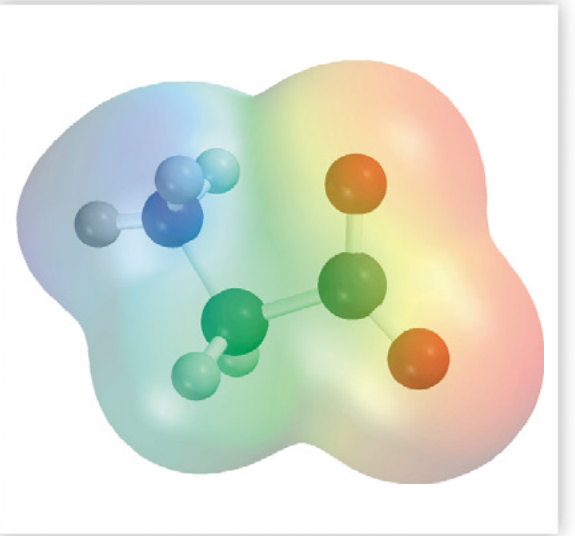
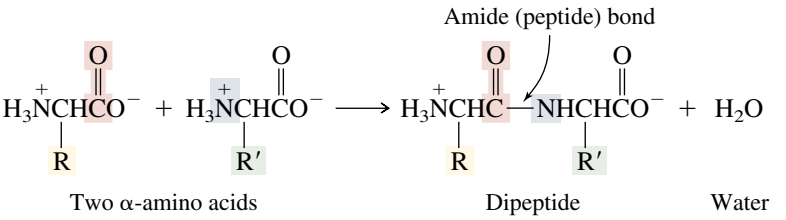


CHAPTER 27

AMINO ACIDS, PEPTIDES, AND PROTEINS. NUCLEIC ACIDS

The relationship between structure and function reaches its ultimate expression in the chemistry of amino acids, peptides, and proteins.

Amino acids are carboxylic acids that contain an amine function. Under certain conditions the amine group of one molecule and the carboxyl group of a second can react, uniting the two amino acids by an amide bond.



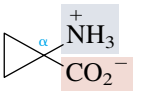
Amide linkages between amino acids are known as **peptide bonds,** and the product of peptide bond formation between two amino acids is called a **dipeptide.** The peptide chain may be extended to incorporate three amino acids in a **tripeptide,** four in a **tetrapeptide,** and so on. **Polypeptides** contain many amino acid units. **Proteins** are naturally occurring polypeptides that contain more than 50 amino acid units—most proteins are polymers of 100–300 amino acids.

The most striking thing about proteins is the diversity of their roles in living systems: silk, hair, skin, muscle, and connective tissue are proteins, and almost all enzymes are proteins. As in most aspects of chemistry and biochemistry, structure is the key to function. We'll explore the structure of proteins by first concentrating on their fundamental building block units, the α-amino acids. Then, after developing the principles of peptide structure, we'll see how the insights gained from these smaller molecules aid our understanding of proteins.

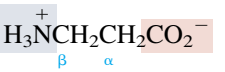
The chapter concludes with a discussion of the **nucleic acids,** which are the genetic material of living systems and which direct the biosynthesis of proteins. These two types of biopolymers, nucleic acids and proteins, are the organic chemicals of life.

27.1 CLASSIFICATION OF AMINO ACIDS

Amino acids are classified as α, β, γ, and so on, according to the location of the amine group on the carbon chain that contains the carboxylic acid function.



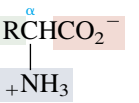
1-Aminocyclopropanecarboxylic acid: an α-amino acid that is the biological precursor to ethylene in plants



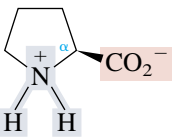
3-Aminopropanoic acid: known as β-alanine, it is a β-amino acid that makes up one of the structural units of coenzyme A

4-Aminobutanoic acid: known as γ-aminobutyric acid (GABA), it is a γ-amino acid and is involved in the transmission of nerve impulses

Although more than 700 different amino acids are known to occur naturally, a group of 20 of them commands special attention. These 20 are the amino acids that are normally present in proteins and are shown in Figure 27.1 and in Table 27.1. All the amino acids from which proteins are derived are α-amino acids, and all but one of these contain a primary amino function and conform to the general structure



The one exception is proline, a secondary amine in which the amino nitrogen is incorporated into a five-membered ring.



Proline

Table 27.1 includes three-letter and one-letter abbreviations for the amino acids. Both enjoy wide use.

