of the active site is not affected significantly by the mutation. This is a reasonable assumption for mutations that involve change of only one or a few atoms, for example, a Tyr → Phe or a Ser → Ala mutation. For more dramatic mutations it might be difficult to predict how the conformation of the active site would change and the computational results might thus be less useful for predicting the behavior of a putative experimental mutant. Nonetheless, these mutations can still provide insight into the role of a given group.

One particular advantage of *in silico* mutations is the fact that 'nonnatural' mutations can also be easily introduced, for example, mutations of the backbone to assess the importance of a backbone hydrogen bond donor or acceptor. For example, in the study of the epoxide-opening reaction of the HheC, a backbone amide in the halide-binding site was mutated into an ester to evaluate the importance of the hydrogen bond between the backbone and the nucleophile (see Section 8.20.4.3.4).

A final comment about the model's size is that during optimizations of the geometries along a given reaction pathway it is important to make sure that the groups that do not directly take part in the chemical transformation are in the same local minima. A change in hydrogen-bonding patterns can, for example, significantly affect the obtained energies and give erroneous barriers and reaction energies. This problem becomes very significant as the model size increases.

The central question in modeling enzyme reactions using quantum chemistry is how an enzyme consisting of thousands of atoms can possibly be modeled with an active site model of say 100 atoms. What are the effects of the rest of the enzyme and how can one model them? The protein surrounding can affect the reactions in two major ways, namely by long-range polarization and by imposing steric restraints on the active site. In the quantum chemical approach used here, these effects are taken care of in different ways, which will be discussed in the following sections.

8.20.3.2 Solvation Effects

The solvation effects that the protein environment surrounding the active site would provide are in this approach modeled by assuming that the surrounding is a homogenous polarizable medium with some dielectric constant. Polarizable continuum model (PCM) techniques can then be used to estimate the solvation energy. In this technique, a cavity around the solute (in this case, the active site model) is created based on its shape. The cavity shape is based on the van der Waals radii of the atoms that make up the model. The main parameter here is the dielectric constant, ε. The choice is somewhat arbitrary, but a typical value in modeling protein surroundings is $\varepsilon = 4$, which roughly corresponds to a mixture of a protein medium (dielectric constant of 2–3) and water (dielectric constant of 80). The choice of ε will be further discussed in connection with the study of one of the enzymes below, where different dielectric constants were investigated. However, some general principles can be mentioned here. The solvation effects on the energies saturate very quickly as a function of the dielectric constant. This is easily rationalized by considering the simple Onsager model of a dipole in a spherical cavity, for which the solvation energy is proportional to $(\varepsilon - 1)/(2\varepsilon + 1)$. The difference between using $\varepsilon = 4$ and $\varepsilon = 5$, for example, is rather small, which makes the procedure quite robust. However, in cases where the conclusions depend on the choice of the dielectric constant, then they cannot be trusted and a larger model has to be devised.

It is important to note that although the solvation energy itself can be large, relative solvation between stationary points on the potential energy surface could be very small. If the charge distribution at the active site is not changed much or if the change is accommodated by the model, that is, the change is not close to the edge of the model, the solvation effects will be rather small. In other words, as the size of the model increases, the solvation effects will diminish. Hence, the inaccuracy of the approximation of using a homogenous PCM decreases as the model size increases. The further the truncation is made from the active site, the better this approximation performs because most of the polarization effects on the reactive parts are already explicitly included in the quantum calculations. This will be illustrated by the examples below.

8.20.3.3 Locking Truncations

As noted above, the quantum chemical active site model is usually composed only of parts of the active site residues, while the remaining protein cannot be included. The geometrical constraints that would be imposed by the backbone and other residues are thus not present.

One way to at least partially account for this is to use a freezing scheme. The enzyme environment can prevent a certain group from moving in a certain way or from making a certain rotation. Groups that are not bound to a metal center or linked by bonds or hydrogen bonds to some other groups at the active site move typically in shallow potentials. In the quantum chemical models, if not constrained somehow during the geometry optimizations, these groups can move a lot, even if the energetic driving force is very small. In order to keep the various groups in place to as much as possibly resemble the crystal structure, certain atoms in the model, typically where the truncation is done, are kept fixed to their X-ray positions. This approach ensures structural integrity of the model, yet allowing for some flexibility of various groups. If the freezing scheme is

overdone, for example, by locking too many centers or by locking them close to the reacting centers, the model can become too rigid and yield wrong energies. As for the solvation effects, the error made by this approximation becomes smaller as the freezing points move further away from the active site, that is, as the size of the quantum model increases.

The combination of continuum solvation and the freezing scheme has proven to be accurate enough to test mechanistic proposals. It represents a very good alternative to the demanding quantum mechanics (QM)/molecular mechanics (MM) calculations. The calculated energies are often sufficient to substantiate or rule out a suggested reaction mechanism.

8.20.3.4 Transition State Theory

The calculated energies and barriers can be readily related to the reaction rates using classical transition state theory (TST). The relationship between the rate constant k of a reaction and the free energy of activation $(\Delta G^{\neq})$ can be expressed as:

$$k = \frac{k_B T}{h} \exp\left(\frac{-\Delta G^{\neq}}{RT}\right)$$

where k is the rate constant (s^{-1}), k_B Boltzmann's constant $(1.38 \times 10^{-23}\,J\,K^{-1})$, T the temperature (in kelvin, 298.15 at room temperature), h Planck's constant $(6.626 \times 10^{-34}\,Js)$, $\Delta G^{\neq}$ the Gibbs free energy difference between the reactant and the transition states, and R the gas constant $(8.314\,J\,K^{-1}\,mol^{-1})$.

Despite its simplicity, this expression provides a very powerful tool to compare the theoretical results to available experimental rates. One can, for example, easily derive that a suggested enzymatic reaction pathway is not feasible if the barrier is calculated to be above $20\,kcal\,mol^{-1}$. It is also important to note the exponential relationship between the rate and the barrier in the above equation. At room temperature, for every $1.4\,kcal\,mol^{-1}$ change in barrier, the rate changes with a factor of 10. A small error in the calculated barrier therefore yields a large error in the rate. The approximation works quite well for the purpose of substantiating or refuting reactions mechanisms, or to compare energies of different reaction pathways. However, it cannot be used to predict accurate reaction rates.

It should be noted that the barriers presented here only correspond to the enthalpy part $(\Delta H^{\neq})$ of the free energy of activation $(\Delta G^{\neq} = \Delta H^{\neq} - T\Delta S^{\neq})$. The entropic part is usually much harder to calculate accurately. It is therefore assumed that $\Delta G^{\neq} \approx \Delta H^{\neq}$, which in many cases is a quite valid assumption because the change in entropy $(\Delta S^{\neq})$ in going from the reactant to the transition state often is small. Obvious exceptions are when a gas molecule is bound or released in the reaction. Special care has to be taken in these cases.

8.20.3.5 Computational Details

Geometry optimization are typically performed using a medium-sized double-ζ basis set, such as 6–31G(d,p) or LANL2DZ. This gives usually quite accurate geometries. However, to obtain accurate energies, larger basis sets have to be used, typically of triple-ζ nature with additional diffuse and polarization functions, such as 6-311+G(2d,2p). These calculations are performed as single points based on the optimized geometries.

Typically, Hessian calculations are then performed to confirm the nature of the stationary point, with no negative eigenvalues for minima, and one negative eigenvalue for transition states. The Hessian calculations also give the zero-point vibrational energies.

Hessian calculations are very demanding, and for very large models they are many times not possible to calculate. In these cases, zero-point vibrational effects can be calculated for smaller models and then transferred to the large models.

Solvation corrections are calculated using the cavity techniques discussed above. The conductor-like polarizable continuum model (CPCM) is used.[45–48]

8.20.4   Applications to Epoxide-Transforming Enzymes

We will provide here three examples of recent applications concerned with the modeling of enzymatic epoxide transformation. Epoxides are versatile compounds and understanding their enzymatic transformation in detail is not only important from a fundamental enzymology point of view but also practically useful for biocatalytic applications. The three enzymes considered are LEH, the human sEH, and HheC.

8.20.4.1   Limonene Epoxide Hydrolase

8.20.4.1.1   *Background*

LEH originates from the bacterium *Rhodococcus erythropolis* DCL14.[49,50] LEH is part of a limonene degradation pathway where it catalyzes the conversion of limonene-1,2-epoxide to limonene-1,2-diol (**Scheme 1**).[49,51,52]

The crystal structure of LEH revealed a six-stranded β-sheet with three α-helices packed on one side (**Figure 2**).[53] The putative active site is composed of a cluster of five charged and polar residues and contains a proposed Asp–Arg–Asp catalytic triad.[53] This topology and active site composition were fundamentally different from other EHs known at the time and LEH was therefore suggested to be the founding member of a new protein family.[49] A few additional members of this family have been identified meanwhile.[54,55]

LEH displays relatively narrow substrate specificity and accepts only few substrates. These include both enantiomers of 1-methylcyclohexene oxide (**1** and **2**, **Scheme 2**) and all four stereoisomers of the natural substrate limonene-1,2-epoxide (**3–6**, **Scheme 2**). The substrates are converted with different enantioselectivities and regioselectivities.[50,56] The four stereoisomers of limonene-1,2-epoxide are hydrolyzed in an enantioconvergent fashion. Conversion of the diastereomeric mixture of **3** and **4** leads to enantioconvergent formation of (1*R*,2*R*,4*S*)-limonene-1,2-diol, whereas conversion of **5** and **6** leads to enantioconvergent

Limonene-1,2-*epoxide*   *Limonene*-1,2-*diol*
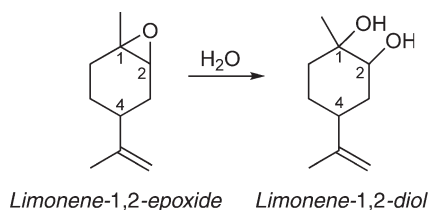
**Scheme 1**   Epoxide hydrolysis reaction catalyzed by LEH.
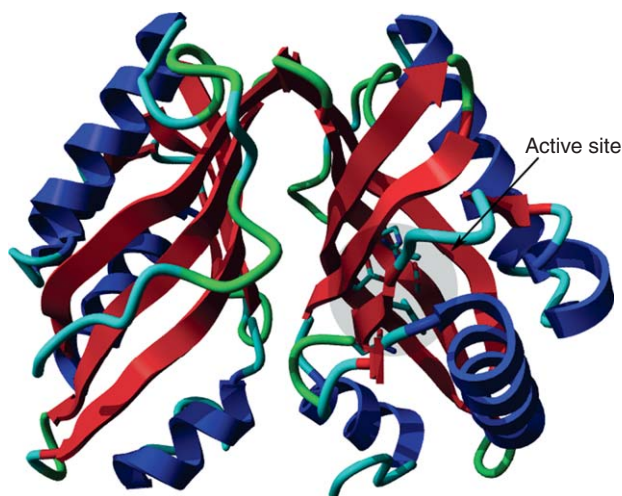
Figure 2   X-ray crystal structure of LEH (based on PDB 1NU3). The active site of one of the two subunits of the homodimer is highlighted.
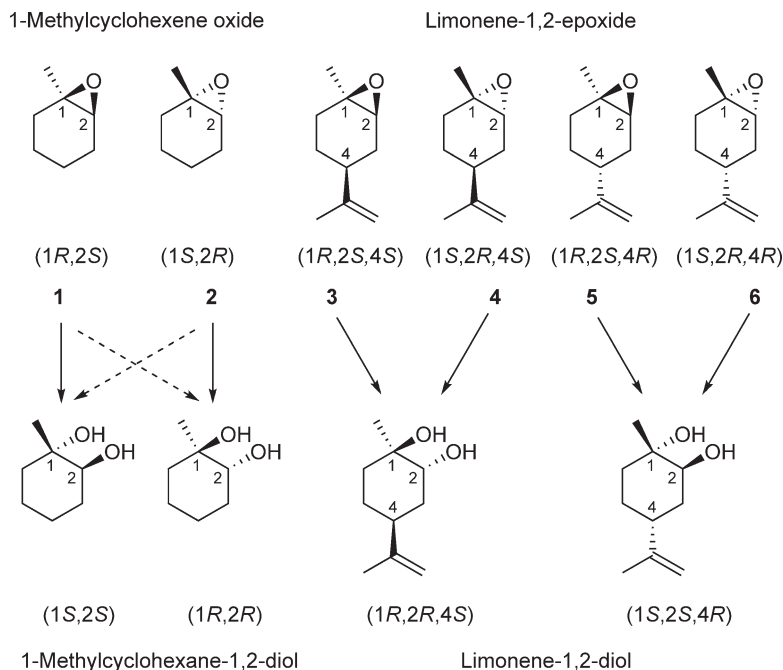
Scheme 2 Conversion of the two enantiomers of 1-methylcyclohexene oxide and the four stereoisomers of limonene-1,2-epoxide by LEH. Adapted from K. H. Hopmann; B. M. Hallberg; F. Himo, *J. Am. Chem. Soc.* **2005**, *127*, 14339–14347.

formation of (1*S*,2*S*,4*R*)-limonene-1,2-diol (**Scheme 2**).[50] The ability to catalyze formation of optically pure products makes LEH interesting from a biocatalytic point of view.