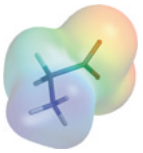
Our bodies can make some of the amino acids shown in the table. The others, which are called **essential amino acids,** we have to get from what we eat.

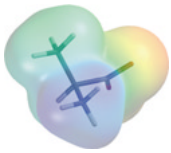
27.2 STEREOCHEMISTRY OF AMINO ACIDS

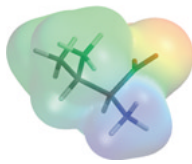
The graphic that opened this chapter is an electrostatic potential map of glycine.

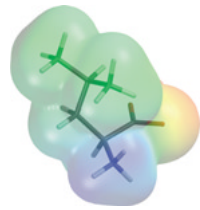
Glycine is the simplest amino acid and the only one in Table 27.1 that is achiral. The α-carbon atom is a stereogenic center in all the others. Configurations in amino acids are normally specified by the D, L notational system. All the chiral amino acids obtained from proteins have the L configuration at their α-carbon atom.

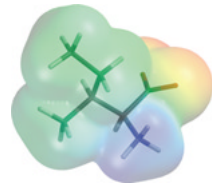
Amino acids with nonpolar side chains

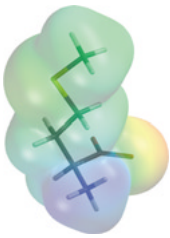

Glycine

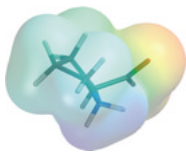

Alanine

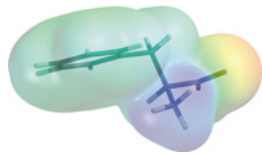

Valine

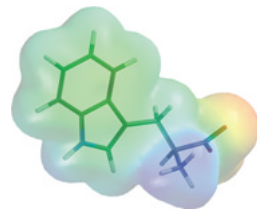

Leucine


Isoleucine

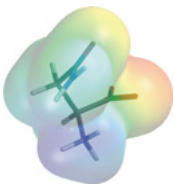

Methionine

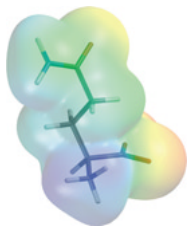

Proline


Phenylalanine

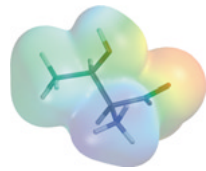

Tryptophan

Amino acids with polar but nonionized side chains

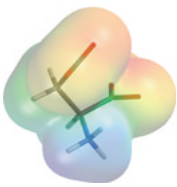

Asparagine

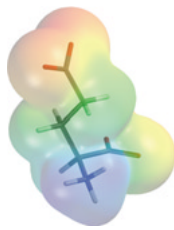

Glutamine

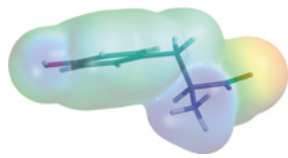

Serine

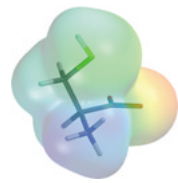

Threonine

Amino acids with acidic side chains

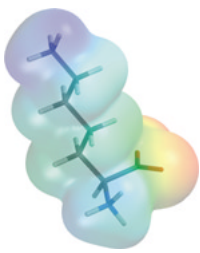

Aspartic acid

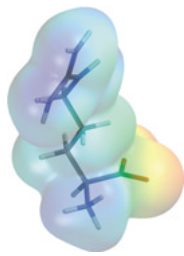

Glutamic acid


Tyrosine


Cysteine

Amino acids with basic side chains


Lysine


Arginine

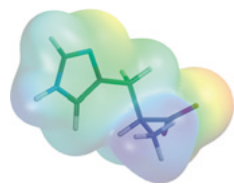
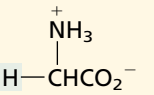
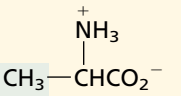
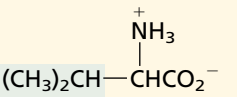
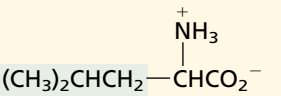
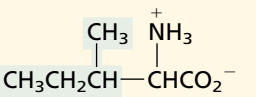
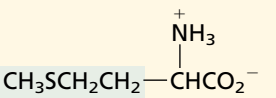
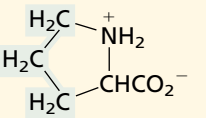
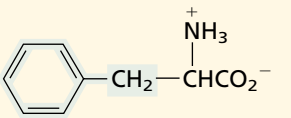
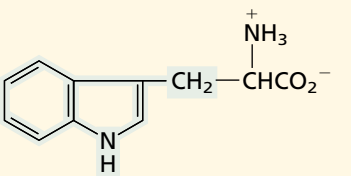
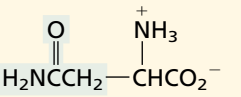