LEH constitutes a novel enzyme and clues about its possible reaction mechanism were limited. A putative active site was identified from the crystal structure and mutational studies were performed on several of the active site residues.[53] Tyr53, Asn55, Arg99, Asp101, and Asp132 were identified as being important for the catalytic activity of LEH.[53] Based on the experimental results, a possible Asp–Arg–Asp catalytic triad was suggested (**Scheme 3**).[53] Asp132 is proposed to activate the nucleophilic water molecule, which attacks the oxirane ring. Asp101 is suggested to donate a proton to the emerging oxyanion of the substrate. Arg99 is proposed to interact with and orient the two aspartate residues. Tyr53 and Asn55 are suggested to be important for binding of the nucleophilic water molecule. In the X-ray crystal structure of LEH, an ordered water molecule is observed close to these two residues.[53]
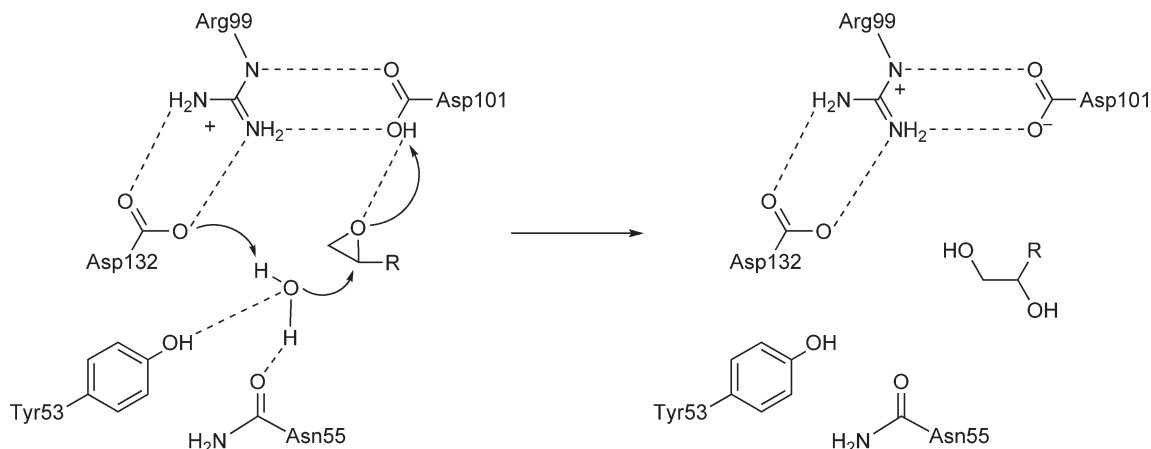
Scheme 3 Proposed epoxide hydrolysis mechanism of LEH.[53]

8.20.4.1.2 Modeling the reaction mechanism of LEH

We studied the reaction mechanism and regioselectivity of LEH with a quantum chemical model based on the crystal structure of LEH in complex with the inhibitor valpromide (PDB 1NU3[53]).[22] The model was composed of the catalytically important residues Tyr53, Asn55, Arg99, Asp101, and Asp132 (**Figure 3**). Also included was the crystallographically observed water molecule. The overall charge of the model was zero. With this model, we studied the conversion of **1**–**6** (**Scheme 2**).

Hydrolysis of **5** in the LEH active site model is shown in **Figure 3**. For formation of the experimentally observed diol, (1*S*,2*S*,4*R*)-limonene-1,2-diol, a barrier of 14.9 kcal mol^{-1} and a reaction energy of -9.7 kcal mol^{-1} were computed. The experimental activation energy for LEH-mediated conversion of **5** is 12.4 kcal mol^{-1}, which is close to the computed value.[51] The activation energy was determined with a diastereomeric mixture of **5 and 6**. However, as their hydrolysis occurs sequentially, it is assumed that the determined value is based on the substrate that is converted first, that is, **5**.

Our results show that the suggested catalytic triad indeed is able to catalyze the conversion of limonene-1,2-epoxide to limonene-1,2-diol. Asp132 abstracts a proton from the nucleophilic water molecule, which attacks the substrate. Asp101 donates a proton to the epoxide oxygen in concert with epoxide opening. Arg99 stabilizes and orients the two aspartate residues. It could also be important for transferring a proton from Asp132 to Asp101 in a subsequent step to restore the active site for the next catalytic cycle. Tyr53 and Asn55 bind and orient the catalytic water molecule.

### 8.20.4.1.3 Analysis of the regioselectivity of LEH

The regioselectivity of LEH has been studied experimentally with different substrates. Studies with the enantiomers of 1-methylcyclohexene oxide (**1 and 2**, **Scheme 2**) revealed preferred attack at the methyl-substituted oxirane carbon, C1, with a regioselectivity of 85(C1):15(C2).[56] This indicated an acid-catalyzed mechanism, which would result in preferred attack at the more substituted carbon.[56] However, conversion of limonene-1,2-epoxide did not support this conclusion and showed somewhat intriguing results. Exclusive attack at the more substituted carbon (C1) is seen for the stereoisomers **4 and 5**, while exclusive attack at the less substituted carbon (C2) is observed for stereoisomers **3 and 6** (**Scheme 2**).[50,52] Interestingly, the two limonene-1,2-epoxide stereoisomers with the same stereochemistry at the oxirane ring, (1*R*,2*S*) for **3 and 5** and (1*S*,2*R*) for **4 and 6**, exhibit attack at opposite carbons (**Scheme 2**). A suggested explanation for the differences was differential binding of the substrates in the active site, which would lead to attack at different carbons.[53]

We investigated the regioselectivity of epoxide opening with the LEH active site model.[22] This model is able to reproduce the experimentally observed regioselectivity for all studied substrates. Our results show that it is not differential binding of the substrate that causes the differences in regioselectivity but that the decisive factor is the conformation of the substrate. Both 1-methylcyclohexene oxide and limonene-1,2-epoxide are cyclic epoxides. While a cyclohexane ring typically prefers to be in a chair conformation, the presence of the oxirane ring in cyclohexene oxide forces the epoxide into a half-chair conformation (**Scheme 4(a)**). The half-chair conformation can exist in two different forms, which are best described by the helicity about the 3,4 bond. These will here be referred to as the *P* and the *M* helicities, respectively (**Scheme 4(a)**). The two helicities are in equilibrium and are present in ratios depending on the energy difference between the two conformers. While unsubstituted cyclohexene oxide can be expected to exist in an equal mixture of the two helicities, this is different for epoxides with large substituents, such as the isopropyl substituent of limonene-1,2-epoxide (**Scheme 4(b)**).

A second important observation regards the reactivity of the epoxide carbons of a cyclic epoxide. In a given half-chair conformation, one of the two carbons will display higher reactivity than the other. This is because of the conformational changes that are required to proceed through the transition state. Attack at one of the two carbons will result in a chair-like conformation whereas attack at the other carbon will distort the half-chair into a twist-boat conformation (**Scheme 5**). The chair-like conformation is energetically preferred and the transition state proceeding through this conformation is hence lower in energy.[22] This is independent of the substitution at the oxirane carbons, that is, the preferred carbon can be the more or the less substituted carbon. These two observations, the half-chair conformers and their intrinsic reactivity, are sufficient to explain the regioselectivity of LEH-mediated epoxide conversion.
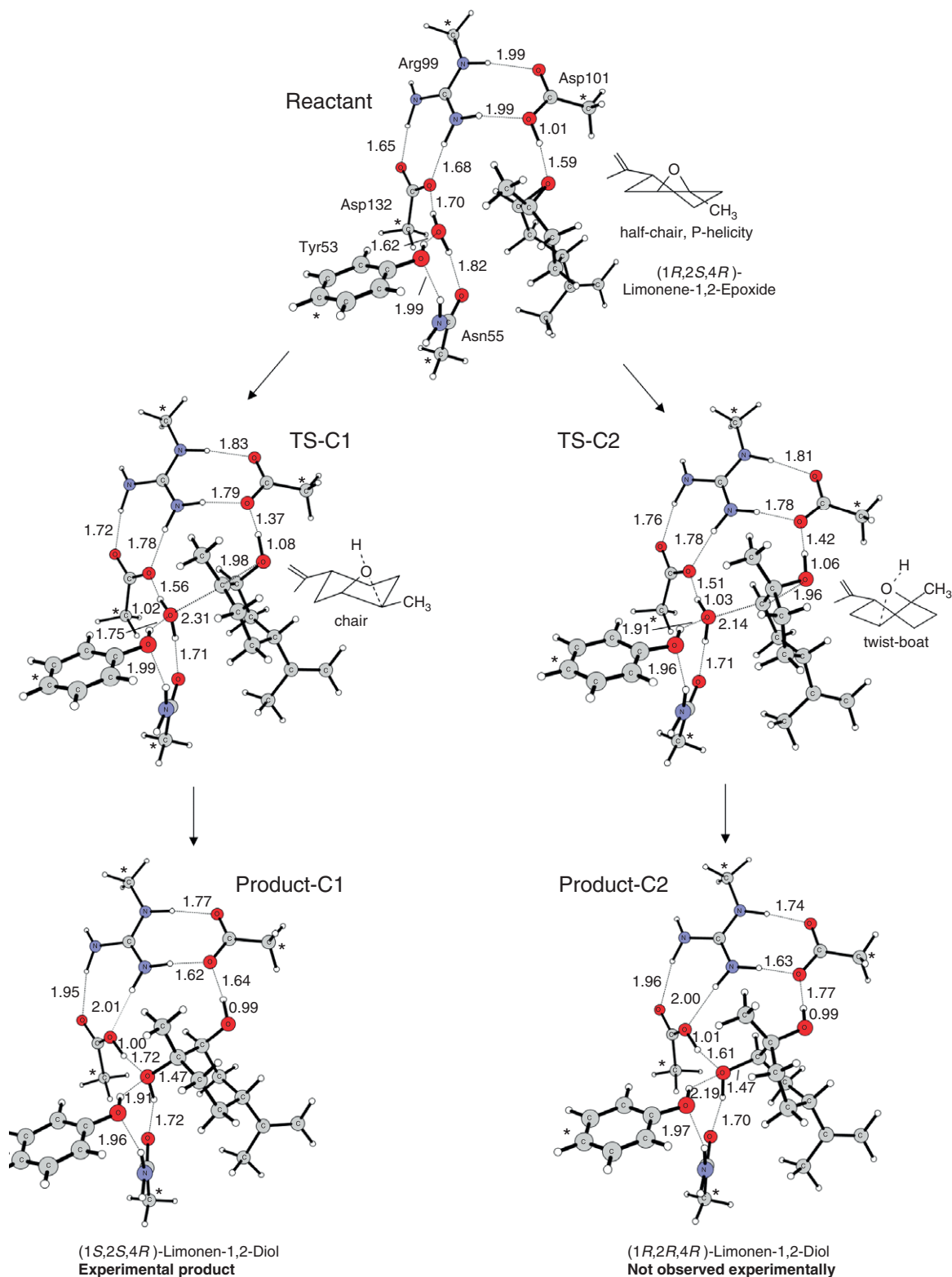
Figure 3   Conversion of **5** (helicity 3,4 *P*) in the LEH active site model. Two possible reaction pathways are shown. Only the product (1*S*,2*S*,4*R*)-limonene-1,2-diol is observed experimentally. Insets show the substrate conformation. Asterisks indicate atoms, which in calculations were fixed to their crystallographically observed position.
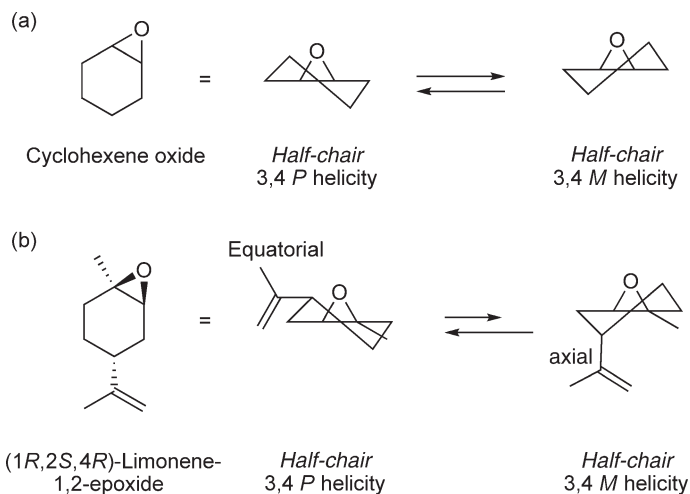
Scheme 4 (a) Cyclohexene oxide adopts a half-chair conformation that exists in two different helicities. (b) The isopropyl substituent of limonene-1,2-epoxide makes one of the two half-chair conformations energetically more favorable.
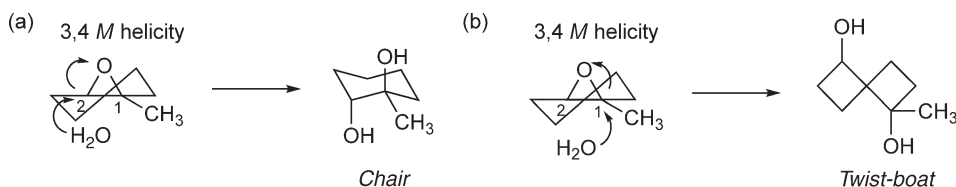
Scheme 5 Attack of water on **1**. Attack at C2 leads to a chair conformation, while attack at C1 leads to a twist boat.

8.20.4.1.3(i) Regioselectivity of 1-methylcyclohexene oxide hydrolysis Of the two stereoisomers of 1-methylcyclohexene oxide, **1** is the preferred substrate of LEH.[56] The two helicity conformers of **1** have almost the same energy and can be assumed to exist in equal amounts.[22] For each conformer, attack is preferred at the carbon that will lead to a chair transition state, for **1** in the *M*-helicity this is C2 and for **1** in the *P*-helicity this is C1. The two helicity forms of **1** are competing substrates of LEH, but because attack at the more substituted carbon C1 of the *P*-helicity has a lower barrier (14.9 kcal mol^{-1}, **Table 2**) than attack at C2 of the *M*-helicity (15.9 kcal mol^{-1}, **Table 2**), attack occurs preferentially, but not exclusively, at C1 of the *P*-helicity. A mixture of products is thus expected, with preferred attack at C1 (of the *P*-helicity) and minor attack at C2 (of the *M*-helicity). The difference in barrier of 1 kcal mol^{-1} corresponds well to the experimentally observed regioselectivity of 85(C1):15(C2) for 1-methylcyclohexene oxide.[56]

Table 2 Calculated barriers and reaction energies (kcal mol^{-1}) for LEH-mediated conversion of 1-methylcyclohexene oxide to 1-methylcyclohexane-1,2-diol

Substrate[a]	Attack at carbon	TS[b]	Product[c]	Barrier	Reaction energy
(1*R*,2*S*), 3,4 *M*	C1	Twist-boat	(1*S*,2*S*)	17.5	−3.4
(1*R*,2*S*), 3,4 *M*	C2	Chair-like	(1*R*,2*R*)	15.9	−9.9
(1*R*,2*S*), 3,4 *P*	C1	Chair-like	(1*S*,2*S*)	14.9	−9.5
(1*R*,2*S*), 3,4 *P*	C2	Twist-boat	(1*R*,2*R*)	19.2	−4.0
(1*S*,2*R*), 3,4 *M*	C1	Chair-like	(1*R*,2*R*)	16.0	−9.0
(1*S*,2*R*), 3,4 *M*	C2	Twist-boat	(1*S*,2*S*)	19.1	−3.2
(1*S*,2*R*), 3,4 *P*	C1	Twist-boat	(1*R*,2*R*)	19.0	−2.8
(1*S*,2*R*), 3,4 *P*	C2	Chair-like	(1*S*,2*S*)	15.7	−9.5

[a] Epoxide stereochemistry and helicity around the 3,4 bond.
[b] Conformation of the substrate in the transition state.
[c] Stereochemistry of resulting diol.
TS, transition state.

8.20.4.1.3(ii) Regioselectivity of limonene-1,2-epoxide hydrolysis For limonene-1,2-epoxide, the natural substrate of LEH, the situation is less complex than for 1-methylcyclohexene oxide. The isopropyl substituent at C4 of limonene-1,2-epoxide determines the preferred helicity of the half-chair, and for each stereoisomer, only the helicity with the substituent in an equatorial position will be observed (**Scheme 4**). For this helicity, attack is preferred at the carbon that leads to a chair-like transition state. The transition states for attack at either C1 or C2 of **5** are shown in **Figure 3**. Attack at C1 of **5** leads to a chair-like transition state and exhibits a barrier of 14.9 kcal mol^{-1}. Attack at C2 of **5** results in a twist-boat transition state and exhibits a barrier of 19.5 kcal mol^{-1} (**Table 3**). The computed barriers indicate that exclusive attack at C1 is expected, in perfect agreement with experimental results.[50,51] Calculated regioselectivities for the other stereoisomers also agree with experimental results, with **4** exhibiting preferred attack at C1, and **3 and 6** exhibiting preferred attack at C2 (**Table 3**).[50,51] Our results clearly show that the regioselectivity of limonene-1,2-epoxide opening is not determined by binding of the substrate in the LEH active site but by intrinsic conformational factors.

8.20.4.2 Soluble Epoxide Hydrolase

8.20.4.2.1 Background

The sEH enzyme catalyzes the hydrolysis of various epoxides to their corresponding diols. Homologues of sEH have been identified in almost all organisms, including bacteria, fungi, plants, insects, and mammals. The human sEH is an interesting enzyme that possesses two different active sites, one for epoxide hydrolysis and one for phosphate ester hydrolysis.[44] In humans, sEH is mainly found in the liver, where it converts various xenobiotic epoxides into their corresponding vicinal diols.[57,58] sEH exhibits a broad substrate specificity but seems to prefer *trans*-substituted epoxides.[57]

sEH belongs to the superfamily of α/β-hydrolase fold enzymes.[44,59] These enzymes do not only share structural but also mechanistic similarities. The mechanism of human sEH was proposed based on studies of sEH homologues from other organisms as well as the closely related microsomal EH. Epoxide hydrolysis is proposed to occur through a covalent mechanism, where an active site aspartate becomes covalently bound to the substrate (**Scheme 6**).[60–63] The formed ester bond is subsequently hydrolyzed by water, which is activated by a His–Asp charge relay, similar to the charge relay in serine hydrolases.[64–69] The sEHs also contain two active site tyrosines that seem to be of importance for stabilization and protonation of the emerging epoxide oxyanion.[44,70–75] A conserved HG*X*P motif ($X =$ any residue) was also identified, which is implicated in formation of the oxyanion hole that stabilizes the tetrahedral intermediate formed in the hydrolytic half-reaction (**Scheme 6**).[76]

Table 3 Calculated barriers and reaction energies (kcal mol^{-1}) for LEH-mediated conversion of limonene-1,2-epoxide to limonene-1,2-diol

Substrate[a]	Attack at carbon	TS[b]	Product[c]	Barrier	Reaction energy
(1*R*,2*S*,4*S*), 3,4 *M*	C1	Twist-boat	(1*S*,2*S*,4*S*)	17.6	−3.5
(1*R*,2*S*,4*S*), 3,4 *M*	C2	Chair-like	(1*R*,2*R*,4*S*)	16.5	−9.7
(1*S*,2*R*,4*S*), 3,4 *M*	C1	Chair-like	(1*R*,2*R*,4*S*)	16.1	−9.5
(1*S*,2*R*,4*S*), 3,4 *M*	C2	Twist-boat	(1*S*,2*S*,4*S*)	19.0	−3.6
(1*R*,2*S*,4*R*), 3,4 *P*	C1	Chair-like	(1*S*,2*S*,4*R*)	14.9	−9.7
(1*R*,2*S*,4*R*), 3,4 *P*	C2	Twist-boat	(1*R*,2*R*,4*R*)	19.5	−4.1
(1*S*,2*R*,4*R*), 3,4 *P*	C1	Twist-boat	(1*R*,2*R*,4*R*)	19.0	−2.8
(1*S*,2*R*,4*R*), 3,4 *P*	C2	Chair-like	(1*S*,2*S*,4*R*)	16.3	−9.4

[a] Epoxide stereochemistry and helicity around the 3,4 bond.
[b] Conformation of the substrate in the transition state.
[c] Stereochemistry of resulting diol.
TS, transition state.
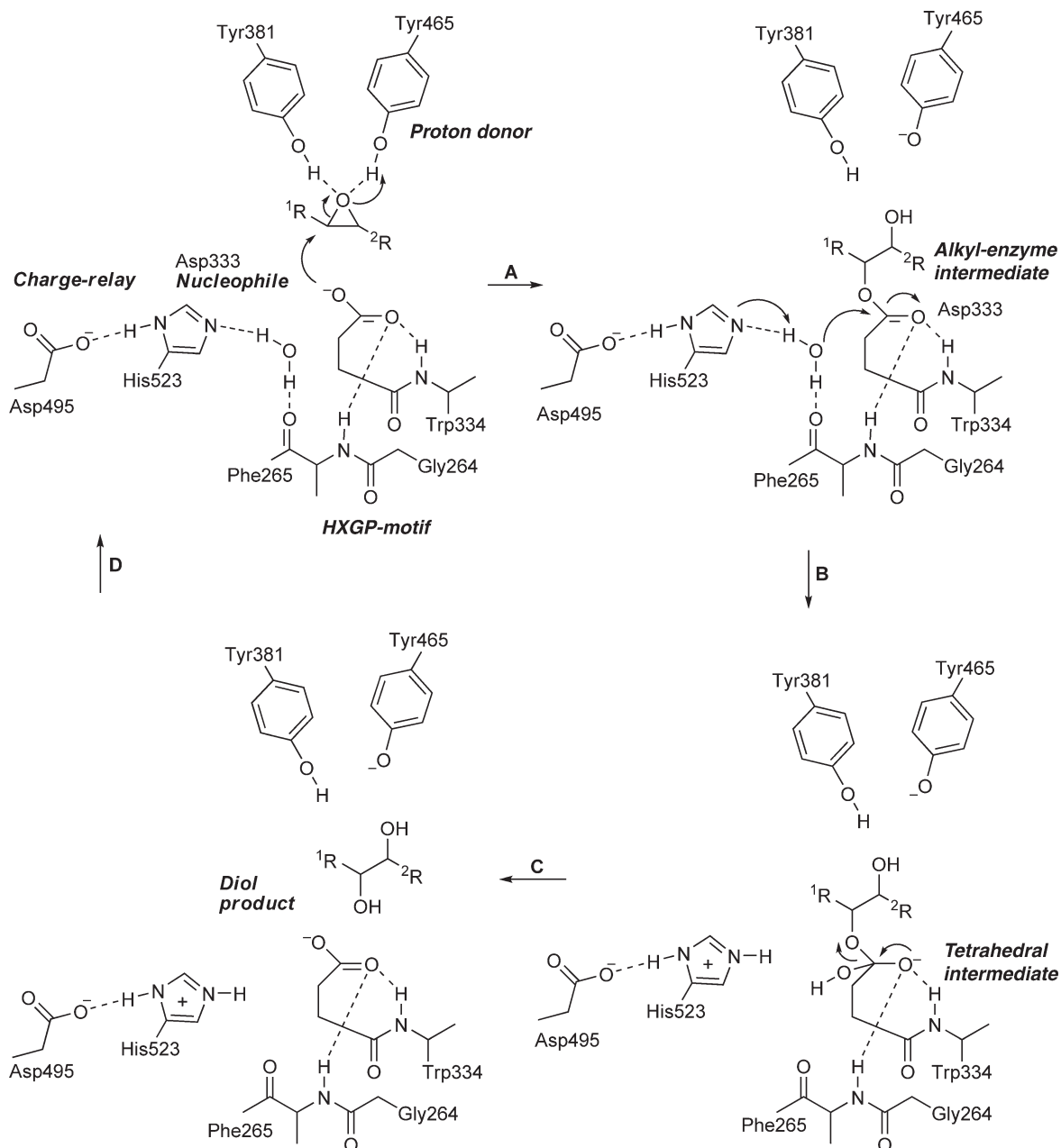
Scheme 6   Epoxide hydrolysis mechanism of soluble epoxide hydrolase (residue numbering as in human sEH). Step A is referred to as the alkylation half-reaction while steps B and C comprise the hydrolytic half-reaction.

8.20.4.2.2   *Quantum chemical studies of human sEH*

We studied the mechanism of sEH using a quantum chemical model based on the X-ray crystal structure of the human sEH in complex with the inhibitor CIU, PDB 1VJ5[44] (**Figure 1**). The model is composed of parts of the residues Gly264, Phe265, Asp333, Trp334, Tyr381, Tyr465, Asp495, Val497, and His523. Also included are a crystallographically observed water molecule and parts of CIU, which was remodeled into the substrate MSO (**Scheme 7**). The model was used to study the full reaction pathway for attack at C1 and C2 of MSO (**Figure 4**).[23]
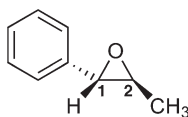
Scheme 7 (1S,2S)-β-methylstyrene oxide substrate used in quantum chemical studies of sEH.
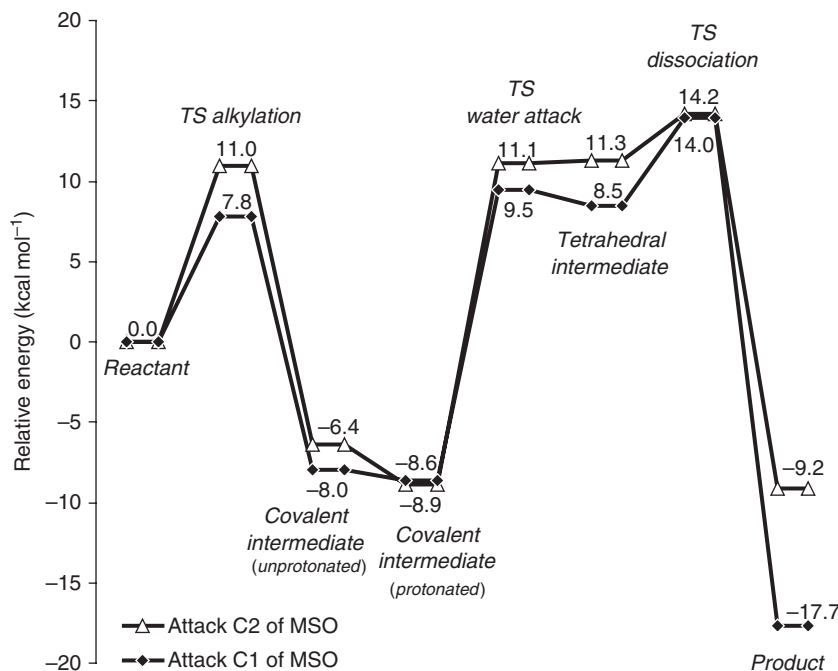
Figure 4 Potential energy curves for hydrolysis of (1S,2S)-β-methylstyrene oxide (MSO) in the human sEH active site model. Adapted with permission from K. H. Hopmann; F. Himo, *Chem. Eur. J.* **2006**, *12*, 6898–6909.