Histidine

FIGURE 27.1 Electrostatic potential maps of the 20 common amino acids listed in Table 27.1. Each amino acid is oriented so that its side chain is in the upper left corner. The side chains affect the shape and properties of the amino acids.

Learning By Modeling contains electrostatic potential maps of all the amino acids in this table.

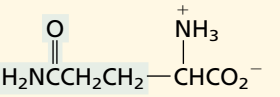
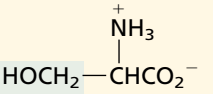
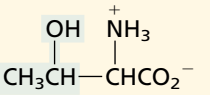
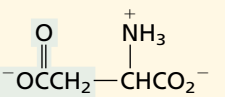
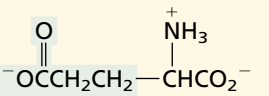
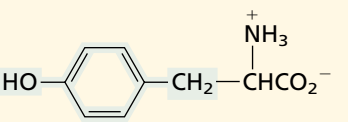
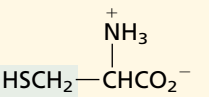
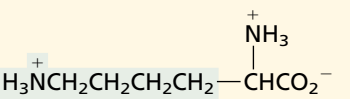
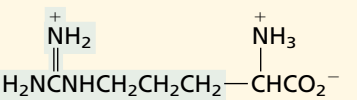
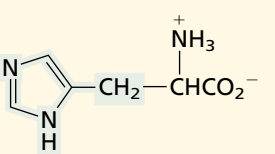
TABLE 27.1 α-Amino Acids Found in Proteins

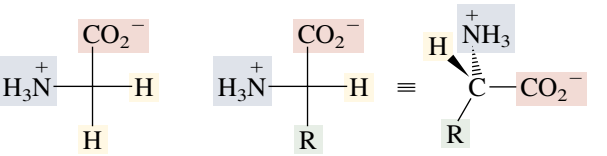
Name	Abbreviation	Structural formula*
Amino acids with nonpolar side chains		
Glycine	Gly (G)	
Alanine	Ala (A)	
Valine†	Val (V)	
Leucine†	Leu (L)	
Isoleucine†	Ile (I)	
Methionine†	Met (M)	
Proline	Pro (P)	
Phenylalanine†	Phe (F)	
Tryptophan†	Trp (W)	
Amino acids with polar but nonionized side chains		
Asparagine	Asn (N)	

*All amino acids are shown in the form present in greatest concentration at pH 7.
†An essential amino acid, which must be present in the diet of animals to ensure normal growth.

(Continued)

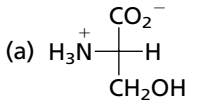
TABLE 27.1 α-Amino Acids Found in Proteins ***(Continued)***

Name	Abbreviation	Structural formula*
Amino acids with polar but nonionized side chains		
Glutamine	Gln (Q)	
Serine	Ser (S)	
Threonine†	Thr (T)	
Amino acids with acidic side chains		
Aspartic acid	Asp (D)	
Glutamic acid	Glu (E)	
Tyrosine	Tyr (Y)	
Cysteine	Cys (C)	
Amino acids with basic side chains		
Lysine†	Lys (K)	
Arginine†	Arg (R)	
Histidine†	His (H)	

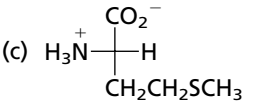


Glycine (achiral) | Fischer projection of an L-amino acid

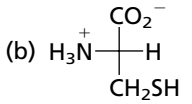
PROBLEM 27.1 What is the absolute configuration (*R* or *S*) at the α carbon atom in each of the following L-amino acids?