Our results give general support to the suggested sEH mechanism (**Scheme 6**) with some minor modifications. The optimized transition states for the pathway involving attack at C1 are shown in **Figure 5**. The calculations indicate that the alkylation half-reaction occurs in two steps. In the first step, the Asp333 carboxylate attacks the epoxide ring, resulting in the formation of a covalent enzyme–substrate intermediate (**Figure 5(b)**). The barrier for this step is 7.8 kcal mol^{-1} (**Figure 4**). The emerging epoxide oxyanion is stabilized by the two tyrosine residues but a proton transfer to the substrate occurs in a second, subsequent step. Experimental results for the sEH homologue from potato, StEH, give support to a scenario in which alkylation and protonation do not occur concertedly.[77]

Also, the hydrolytic half-reaction occurs in two steps. In the first part, Asp495 and His523 act in concert to activate a water molecule, which attacks the Asp333–substrate bond (**Figure 5(c)**). The barrier for this step is 18.1 kcal mol^{-1}. The formed tetrahedral intermediate is stabilized by the hydrogen bonds from the backbone nitrogens of Trp334 and Phe265. In our calculations, dissociation of the tetrahedral intermediate occurs in concert with proton transfer from Nε of His523 to the forming diol (**Figure 5(d)**). The barrier for this step is 5.5 kcal mol^{-1}. The calculations thus indicate that His523 has a dual role in sEH-mediated epoxide hydrolysis, acting first as a general base and then as general acid. The energetics computed for conversion of MSO show that the hydrolytic half-reaction is rate limiting (**Figure 4**). This is in line with experimental results for the sEH homologues EchA and StEH.[78,79]

We also tested another mechanistic proposal, in which His523 was protonated at Nε during the first step of the reaction.[24] It has been suggested that the protonated His523 would orient and activate Asp333 for nucleophilic attack.[80] Protonation of His523 during the alkylation half-reaction was found to be unlikely based on the calculated energetics.[24]
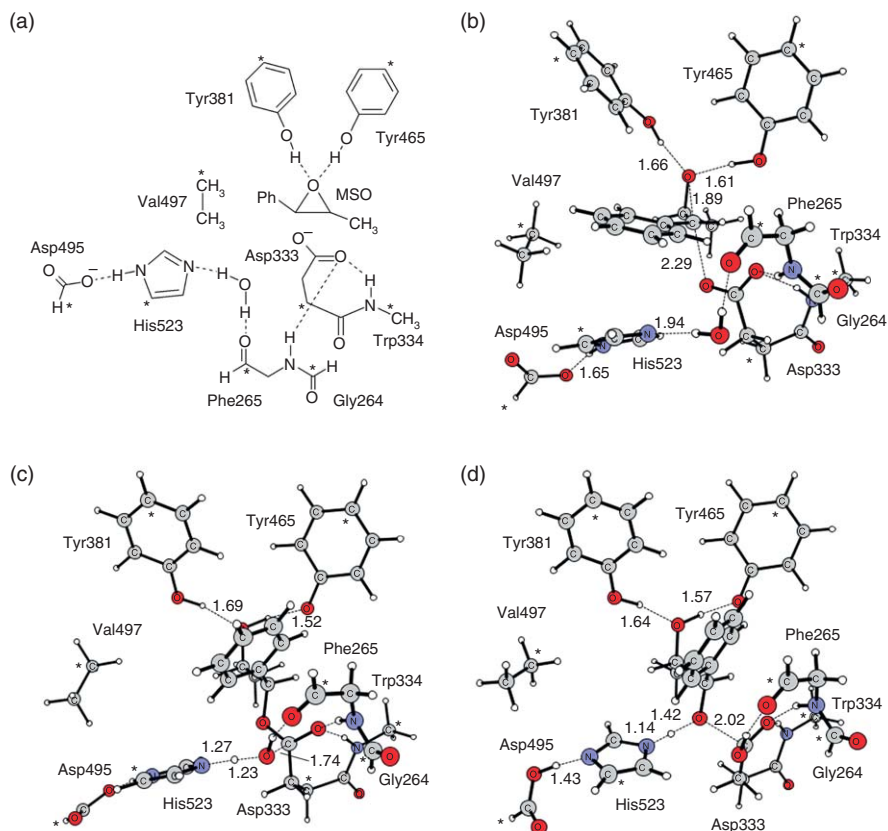
Figure 5 Optimized transition states for the hydrolysis of (1S,2S)-β-methylstyrene oxide (MSO, attack at C1) in the sEH active model. (a) Schematic representation of the model; (b) Transition state of the alkylation step; (c) Transition state for water attack; and (d) Transition state for dissociation of the enzyme–product bond.

The roles of the two active site tyrosine residues of sEH were investigated by *in silico* mutations to phenylalanine.[24] Full reaction pathways were calculated for the single mutants Y381F and Y465F whereas only the alkylation step was investigated for the double mutant Y381F/Y465F.[24] The barrier of 7.8 kcal mol^{-1} observed for Asp333 attack at C1 of MSO in the wild-type sEH model increased to 13.2 and 15.7 kcal mol^{-1} in the Y465F and Y381F models, respectively (**Table 4**). If both tyrosines are mutated, a barrier of 24.8 kcal mol^{-1} is obtained. The results indicate that the single mutants would remain catalytically active whereas the double mutant would not. This is in agreement with experimental results for the mouse sEH and the EHs from *Agrobacterium radiobacter* and potato, EchA and StEH.[73,74,79]

Table 4   Calculated barriers (kcal mol^{-1}) for MSO hydrolysis with the wild-type and tyrosine mutant models of human sEH (pathway involving attack at C1)

Model	TS alkylation	TS dealkylation[a]	TS water attack	TS dissociation	Overall barrier
Wild type	7.8	16.4	18.1	5.5	22.6
Y381F	15.7	15.7	16.1	6.2	20.9
Y465F	13.2	16.2	16.6	4.7	19.8
Y381F/Y465F	24.8	11.7	–	–	–

[a] The reverse reaction of the reversible alkylation step.
TS, transition state.

It should be noted that the effect of the single mutations appears to be overestimated in our model. For EchA, the alkylation rate of (R)-styrene oxide (RSO) changed from $1100 \, \text{s}^{-1}$ in the wild type to $60 \, \text{s}^{-1}$ in the Y215F mutant (Tyr215 in EchA is equal to Tyr465 in human sEH).[73] This corresponds to a barrier increase of about $1.8 \, \text{kcal} \, \text{mol}^{-1}$, while in our calculations, we observe an increase of $5.4 \, \text{kcal} \, \text{mol}^{-1}$ for mutation of Tyr465. It is possible that in the EchA active site, other polarization effects compensate for loss of the tyrosine hydrogen bond and thus reduce the effect of the mutation. The barriers for the subsequent steps of the reaction mechanism, that is, for attack by water on the Asp333-ester and for dissociation of the tetrahedral intermediate, changed only $1–2 \, \text{kcal} \, \text{mol}^{-1}$ compared to the wild-type model (**Table 4**). The major effect of the tyrosine residues is thus in the first step of the reaction, the alkylation.

8.20.4.2.3 Analysis of the regioselectivity of sEH

We used the sEH model to investigate the regioselectivity of epoxide opening for two substrates, MSO and (S)-styrene oxide (SSO). Regioselectivities were studied with both the wild-type and the tyrosine mutant models.[23,24] The results are summarized in **Tables 5** and **6**.

The regioselectivity of epoxide opening is in general a complex process, which is governed by multiple intrinsic and external factors.[81] Nucleophilic attack at the less substituted carbon is often preferred, due to steric effects (**Scheme 8**). However, there is an electronic advantage for attack at the more substituted carbon, since the substituent will mediate better stabilization of the partial positive charge that will develop during epoxide opening. The competition between steric and electronic factors will determine at which carbon attack occurs. The surrounding conditions can affect this competition. For example, hydrogen bonding or proton donation to the epoxide oxygen will polarize the oxygen–carbon bond of the epoxide further, thus enhancing the electronic preference for the more substituted center (**Scheme 8**).

The MSO substrate is singly substituted at each oxirane carbon, with a methyl group on C2 and a phenyl group on C1. The phenyl ring can be expected to be superior in stabilizing the evolving positive charge, and attack will thus be preferred at C1. In addition, the epoxide oxygen forms hydrogen bonds with the two active site tyrosines, Tyr381 and Tyr465, which might enhance attack at the benzylic carbon. The calculated regioselectivity of the wild-type model (defined as the barrier difference, barrier C2 − barrier C1), is $3.2 \, \text{kcal} \, \text{mol}^{-1}$, that is, significantly in favor of attack at the benzylic carbon (**Table 5**). The mutant models exhibit a slightly increased regioselectivity, $3.7 \, \text{kcal} \, \text{mol}^{-1}$ for Y465F and $4.4 \, \text{kcal} \, \text{mol}^{-1}$ for Y381F. While this seems puzzling at first, it is explainable with the concerted nature of the alkylation step in the single mutants. As

Table 5 Regioselectivity of MSO attack in different sEH active site models. Barriers (in kcal mol^{-1}) for the alkylation TS for attack at C1 or C2 of MSO

Model	Barrier C1 (benzylic)	Barrier C2 (homo-benzylic)	Regioselectivity[a]
Wild type	7.8	11.0	+3.2
Y465F	13.2	16.9	+3.7
Y381F	15.7	20.1	+4.4
Y381F/Y465F	24.8	27.7	+2.9

[a] Defined as difference in barrier (barrier C2 – barrier C1).
TS, transition state.

Table 6 Regioselectivity of SSO attack in different sEH active site models. Barriers (in kcal mol^{-1}) for the alkylation TS for attack at C1 or C2 of SSO
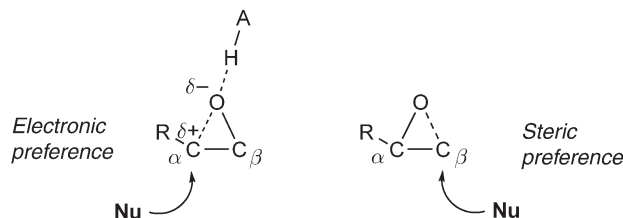
Model	Barrier C1 (benzylic)	Barrier C2 (terminal)	Regioselectivity[a]
Wild type	6.3	7.0	+0.7
Y465F	14.0	14.8	+0.8
Y381F	13.7	15.5	+1.8
Y381F/Y465F	24.2	23.6	−0.6

[a] Defined as difference in barrier (barrier C2 – barrier C1).
TS, transition state.

Scheme 8   The regioselectivity of epoxide opening is determined by competing electronic and steric factors. Interactions with a hydrogen bond- or proton-donor can enhance the electronic preference for the more substituted carbon (Nu = nucleophile).

discussed above, in the wild-type sEH model, the epoxide oxygen is hydrogen-bonded to Tyr465 and Tyr381, but alkylation does not occur in concert with proton transfer. In the single mutants, proton transfer occurs concertedly with alkylation and the interaction of the substrate with the proton is thus stronger at the transition state than it is in the wild-type model, where no proton is transferred. This will enhance polarization and increase the preference for the benzylic carbon in the single mutants. In the double mutant, Y381F/Y465F, the computed regioselectivity is 2.9 kcal mol^{-1}, which is similar to wild-type regioselectivity (**Table 5**). It can be concluded that the preference for nucleophilic attack at the benzylic carbon of MSO is intrinsic to the substrate, but that it can be enhanced in presence of the active site tyrosines.

For SSO, the results are somewhat different (**Table 6**). Removal of the methyl group from C2 reduces the steric hindrance for attack at this carbon and the barrier for opening of SSO at C2 is thus only 7.0 kcal mol^{-1}, compared to 11.0 kcal mol^{-1} for opening at C2 of MSO (**Table 5**). However, the steric advantage for attack at C2 is still lower than the electronic advantage for attack at C1 of SSO, and attack at C1 remains preferred, albeit only with 0.7 kcal mol^{-1}. The small difference in barriers suggests that a mixture of products would be formed. The single tyrosine mutants also exhibit preferred attack at C1 of SSO, while in the double mutant, steric factors dominate, resulting in preferred attack at C2 (**Table 6**).

To our knowledge, the regioselectivities of MSO and SSO openings have not been studied experimentally with human sEH. However, studies have been performed with a number of homologues. Conversion of MSO by the *Aspergillus terreus* EH leads to 95% attack at the benzylic position.[82] Conversion of the same substrate by the EH from *Rhodotorula glutinis*, the rabbit sEH or the mouse sEH yields virtually exclusive attack at the benzylic position.[83–85] These regioselectivities are very much in line with the energetics calculated for MSO.

Hydrolysis of styrene oxide with different sEH homologues has yielded a range of regioselectivities. Conversions are often performed on racemic mixtures of SSO and RSO, but in certain cases, stereoisomers are tested individually. Several EHs exhibit formation of a mixture of products for the transformation of SSO alone, as suggested from the calculated energetics. For example, the rabbit sEH exhibits a regioselectivity of 45%(C1):55%(C2) for conversion of SSO,[84] while for the *Aspergillus niger* EH, it is 15%(C1):85%(C2).[86] Exclusive attack at the benzylic carbon of SSO is mediated by the *Beauveria sulfurescens* EH[86] while EchA from *A. radiobacter* AD1 seems to mediate exclusive attack at the terminal carbon.[78] The potato *Solanum tuberosum* StEH converts styrene oxide in an enantioconvergent manner, with preferred attack at C1 for SSO (98%) and preferred attack at C2 for RSO (92%).[87] A possible explanation of the experimental variations in regioselectivity can be given based on the computed energetics. The steric preference for C2 and the electronic preference for C1 of SSO seem to be comparable in magnitude, with a computed barrier difference for attack at either center of only 0.7 kcal mol^{-1}. The results indicate that the regioselectivity of SSO can easily be influenced by interaction in the active site. Binding that reduces the steric advantage for attack at C2 could thus increase the amount of benzylic attack whereas binding that reduces the stabilization through the active site tyrosines would enhance attack at the terminal position.

8.20.4.3   Haloalcohol Dehalogenase HheC

8.20.4.3.1   *Background*

HheC is a bacterial enzyme, which was isolated from *A. radiobacter* AD1.[88] Its natural reaction is the conversion of halohydrins into epoxides (**Scheme 9**). However, HheC is also able to catalyze the reverse reaction, the

Haloalcohol   Epoxide
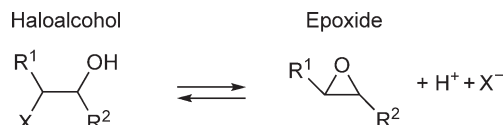
Scheme 9   Haloalcohol dehalogenase reaction.

transformation of epoxides into various substituted alcohols. Irreversible epoxide opening can occur with a number of interesting nonhalide nucleophiles, including CN^-, N_3^-, and NO_2^-.[88–94]

Structural and biochemical studies of HheC revealed that it is related to the superfamily of nicotinamide adenine dinucleotide phosphate (NAD(P)H)-dependent short-chain dehydrogenase/reductases (SDRs).[88,95] However, the HheC reaction is cofactor-independent and instead of the NAD(P)H-binding site in the SDRs, HheC exhibits a halide-binding site.[95] SDRs possess a Lys–Tyr–Ser catalytic triad and based on mutual similarities, HheC was proposed to employ an Arg145–Tyr149–Ser132 catalytic triad. Mutational studies confirmed the importance of these residues for the activity of HheC.[88,95]

HheC has the potential to become a useful biocatalyst due to its broad substrate specificity and its regio- and enantioselective properties. HheC displays high enantioselectivity for both aromatic and nonaromatic halohydrins.[94,96,97] HheC and a HheC mutant (W249F) have successfully been used in the kinetic resolution of various bromo- and chloroalcohols.[94,96,97] HheC also displays high enantioselectivity with epoxide substrates and the epoxide-opening reaction of HheC has been employed to mediate kinetic resolution of epoxides.[91,93] An interesting feature of the irreversible epoxide opening with the nonhalide nucleophiles CN^-, N_3^-, and NO_2^- is the high β-regioselectivity of HheC with various 1,2-epoxides.[90,91,93] In certain cases, the regioselectivity of the HheC reaction differs from the regioselectivity of the nonenzymatic reaction in water. For example, azidolysis of styrene oxide and p-chloro-styrene oxide in water results in, respectively, 98% and 97% attack at the benzylic α-carbon of the substrate.[93] Using HheC, this regioselectivity is changed to 79% and 89% attack at the terminal β-carbon.[93]

The putative Arg149–Tyr145–Ser132 catalytic triad in HheC is proposed to mediate a concerted dehalogenation reaction (**Scheme 10**). Tyr145 is suggested to function as the catalytic base, which abstracts a proton from the halohydrin substrate. The alcohol oxygen of the substrate then attacks the vicinal carbon, resulting in displacement of the halide ion. Arg149 is suggested to activate Tyr145. The role of Ser132 is suggested to be primarily substrate binding.[88,95] Mutational studies also indicated an important role for Asp80.[95] The carbonyl backbone of Asp80 might be important for the positioning of Arg149 (**Scheme 8**). Additionally, the Asp80 side chain might possibly function as a proton relay between Arg149 and the solvent.[95]
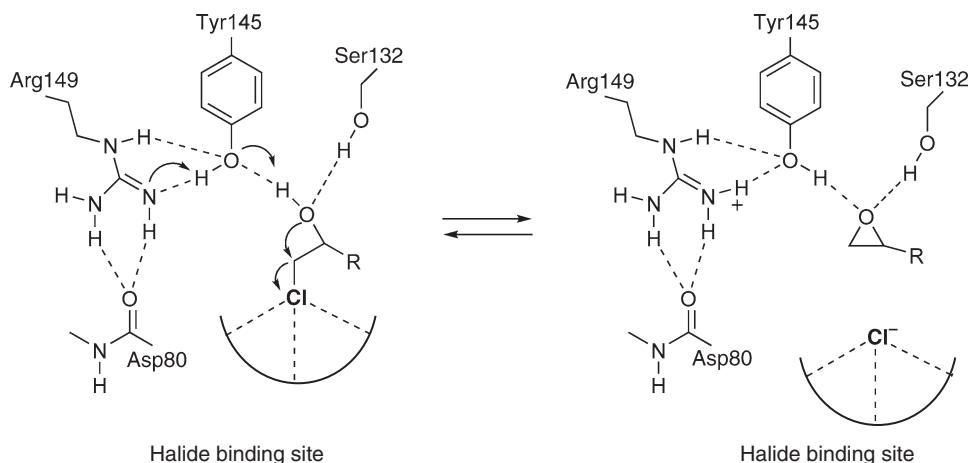
Scheme 10   Proposed dehalogenation mechanism of HheC.[95]
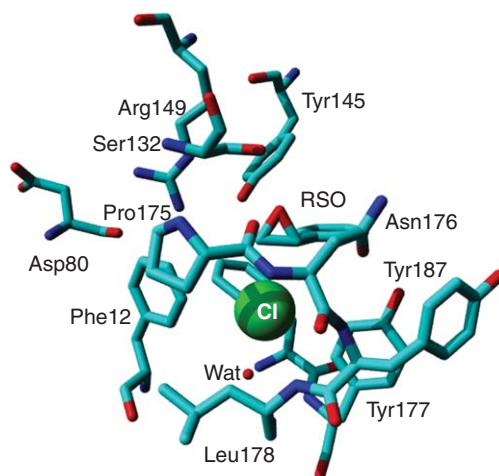
Figure 6 Active site of HheC (from the X-ray crystal structure of HheC in complex with (*R*)-styrene oxide and chloride, PDB 1PWZ).[95] Chloride is shown in ball-representation.

The halide-binding site is involved in binding of the halide side chain of the substrate and subsequently in stabilization of the displaced halide ion. The crystal structures of HheC in complex with either bromide or chloride revealed several interactions of the free halide in the halide-binding site; these include hydrogen-bonding interactions with the backbone amides of Tyr177 and Leu178, as well as with an ordered water molecule.[95] Other residues suggested to be important for formation of the halide-binding site are Phe12, Pro175, Asn176, Phe186, and Tyr187 (**Figure 6**).

8.20.4.3.2 Modeling of the dehalogenation reaction of HheC

We have studied the dehalogenation mechanism of HheC with active site models based on the crystal structure of HheC in complex with RSO and chloride (PDB 1PWZ[95], **Figure 6**).[26] RSO and chloride can be considered the product state of the HheC-mediated dehalogenation of (*R*)-2-chloro-1-phenylethanol (RCPE). We studied the conversion of RCPE using models of different size, varying from 83 to 161 atoms. All models include the proposed catalytic triad and the backbone of Asp80, which is interacting with Arg149 (**Scheme 10**). The halide-binding site was modeled differently in the three models. In the smallest model, here referred to as Model A, only the backbone amide of Leu178 and a crystallographically observed water molecule were included (**Figure 7(a)**). In Model B, the halide-binding site was extended, including parts of Pro175, Asn176, Tyr177, Leu178, Phe12, and Phe186 (**Figure 7(b)**). In Model C, the halide-binding site was modeled fully, including the side chains of the residues Pro175, Asn176, Leu178, Phe186, Tyr187, and Phe12, and the backbone part of Tyr177 (**Figure 8**).

Calculations with all three models show that the reaction proceeds as proposed through a concerted proton abstraction by Tyr145 and an intramolecular attack of the epoxide oxygen on the carbon–halide bond. Ser132 seems to be involved in stabilizing the transient oxyanion of the substrate. All three models support this reaction mechanism.

Different dielectric constants were employed to compute the energetics for the three models (**Table 7**). Relatively large solvent effects are observed for Models A and B, while the energies computed for Model C are only slightly affected when the dielectric constant is increased (**Table 7**). Model C thus provides adequate explicit solvation of the chloride ion.

Using the typical value of $\varepsilon = 4$, barriers of 14.6–17.4 kcal mol^{-1} are obtained for the three models (**Table 7**). Experimental results show that the rate-limiting step in the dehalogenation reaction is halide release, and the barrier of the chemical step should thus be equal to or below the overall barrier.[98] Dehalogenation of RCPE has an experimental rate constant of 48.5 s^{-1}, corresponding to a barrier of around 15.5 kcal mol^{-1}.[88] The models thus appear to have relatively good agreement with the experimental barrier. However, the computed reaction energies are all positive for Models B and C, indicating that epoxide opening would be more favorable than epoxide formation. Although the HheC reaction with halides is reversible, the equilibrium of HheC is toward
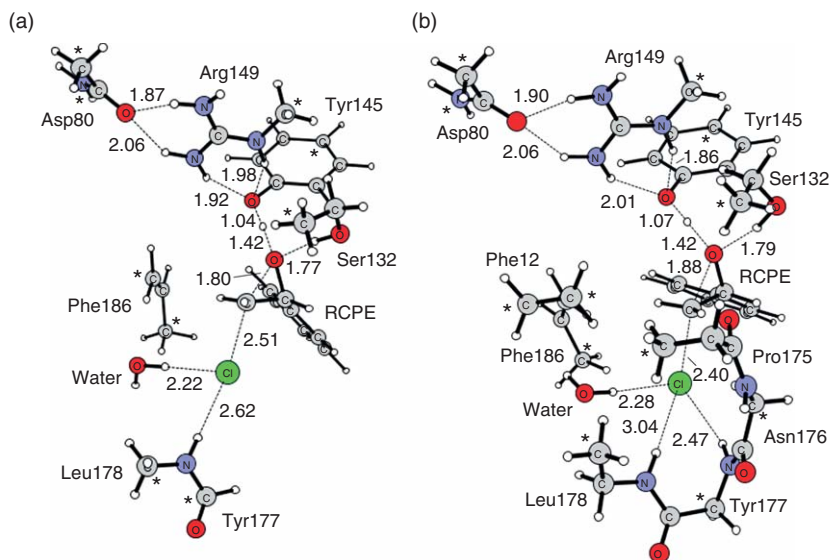
Figure 7 Optimized transition states for dehalogenation of RCPE in different HheC active site models. (a) Model A (83 atoms); (b) Model B (112 atoms).
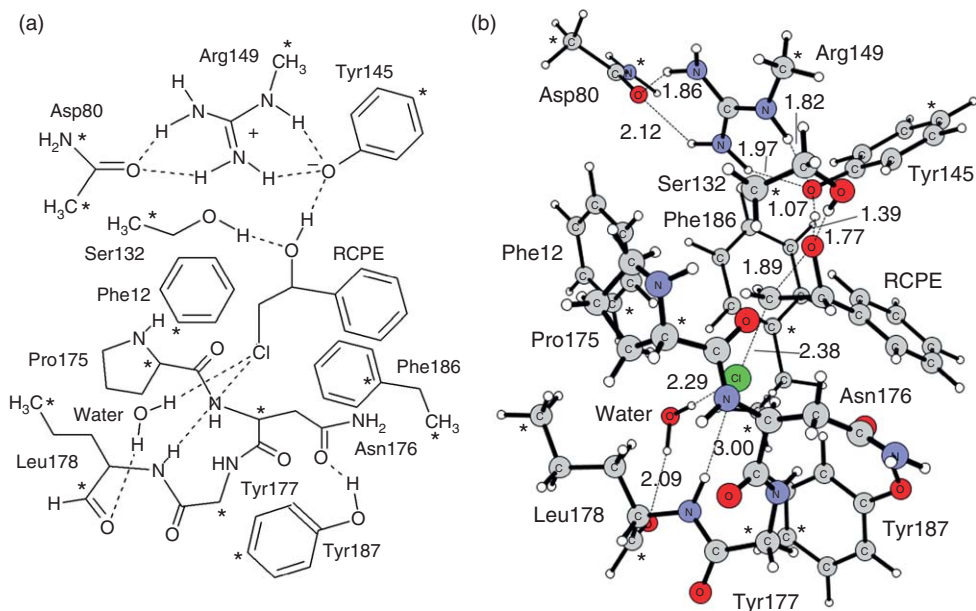
Figure 8 HheC Model C (161 atoms). (a) Schematic representation; (b) Optimized transition state for dehalogenation of RCPE.