(a) 

L-Serine

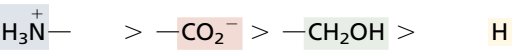
(c) 

L-Methionine

(b) 

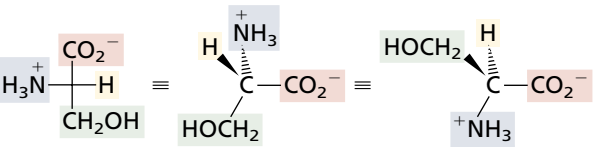
L-Cysteine

SAMPLE SOLUTION (a) First identify the four groups attached directly to the stereogenic center, and rank them in order of decreasing sequence rule precedence. For L-serine these groups are

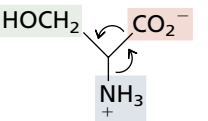


Highest ranked — Lowest ranked

Next, translate the Fischer projection of L-serine to a three-dimensional representation, and orient it so that the lowest ranked substituent at the stereogenic center is directed away from you.



In order of decreasing precedence the three highest ranked groups trace an anticlockwise path.



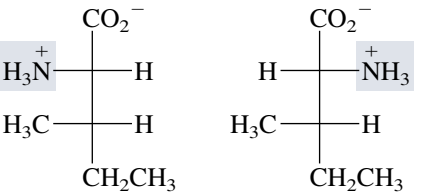
The absolute configuration of L-serine is *S*.

PROBLEM 27.2 Which of the amino acids in Table 27.1 have more than one stereogenic center?

Although all the chiral amino acids obtained from proteins have the L configuration at their α carbon, that should not be taken to mean that D-amino acids are unknown. In fact, quite a number of D-amino acids occur naturally. D-Alanine, for example, is a

constituent of bacterial cell walls. The point is that D-amino acids are not constituents of proteins.

A new technique for dating archaeological samples called *amino acid racemization* (AAR) is based on the stereochemistry of amino acids. Over time, the configuration at the α-carbon atom of a protein's amino acids is lost in a reaction that follows first-order kinetics. When the α carbon is the only stereogenic center, this process corresponds to racemization. For an amino acid with two stereogenic centers, changing the configuration of the α carbon from L to D gives a diastereomer. In the case of isoleucine, for example, the diastereomer is an amino acid not normally present in proteins, called *alloisoleucine.*



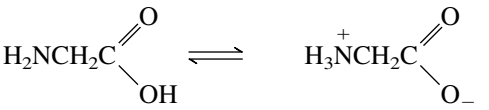
L-Isoleucine D-Alloisoleucine

By measuring the L-isoleucine/D-alloisoleucine ratio in the protein isolated from the eggshells of an extinct Australian bird, a team of scientists recently determined that this bird lived approximately 50,000 years ago. Radiocarbon (^{14}C) dating is not accurate for samples older than about 35,000 years, so AAR is a useful addition to the tools available to paleontologists.

27.3 ACID–BASE BEHAVIOR OF AMINO ACIDS

The physical properties of a typical amino acid such as glycine suggest that it is a very polar substance, much more polar than would be expected on the basis of its formulation as $H_2NCH_2CO_2H$. Glycine is a crystalline solid; it does not melt, but on being heated it eventually decomposes at 233°C. It is very soluble in water but practically insoluble in nonpolar organic solvents. These properties are attributed to the fact that the stable form of glycine is a **zwitterion,** or **inner salt.**

The zwitterion is also often referred to as a *dipolar ion.* Note, however, that it is not an ion, but a neutral molecule.



Zwitterionic form of glycine

The equilibrium expressed by the preceding equation lies overwhelmingly to the side of the zwitterion.

Glycine, as well as other amino acids, is *amphoteric,* meaning it contains an acidic functional group and a basic functional group. The acidic functional group is the ammonium ion $H_3\overset{+}{N}$—; the basic functional group is the carboxylate ion $—CO_2^-$. How do we know this? Aside from its physical properties, the acid–base properties of glycine, as illustrated by the titration curve in Figure 27.2, require it. In a strongly acidic medium the species present is $H_3\overset{+}{N}CH_2CO_2H$. As the pH is raised, a proton is removed from this species. Is the proton removed from the positively charged nitrogen or from the carboxyl group? We know what to expect for the relative acid strengths of $R\overset{+}{N}H_3$ and RCO_2H. A typical ammonium ion has $pK_a \approx 9$, and a typical carboxylic acid has $pK_a \approx 5$. The