Table 7 Summary of the calculated energetics (kcal mol^{-1}) for the HheC-catalyzed transformation of RCPE to RSO

Model		No solvation	$\varepsilon = 2$	$\varepsilon = 4$	$\varepsilon = 8$	$\varepsilon = 16$	$\varepsilon = 80$
Model A (83 atoms)	Barrier	23.0	17.8	15.0	13.5	12.8	12.2
	Reaction energy	+17.5	+8.2	+3.4	+1.0	−0.3	−1.2
Model B (112 atoms)	Barrier	17.9	15.8	14.6	14.0	13.7	13.5
	Reaction energy	+14.1	+8.3	+5.2	+3.6	+2.8	+2.2
Model C (161 atoms)	Barrier	18.2	17.7	17.4	17.1	17.0	17.0
	Reaction energy	+5.5	+4.9	+4.5	+4.3	+4.2	+4.2

epoxide formation and a slightly negative reaction energy should thus be expected.[93] Halide release to the solvent could be energetically favorable and might thus be the driving force. However, it is noteworthy that HheC has been successfully cocrystallized with RSO and chloride (**Figure 6**),[95] indicating that already the enzyme-bound product state is more favorable than the reactant state (as formation of the haloalcohol otherwise would be expected). It is possible that release of the proton from Tyr145 to the bulk is exothermic, thus reducing the reaction energy. Another possibility is that the freezing scheme employed in our calculations makes the halide-binding site too rigid, that is, the product state is not allowed to relax as much as it would in the enzyme active site, thus resulting in a positive reaction energy for Models B and C.

8.20.4.3.3   The epoxide-opening reaction of HheC

The epoxide-opening reaction of HheC can be assumed to operate in reverse fashion with respect to the dehalogenation mechanism. We studied epoxide opening of RSO with the nucleophiles cyanide and azide. The quantum chemical models were based on the X-ray crystal structure of HheC in complex with (*R*)-1-*p*-*N*-phenyl-2-azido-ethanol, (PDB code 1PXO[95]).[25] The model includes the catalytic triad, and parts of the residues comprising the halide-binding site, Phe12, Pro175, Asn176, Tyr177, Leu178, Phe186, and Tyr187. The overall size including substrate is 130 atoms for the azide model and 129 atoms for the cyanide model.

For both azidolysis and cyanolysis of RSO, we find a one-step mechanism, where attack of the nucleophile on the epoxide and proton transfer from Tyr145 occur in concert.[25] The transition state for nucleophilic attack of azide at the β-carbon of RSO is shown in **Figure 9(a)**. Attack of azide at the terminal carbon occurs at a distance of 2.16 Å. The Cβ–O bond is elongated to 1.79 Å. No proton transfer from Arg149 to Tyr145 was observed, and in the product geometry, a tyrosinate has been formed. Arg149 stabilizes the tyrosinate through hydrogen bonding. The computed barrier for the concerted proton transfer from Tyr145 and epoxide opening at Cβ is 8.1 kcal mol^{-1}. The overall reaction energy is calculated to -25.2 kcal mol^{-1}, which agrees well with the observed irreversibility of HheC-mediated azidolysis.[93]

The transition state for attack at the β-carbon of RSO by cyanide is shown in **Figure 9(c)**. The optimized distances are similar to the transition state for azidolysis, with the main difference being the nucleophile-Cβ distance, which is slightly longer (2.32 Å). The computed energies for cyanolysis are somewhat different than for azidolysis. The barrier for nucleophilic attack of cyanide at Cβ is calculated to be 7.0 kcal mol^{-1}, while the overall reaction energy is -46.0 kcal mol^{-1}. The computed energies clearly show that the proposed epoxide-opening mechanism is energetically feasible.

8.20.4.3.4   Analysis of the regioselectivity of HheC

Experimental studies show that HheC mediates attack at the less substituted carbon of various epoxides with the nucleophiles azide, cyanide, and nitrite.[89–91,93,94] For azidolysis of styrene oxide, the observed regioselectivity is different from that in solution. While chemical conversion of styrene oxide with azide results in 2% attack at the terminal carbon (Cβ), HheC-mediated azidolysis increases this value to 79%.[93] This observation made us interested in studying the regioselectivity of HheC-mediated epoxide opening and to investigate the factors controlling it.

As a reference reaction, we first studied the regioselectivities of azidolysis and cyanolysis of RSO in a small phenol-catalyzed model (**Figure 10**).[25] The calculated barriers for attack of azide at Cα and Cβ of RSO are 10.7 and 13.5 kcal mol^{-1}, respectively (**Table 8**, **Figures 10(a) and 10(b)**). The observed regioselectivity is in line with the strong preference for the Cα in the chemical azidolysis of styrene oxide.[93] With the HheC wild-type active site model of 130 atoms, the barriers for azidolysis were computed to 8.6 kcal mol^{-1} for attack at Cα (**Figure 9(b)**) and 8.1 kcal mol^{-1} for attack at Cβ (**Figure 9(a)**, **Table 8**). The regioselectivity (defined as barrier Cβ − barrier Cα) has thus changed significantly, from $+2.8$ kcal mol^{-1} in the phenol-catalyzed model to -0.5 kcal mol^{-1} in the HheC active site model. The slight preference for Cβ in the HheC active site model is in line with experimental results.[93]

For cyanolysis of RSO, we calculated a similar pattern. With the phenol-catalyzed model (**Figures 10(c) and 10(d)**), the barriers for attack at Cα and Cβ are 10.5 and 11.0 kcal mol^{-1}, respectively, that is, a slight preference for the Cα is observed. With the HheC active site model, the barriers for cyanolysis are 10.6 kcal mol^{-1} for attack at Cα (**Figure 9(d)**) and 7.0 kcal mol^{-1} for attack at Cβ (**Figure 9(c)**, **Table 8**). The regioselectivity has
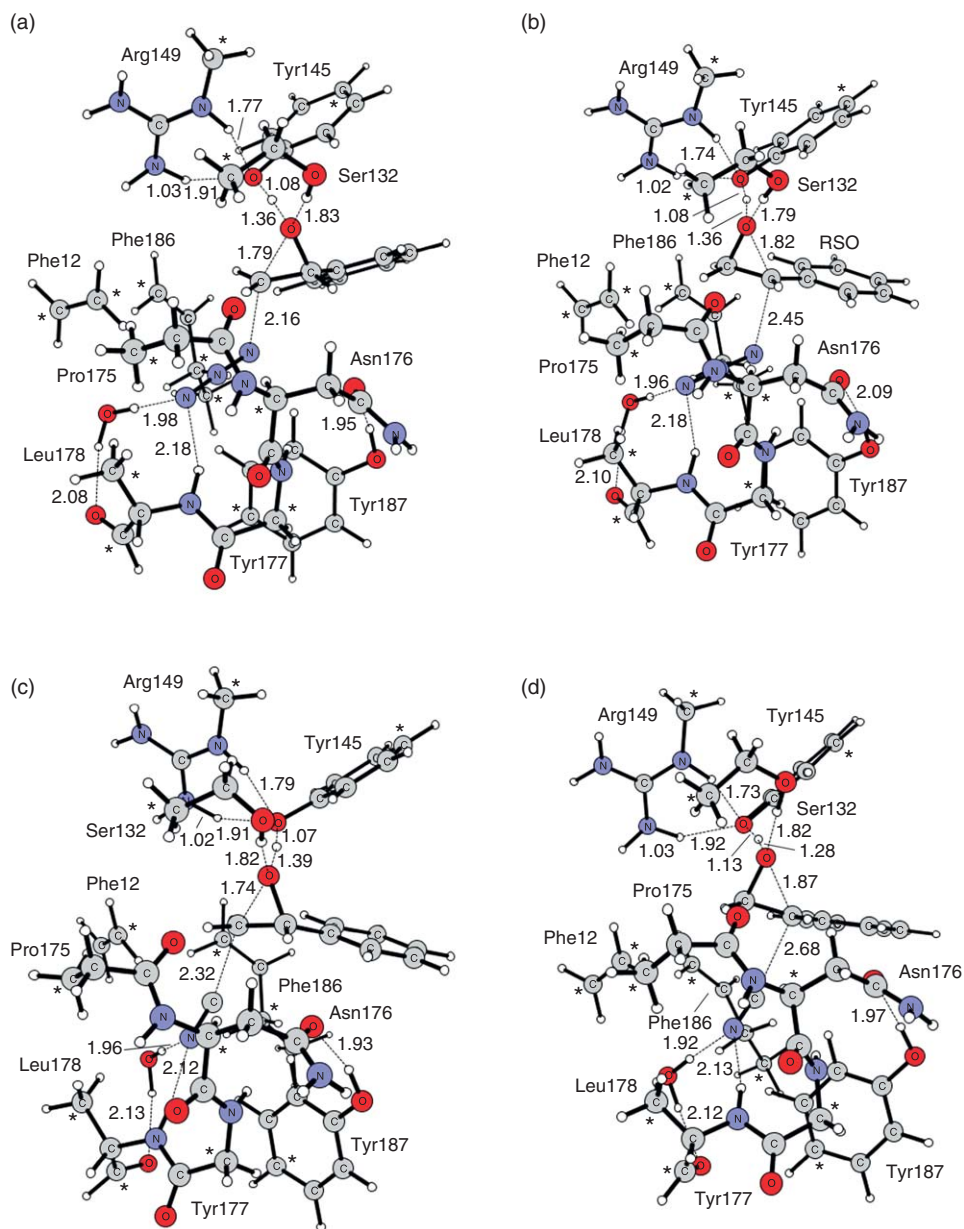
Figure 9 Optimized transition states for cyanolysis and azidolysis of RSO in the HheC active site model. (a) Attack of azide at Cβ; (b) Attack of azide at Cα; (c) Attack of cyanide at Cβ; and (d) Attack of cyanide at Cα.

thus changed from +0.5 in the phenol-catalyzed model to $-3.6 \, \text{kcal mol}^{-1}$ in the HheC active site model (**Table 8**).

The above results shows that for both the azidolysis and the cyanolysis of RSO, HheC promotes attack at Cβ with several kilocalories per mole compared to the phenol-catalyzed epoxide opening. The agreement with experimental results indicates that the HheC active site model captures the elements, which influence the enzymatic regioselectivity. Analysis of the optimized structures for the wild-type HheC model indicates that there are multiple effects that might cause the preference for attack at the β-carbon of the substrate. These include the relative positions and orientations of the nucleophile and the substrate, steric effects mediated by surrounding residues, and also electrostatic stabilization of attack at Cβ. These possible effects were analyzed by making individual *in silico* mutations of several groups (**Figure 11**).
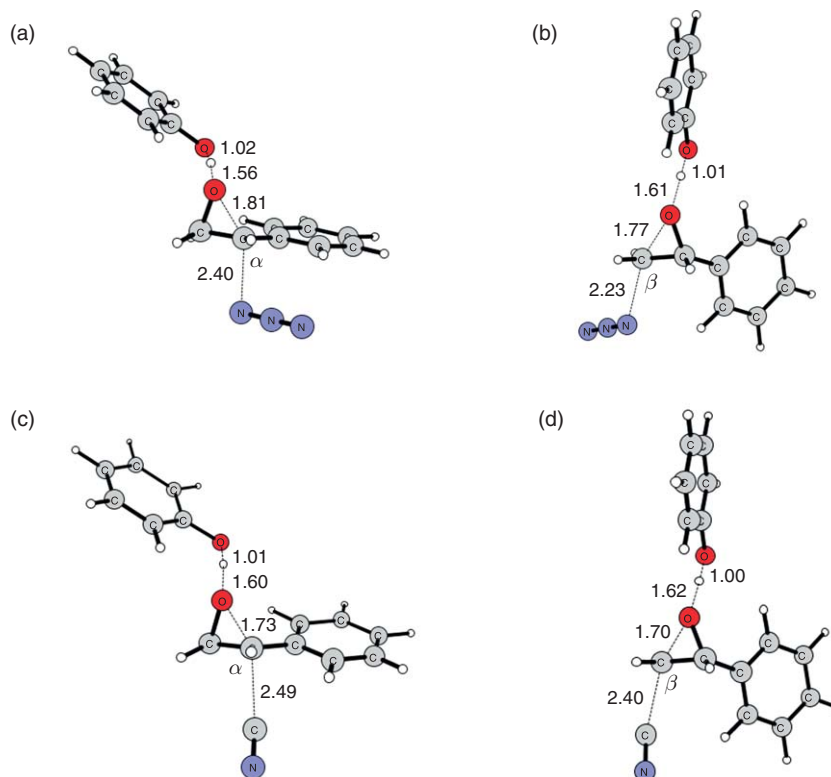
Figure 10   Phenol-catalyzed azidolysis and cyanolysis of RSO. (a) Attack of azide at Cα; (b) Attack of azide at Cβ; (c) Attack of cyanide at Cα; and (d) Attack of cyanide at Cβ.

**Table 8**   Calculated barriers (kcal mol^{-1})[a] for azidolysis and cyanolysis of RSO

Model (# atoms)	Nucleophile	Barrier Cα	Barrier Cβ	Regioselectivity[b]
Phenol-catalyzed (33)	N_3^-	10.7	13.5	+2.8
HheC wild-type (130)	N_3^-	8.6	8.1	−0.5
HheC Ser132 → Ala (129)	N_3^-	13.1	13.8	+0.7
HheC Leu178NH → O (129)	N_3^-	10.4	10.5	+0.1
HheC Pro175CO → CH$_2$ (131)	N_3^-	11.2	11.5	+0.3
HheC RmTyr187 (117)	N_3^-	9.2	11.6	+2.4
Phenol-catalyzed (32)	CN$^-$	10.5	11.0	+0.5
HheC wild-type (129)	CN$^-$	10.6	7.0	−3.6
HheC Ser132 → Ala (128)	CN$^-$	12.1	9.2	−2.9
HheC Leu178NH → O (128)	CN$^-$	9.1	6.5	−2.6
HheC Pro175CO → CH$_2$ (130)	CN$^-$	10.8	8.5	−2.3
HheC RmTyr187 (116)	CN$^-$	7.9	7.8	−0.1

[a] Solvation corrections were calculated with $\varepsilon = 80$ for phenol-catalyzed models and $\varepsilon = 4$ for HheC active site models.
[b] Defined as the difference in barriers (barrier Cβ – barrier Cα).

The effect of the hydrogen bonding to the substrate and the nucleophile were investigated by mutating Ser132 to alanine and the Leu178 backbone amide to an ester, respectively (**Figure 11**). The results indicate that these two groups have moderate effect on the regioselectivity of epoxide opening. For both cyanolysis and azidolysis, the preference for the β-carbon was reduced by 0.6–1.2 kcal mol^{-1} in the Ser132 → Ala and the Leu178NH → CO mutants, respectively (**Table 8**).

The optimized structures of the wild-type model indicated an interaction between the Pro175 backbone carbonyl and one of the hydrogens on β-carbon of the substrate. This could stabilize the evolving positive
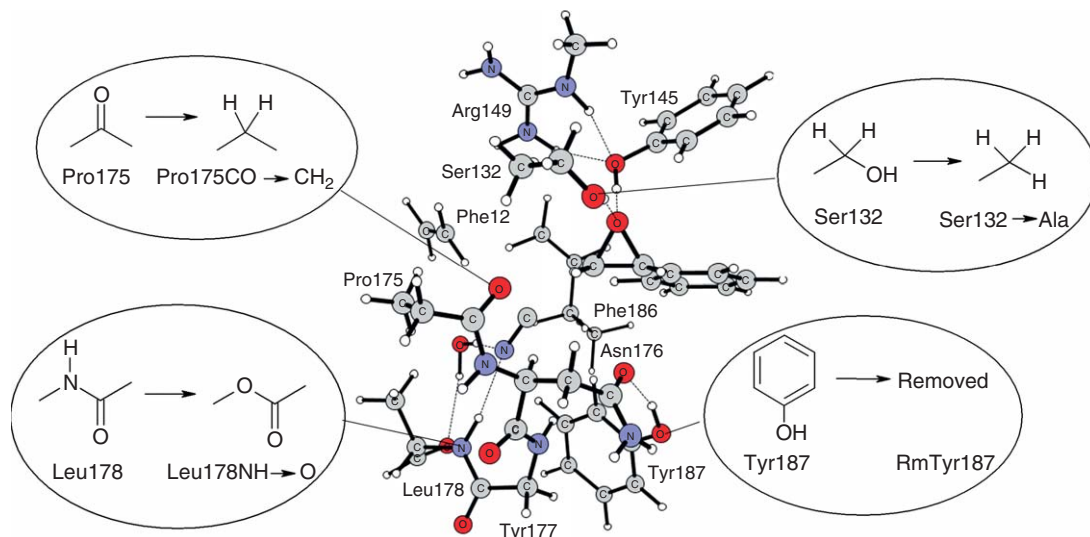
Figure 11 Individual *in silico* mutations studied in the HheC active site model.

charge at the β-carbon and might thus enhance the preference for attack at this center. To evaluate this hypothesis, a $C=O \rightarrow CH_2$ mutation of the proline backbone was performed. The results indicate that this interaction contributes by $0.8-1.3\,\text{kcal mol}^{-1}$ to the β-regioselectivity of HheC (**Table 8**).

One additional residue that was suspected to affect the regioselectivity of epoxide opening is Tyr187. The position of this residue indicated that it might impair attack of the nucleophile at the benzylic carbon of the substrate. We prepared a HheC model in which Tyr187 was removed. In the RmTyr187 model, the regioselectivity for azidolysis is $+2.4\,\text{kcal mol}^{-1}$, which is close to the $+2.8\,\text{kcal mol}^{-1}$ observed for the phenol-catalyzed model (**Table 8**). A similar result is obtained for cyanolysis, where the regioselectivity of the RmTyr187 model is $-0.1\,\text{kcal mol}^{-1}$, which is close to the $+0.5\,\text{kcal mol}^{-1}$ of the phenol-catalyzed model (**Table 8**). For both azidolysis and cyanolysis, Tyr187 is thus identified as being critical for the observed change in regioselectivity in going from the phenol-catalyzed to the HheC active site model. It can be argued that the effect of Tyr187 in the active site model is exaggerated due to the freezing scheme employed. In our model, Tyr187 is truncated and fixed at $C\gamma$ whereas in the native enzyme, the side chain would be extended and might exhibit larger flexibility. However, the interaction of Tyr187 with Asn176 (and also with Trp249) observed in the HheC crystal structure indicates that Tyr187 has little flexibility also in the native enzyme.[95,96]

The factors affecting HheC-mediated regioselectivity are summarized in **Scheme 11**.[25] First, the hydrogen-bonding pattern in the active site affects the regioselectivity by restricting the relative movement of the
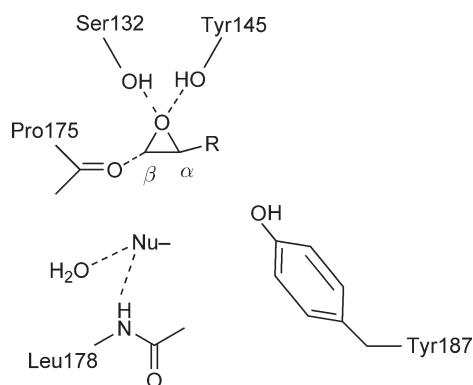
Scheme 11 Schematic illustration of effects influencing the regioselectivity of the HheC-mediated epoxide opening. These include positioning of the nucleophile and the substrate, steric hindrance caused by Tyr187, and electrostatic stabilization mediated by the backbone of Pro175.

substrate and the nucleophile. For example, removal of the hydrogen bond from the Leu178 backbone to cyanide reduces the regioselectivity of RSO cyanolysis by 1.0 kcal mol^{-1}. Also the backbone carbonyl of Pro175 affects the regioselectivity by providing electrostatic stabilization to Cβ, facilitating opening at this carbon. In our model, mutation of the Pro175 backbone reduces the regioselectivity of RSO cyanolysis by 1.3 kcal mol^{-1}. Finally, the most significant effect is exerted by the side chain of Tyr187, which imposes steric hindrance on the nucleophile, making attack at Cα less favorable. Removal of Tyr187 results in a regioselectivity that is very similar to the nonenzymatic model.

8.20.5 Conclusions and Outlook

In this chapter, we have discussed how enzyme active sites and reaction mechanisms can be studied using quantum chemical models. The usefulness of this approach has been validated for a large number of enzymes during the last decade. Here, we have presented three recent examples concerned with the reactivity of epoxides in different enzymes.

The reaction mechanisms and regioselectivities of sEH, LEH, and HheC were analyzed in detail. The computational results support the proposed mechanisms and are able to explain the experimentally observed regioselectivities of these enzymes. The quantum chemical approach also allowed us to identify and quantify the importance of individual functional groups for the reaction mechanism and the regioselectivity of epoxide opening.

Accurate DFT calculations can today routinely handle more than 100 atoms. This has paved the way for a better description of enzyme active sites, and for wider applications. For example, as we have seen in this chapter, it is now possible to perform *in silico* mutations to gain deeper insight into the roles of various groups at the active site. This development is going to continue over the coming years. As computers become faster and cheaper, larger and larger models will be able to be treated. Efficient code parallelization and linear scaling methods will speed up the calculations further. Active site models will thus become more and more realistic. It should be stressed here that much large models are not necessarily more accurate. They can, for example, suffer from multiple minima problems and sampling will be required.

Abbreviations

CIU	cyclohexyl-*N'*-(iodophenyl) urea
CPCM	conductor-like polarizable continuum model
DFT	density functional theory
EH	epoxide hydrolase
HF	Hartree–Fock
MSO	(1*S*,2*S*)-β-methylstyrene oxide
NAD(P)H	nicotinamide adenine dinucleotide phosphate
PCM	polarizable continuum model
PDB	Protein Data Bank
PFL	pyruvate-formate lyase
QM/MM	quantum mechanics/molecular mechanics
RCPE	(*R*)-2-chloro-1-phenylethanol
RSO	(*R*)-styrene oxide
SDR	short-chain dehydrogenase/reductase
sEH	soluble epoxide hydrolase
SPL	spore-photoproduct lyase
SSO	(*S*)-styrene oxide
TS	transition state
TST	transition state theory

Nomenclature

Ala   alanine
Arg   arginine
Asn   asparagine
Asp   aspartate
EchA   epoxide hydrolase from *Agrobacterium radiobacter* AD1
Gly   glycine
HheC   haloalcohol dehalogenase C from *Agrobacterium radiobacter* AD1
His   histidine
LEH   limonene epoxide hydrolase from *Rhodococcus erythropolis* DCL14
Leu   leucine
MP2   second-order Møller–Plesset perturbation theory
Phe   phenylalanine
Pro   proline
Ser   serine
StEH   epoxide hydrolase from potato (*Solanum tuberosum*)
Trp   tryptophan
Tyr   tyrosine
Val   valine

References

1. P. E. M. Siegbahn; M. R. A. Blomberg, *J. Phys. Chem. B* **2001**, *105*, 9375.
2. P. E. M. Siegbahn, *Q. Rev. Biophys.* **2003**, *36*, 91.
3. P. E. M. Siegbahn; T. Borowski, *Acc. Chem. Res.* **2006**, *39*, 729.
4. F. Himo; P. E. M. Siegbahn, *Chem. Rev.* **2003**, *103*, 2421.
5. L. Noodleman; T. Lovell; W.-G. Han; J. Li; F. Himo, *Chem. Rev.* **2004**, *104*, 459.
6. F. Himo, *Theor. Chem. Acc.* **2006**, *116*, 232.
7. F. Himo, *Biochim. Biophys. Acta – Bioenerg.* **2005**, *1707*, 24.
8. M. Leopoldini; N. Russo; M. Toscano, *Chem. Eur. J.* **2007**, *13*, 2109.
9. M. Leopoldini; N. Russo; M. Toscano; M. Dulak; T. A. Wesolowski, *Chem. Eur. J.* **2006**, *12*, 2532.
10. T. Marino; N. Russo; M. Toscano, *J. Am. Chem. Soc.* **2005**, *127*, 4242.
11. S. F. Sousa; P. A. Fernandes; M. J. Ramos, *J. Phys. Chem. B* **2007**, *111*, 9146.
12. S. F. Sousa; P. A. Fernandes; M. J. Ramos, *J. Am. Chem. Soc.* **2007**, *29*, 1378.
13. N. M. F. S.A. Cerqueira; P. A. Fernandes; M. J. Ramos, *J. Phys. Chem. B* **2006**, *110*, 21272.
14. N. M. F. S.A. Cerqueira; P. A. Fernandes; L. A. Eriksson; M. J. Ramos, *Biophys. J.* **2006**, *90*, 2109.
15. H. N. Liu; J. J. Robinet; S. Ananvoranich; J. W. Gauld, *J. Phys. Chem. B* **2007**, *111*, 439.
16. K.-B. Cho; J. W. Gauld, *J. Phys. Chem. B.* **2005**, *109*, 23706.
17. F. Himo; J.-D. Guo; A. Rinaldo-Matthis; P. Nordlund, *J. Phys. Chem. B* **2005**, *109*, 20004.
18. P. Velichkova; F. Himo, *J. Phys. Chem. B* **2005**, *109*, 8216.
19. P. Velichkova; F. Himo, *J. Phys. Chem. B* **2006**, *110*, 16.
20. R. Sevastik; F. Himo, *Bioorg. Chem.* **2007**, *35*, 444.
21. K. H. Hopmann; J.-D. Guo; F. Himo, *Inorg. Chem.* **2007**, *46*, 4850.
22. K. H. Hopmann; B. M. Hallberg; F. Himo, *J. Am. Chem. Soc.* **2005**, *127*, 14339–14347.
23. K. H. Hopmann; F. Himo, *Chem. Eur. J.* **2006**, *12*, 6898–6909.
24. K. H. Hopmann; F. Himo, *J. Phys. Chem. B.* **2006**, *110*, 21299–21310.
25. K. H. Hopmann; F. Himo, *Biochemistry* **2008**, *47*, 4973–4982.
26. K. H. Hopmann; F. Himo, *J. Chem. Theory Comp.* **2008**, *4*, 1129–1137.
27. P. Hohenberg; W. Kohn, *Phys. Rev.* **1964**, *136*, B864–B871.
28. W. Kohn; L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133–A1138.
29. A. D. Becke, *J. Chem. Phys.* **1992**, *96*, 2155.
30. A. D. Becke, *J. Chem. Phys.* **1992**, *97*, 9173.
31. A. D. Becke, *Phys. Rev. A* **1988**, *38*, 3098–3100.
32. C. Lee; W. Yang; R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
33. S. H. Vosko; L. Wilk; M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200–1211.
34. A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
35. C. W. Bauschlicher, Jr., *Chem. Phys. Lett.* **1995**, *246*, 40–44.