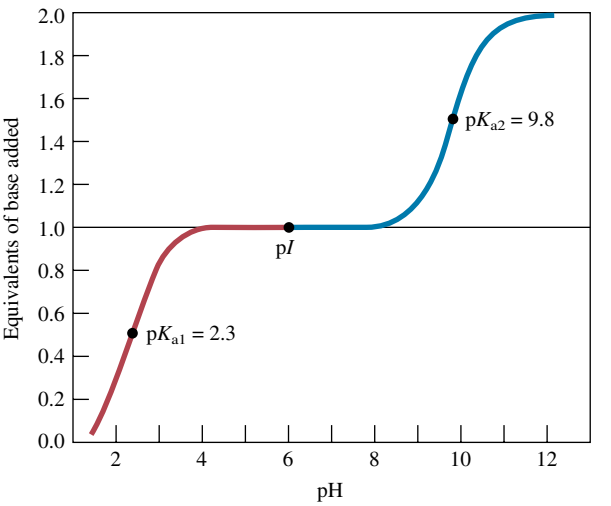
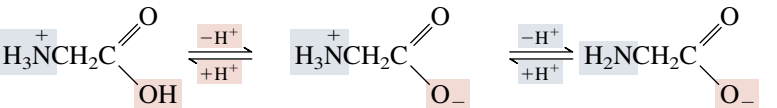


FIGURE 27.2 The titration curve of glycine. At pH values less than pK_{a1}, $H_3\overset{+}{N}CH_2CO_2H$ is the major species present. At pH values between pK_{a1} and pK_{a2}, the principal species is the zwitterion $H_3\overset{+}{N}CH_2CO_2^-$. The concentration of the zwitterion is a maximum at the isoelectric point pI. At pH values greater than pK_{a2}, $H_2NCH_2CO_2^-$ is the species present in greatest concentration.

measured pK_a for the conjugate acid of glycine is 2.35, a value closer to that expected for deprotonation of the carboxyl group. As the pH is raised, a second deprotonation step, corresponding to removal of a proton from nitrogen of the zwitterion, is observed. The pK_a associated with this step is 9.78, much like that of typical alkylammonium ions.



Species present in strong acid | Zwitterion; predominant species in solutions near neutrality | Species present in strong base

Thus, glycine is characterized by two pK_a values: the one corresponding to the more acidic site is designated pK_{a1}, the one corresponding to the less acidic site is designated pK_{a2}. Table 27.2 lists pK_{a1} and pK_{a2} values for the α-amino acids that have neutral side chains, which are the first two groups of amino acids given in Table 27.1. In all cases their pK_a values are similar to those of glycine.

Table 27.2 includes a column labeled pI, which gives **isoelectric point** values. The isoelectric point is the pH at which the amino acid bears no net charge; it corresponds to the pH at which the concentration of the zwitterion is a maximum. For the amino acids in Table 27.2 this is the average of pK_{a1} and pK_{a2} and lies slightly to the acid side of neutrality.

Some amino acids, including those listed in the last two sections of Table 27.1, have side chains that bear acidic or basic groups. As Table 27.3 indicates, these amino acids are characterized by three pK_a values. The "extra" pK_a value (it can be either pK_{a2} or pK_{a3}) reflects the nature of the function present in the side chain. The isoelectric points of the amino acids in Table 27.3 are midway between the pK_a values of the monocation and monoanion and are well removed from neutrality when the side chain bears a carboxyl group (aspartic acid, for example) or a basic amine function (lysine, for example).

TABLE 27.2 Acid–Base Properties of Amino Acids with Neutral Side Chains

Amino acid	pK_{a1}*	pK_{a2}*	pI
Glycine	2.34	9.60	5.97
Alanine	2.34	9.69	6.00
Valine	2.32	9.62	5.96
Leucine	2.36	9.60	5.98
Isoleucine	2.36	9.60	6.02
Methionine	2.28	9.21	5.74
Proline	1.99	10.60	6.30
Phenylalanine	1.83	9.13	5.48
Tryptophan	2.83	9.39	5.89
Asparagine	2.02	8.80	5.41
Glutamine	2.17	9.13	5.65
Serine	2.21	9.15	5.68
Threonine	2.09	9.10	5.60

*In all cases pK_{a1} corresponds to ionization of the carboxyl group; pK_{a2} corresponds to deprotonation of the ammonium ion.