36. C. W. Bauschlicher, jr., H. Patridge, *Chem. Phys. Lett.* **1995**, *240*, 533–540.
37. L. A. Curtiss; C. A. Redfern; K. Raghavachari, *J. Chem. Phys.* **2005**, *123*, 124107.
38. J. K. Kang; C. B. Musgrave, *J. Chem. Phys.* **2001**, *115*, 11040–11051.
39. Y. Zhao; N. González-García; D. G. Truhlar, *J. Phys. Chem. A* **2005**, *109*, 2012–2018.
40. V. Guner; K. S. Khuong; A. G. Leach; P. S. Lee; M. D. Bartberger; K. N. Houk, *J. Phys. Chem. A.* **2003**, *107*, 11445–11459.
41. P. E. M. Siegbahn, *J. Biol. Inorg. Chem.* **2006**, *11*, 695–701.
42. F. Himo; L. A. Eriksson, *J. Am. Chem. Soc.* **1998**, *120*, 11449–11455.
43. J. D. Guo; Y. Luo; F. Himo, *J. Phys. Chem. B* **2003**, *107*, 11188–11192.
44. G. A. Gomez; C. Morisseau; B. D. Hammock; D. W. Christianson, *Biochemistry* **2004**, *43*, 4716–4723.
45. A. Klamt; G. Schüürmann, *J. Chem. Soc. Perkin. Trans.* **1993**, *2*, 799–805.
46. J. Andzelm; C. Kölmel; A. Klamt, *J. Chem. Phys.* **1995**, *103*, 9312–9320.
47. V. Barone; M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
48. M. Cossi; N. Rega; G. Scalmani; V. Barone, *J. Comput. Chem.* **2003**, *24*, 669–681.
49. F. Barbirato; J. C. Verdoes; J. A. M. de Bont; M. J. van der Werf, *FEBS Lett.* **1998**, *438*, 293–296.
50. M. J. van der Werf; R. V. A. Orru; K. M. Overkamp; H. J. Swarts; I. Osprian; A. Steinreiber; J. A. M. de Bont; K. Faber, *Appl. Microbiol. Biotechnol.* **1999**, *52*, 380–385.
51. M. J. van der Werf; K. M. Overkamp; J. A. M. Bont, *J. Bacteriol.* **1998**, *180*, 5052–5057.
52. M. van der Werf; H. Swarts; J. A. M. de Bont, *Appl. Environ. Microbiol.* **1999**, *65*, 2092–2102.
53. M. Arand; B. M. Hallberg; J. Zou; T. Bergfors; F. Oesch; M. van der Werf; J. A. M. de Bont; T. A. Jones; S. L. Mowbray, *EMBO J.* **2003**, *22*, 2583–2592.
54. P. Johansson; T. Unge; A. Cronin; M. Arand; T. Bergfors; T. Jones; S. L. Mowbray, *J. Mol. Biol.* **2005**, *351*, 1048–1056.
55. A. R. Gallimore; C. B. Stark; A. Bhatt; B. M. Harvey; Y. Demydchuk; V. Bolanos-Garcia; D. J. Fowler; J. Staunton; P. F. Leadlay; J. B. Spencer, *Chem. Biol.* **2006**, *13*, 453–460.
56. M. van der Werf; J. A. M. de Bont; H. J. Swarts, *Tetrahedron: Asymmetry* **1999**, *10*, 4225–4230.
57. A. J. Fretland; C. J. Omiecinski, *Chem. Biol. Interact.* **2000**, *129*, 41–59.
58. E. C. Dietze; J. Magdalou; B. D. Hammock, *Int. J. Biochem.* **1990**, *22*, 461–470.
59. D. L. Ollis; E. Cheah; M. Cygler; B. Dijkstra; F. Frolow; S. M. Franken; M. Harel; S. J. Remington; I. Silman; J. Schrag; J. L. Sussman; K. H. G. Verschueren; A. Goldman, *Protein Eng.* **1992**, *5*, 197–211.
60. B. Borhan; A. D. Jones; F. Pinot; D. F. Grant; M. J. Kurth; B. D. Hammock, *J. Biol. Chem.* **1995**, *270*, 26923–26930.
61. F. Müller; M. Arand; H. Frank; A. Seidel; W. Hinz; L. Winkler; K. Hänel; E. Blée; J. K. Beetham; B. D. Hammock; F. Oesch, *Eur. J. Biochem.* **1997**, *245*, 490–496.
62. B. D. Hammock; F. Pinot; J. K. Beetham; D. F. Grant; M. E. Arand; F. Oesch, *Biochem. Biophys. Res. Commun.* **1994**, *3*, 850–856.
63. G. M. Lacouciere; R. N. Armstrong; *J. Am. Chem. Soc.* **1993**, *115*, 10466–10467.
64. L. T. Laughlin; H.-F. Tzeng; S. Lin; R. N. Armstrong, *Biochemistry* **1998**, *37*, 2897–2904.
65. M. Arand; H. Wagner; F. Oesch, *J. Biol. Chem.* **1996**, *271*, 4223–4229.
66. F. Pinot; D. F. Grant; J. K. Beetham; A. G. Parker; B. Borhan; S. Landt; A. D. Jones; B. D. Hammock, *J. Biol. Chem.* **1995**, *270*, 7968–7974.
67. R. Rink; F. Fennema; M. Smids; U. Dehmel; D. B. Janssen, *J. Biol. Chem.* **1997**, *272*, 14650–14657.
68. H.-F. Tzeng; L. T. Laughlin; R. N. Armstrong, *Biochemistry* **1998**, *37*, 2905–2911.
69. P. A. Bell; C. B. Kasper, *J. Biol. Chem.* **1993**, *268*, 14011–14017.
70. M. Nardini; I. S. Ridder; H. J. Rozeboom; K. H. Kalk; R. Rink; D. B. Janssen; B. W. Dijsktra, *J. Biol. Chem.* **1999**, *274*, 14579–14586.
71. M. A. Argiriadi; C. M. Morisseau; M. H. Goodrow; D. L. Dowdy; B. D. Hammock; D. W. Christianson, *J. Biol. Chem.* **2000**, *275*, 15265–15270.
72. J. Zou; B. M. Hallberg; T. Bergfors; F. Oesch; M. Arand; S. L. Mowbray; T. A. Jones, *Structure* **2000**, *8*, 111–122.
73. R. Rink; J. Kingma; J. H. Lutje Spelberg; D. B. Janssen, *Biochemistry* **2000**, *39*, 5600–5613.
74. T. Yamada; C. Morisseau; J. E. Maxwell; M. A. Argiriadi; D. W. Christianson; B. D. Hammock, *J. Biol. Chem.* **2000**, *275*, 23082–23088.
75. S. L. Mowbray; L. T. Elfström; K. M. Ahlgren; C. E. Andersson; M. Widersten, *Protein Sci.* **2006**, *15*, 1628–1637.
76. G. M. Lacouciere; R. N. Armstrong, *Chem. Res. Toxicol.* **1994**, *7*, 121–124.
77. L. T. Elfström; M. Widersten, *Biochemistry* **2006**, *45*, 205–212.
78. R. Rink; D. B. Janssen, *Biochemistry* **1998**, *37*, 18119–18127.
79. L. T. Elfström; M. Widersten, *Biochem. J.* **2005**, *390*, 633–640.
80. B. Schiøtt; T. C. Bruice, *J. Am. Chem. Soc.* **2002**, *124*, 14558–14570.
81. J. March, *Advanced Organic Chemistry: Reactions, Mechanism and Structure*, 4th ed.; John Wiley & Sons: New York, 1992.
82. P. Moussou; A. Archelas; J. Baratti; R. Furstoss, *Tetrahedron: Asymmetry.* **1998**, *9*, 1539–1547.
83. C. A. G. M. Weijers; J. A. M. de Bont, *J. Mol. Catal. B: Enzym.* **1999**, *6*, 199–214.
84. G. Bellucci; C. Chiappe; A. Cordoni; F. Marioni, *Tetrahedron Lett.* **1994**, *35*, 4219–4222.
85. K. C. Williamson; C. Morisseau; J. E. Maxwell; B. D. Hammock, *Tetrahedron: Asymmetry* **2000**, *11*, 4451–4462.
86. S. Pedragosa-Moreau; A. Archelas; R. Furstoss, *Bioorg. Med. Chem.* **1994**, *2*, 609–616.
87. M. I. Monterde; M. Lombard; A. Archelas; A. Cronin; M. Arand; R. Furstoss, *Tetrahedron: Asymmetry* **2004**, *15*, 2801–2805.
88. J. E. T. van Hylckama Vlieg; L. Tang; J. H. Lutje Spelberg; T. Smilda; G. J. Poelarends; T. Bosma; A. E. van Merode; M. W. Fraaije; D. B. Janssen, *J. Bacteriol.* **2001**, *183*, 5058–5066.
89. D. B. Janssen; M. Majerić-Elenkov; G. Hasnaoui; B. Hauer; J. H. Lutje Spelberg, *Biochem. Soc. Trans.* **2006**, *34*, 291–295.
90. M. Majerić-Elenkov; B. Hauer; D. B. Janssen, *Adv. Synth. Catal.* **2006**, *348*, 579–585.
91. G. Hasnaoui; J. H. Lutje Spelberg; E. de Vries; L. Tang; B. Hauer; D. B. Janssen, *Tetrahedron: Asymmetry* **2005**, *16*, 1685–1692.
92. T. Nakamura; T. Nagasawa; F. Yu; I. Watanabe; H. Yamada, *Biochem. Biophys. Res. Commun.* **1991**, *180*, 124–130.
93. J. H. Lutje Spelberg; J. E. T. van Hylckama Vlieg; L. Tang; D. B. Janssen; R. M. Kellogg, *Org. Lett.* **2001**, *3*, 41–43.
94. M. Majerić-Elenkov; L. Tang; B. Hauer; D. B. Janssen, *Org. Lett.* **2006**, *8*, 4227–4229.

95. R. M. de Jong; J. J. Tiesinga; H. J. Rozeboom; K. H. Kalk; L. Tang; D. B. Janssen; B. W. Dijkstra, *EMBO J.* **2003**, *22*, 4933–4944.
96. L. Tang; D. E. T. Pazmiño; M. W. Fraaije; R. M. de Jong; B. W. Dijkstra; D. B. Janssen, *Biochemistry* **2005**, *44*, 6609–6618.
97. R. M. Haak; C. Tarabiono; D. B. Janssen; A. J. Minnaard; J. G. de Vries; B. L. Feringa, *Org. Biomol. Chem.* **2007**, *5*, 318–323.
98. L. Tang; J. H. Lutje Spelberg; M. W. Fraaije; D. B. Janssen, *Biochemistry* **2003**, *42*, 5378–5386.

Biographical Sketches

Kathrin H. Hopmann was born in Stadt Lahn, Germany in 1979. She completed her undergraduate studies in chemistry and molecular biology at the University of Aarhus, Denmark. In 2008, she obtained her Ph.D. degree from the Royal Institute of Technology in Stockholm, Sweden. Her graduate work focused on computational enzymatic catalysis and was conducted under the supervision of Dr. Fahmi Himo. Currently, she is working as a postdoctoral fellow at the Centre for Theoretical and Computational Chemistry at the University of Tromsoe, Norway. Her research activities are in the area of bioinorganic chemistry with the main focus on electronic structure and reactivity of transition metal complexes.

Fahmi Himo was born in 1973. He did his undergraduate studies in physics at Stockholm University, where he also received his Ph.D. degree in 2000 (with Leif A. Eriksson and Per E. M. Siegbahn). He then spent two years as a Wenner-Gren postdoctoral fellow at the Scripps Research Institute in La Jolla, California, working with Louis Noodleman, David Case, and Barry Sharpless. He is currently working as an associate professor at the Royal Institute of Technology in Stockholm, Sweden. The main focus of his research is catalysis. Quantum chemical methods are used to study mechanisms of enzymatic and organic/organometallic reactions.