TABLE 27.3 Acid–Base Properties of Amino Acids with Ionizable Side Chains

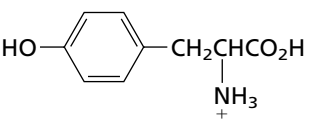
Amino acid	pK_{a1}*	pK_{a2}	pK_{a3}	pI
Aspartic acid	1.88	3.65	9.60	2.77
Glutamic acid	2.19	4.25	9.67	3.22
Tyrosine	2.20	9.11	10.07	5.66
Cysteine	1.96	8.18	10.28	5.07
Lysine	2.18	8.95	10.53	9.74
Arginine	2.17	9.04	12.48	10.76
Histidine	1.82	6.00	9.17	7.59

*In all cases pK_{a1} corresponds to ionization of the carboxyl group of $RCHCO_2H$ (NH_3^+).

PROBLEM 27.3 Write the most stable structural formula for tyrosine:

(a) In its cationic form
(b) In its zwitterionic form
(c) As a monoanion
(d) As a dianion

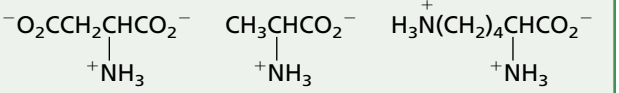
SAMPLE SOLUTION (a) The cationic form of tyrosine is the one present at low pH. The positive charge is on nitrogen, and the species present is an ammonium ion.



ELECTROPHORESIS

Electrophoresis is a method for separation and purification that depends on the movement of charged particles in an electric field. Its principles can be introduced by considering the electrophoretic behavior of some representative amino acids. The medium is a cellulose acetate strip that is moistened with an aqueous solution buffered at a particular pH. The opposite ends of the strip are placed in separate compartments containing the buffer, and each compartment is connected to a source of direct electric current (Figure 27.3*a*). If the buffer solution is more acidic than the isoelectric point (pI) of the amino acid, the amino acid has a net positive charge and migrates toward the negatively charged electrode. Conversely, when the buffer is more basic than the pI of the amino acid, the amino acid has a net negative charge and migrates toward the positively charged electrode. When the pH of the buffer corresponds to the pI, the amino acid has no net charge and does not migrate from the origin.

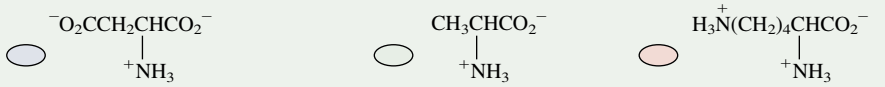
Thus if a mixture containing alanine, aspartic acid, and lysine is subjected to electrophoresis in a buffer that matches the isoelectric point of alanine (pH 6.0), aspartic acid (pI = 2.8) migrates toward the positive electrode, alanine remains at the origin, and lysine (pI = 9.7) migrates toward the negative electrode (Figure 27.3*b*).



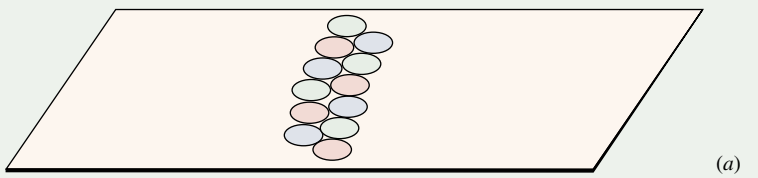
Aspartic acid (monoanion) | Alanine (neutral) | Lysine (monocation)

—Cont.

A mixture of amino acids



is placed at the center of a sheet of cellulose acetate. The sheet is soaked with an aqueous solution buffered at a pH of 6.0. At this pH aspartic acid ⬭ exists as its −1 ion, alanine ⬭ as its zwitterion, and lysine ⬭ as its +1 ion.



Application of an electric current causes the negatively charged ions to migrate to the + electrode, and the positively charged ions to migrate to the − electrode. The zwitterion, with a net charge of zero, remains at its original position.

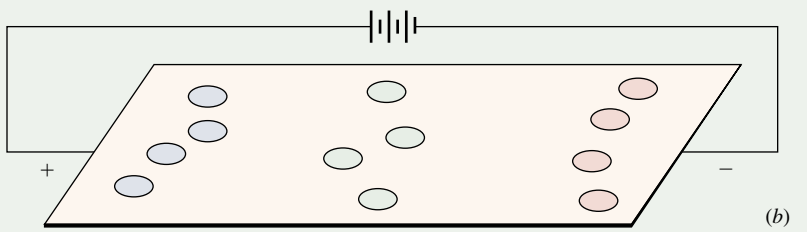


FIGURE 27.3 Application of electrophoresis to the separation of aspartic acid, alanine, and lysine according to their charge type at a pH corresponding to the isoelectric point (p*I*) of alanine.

Electrophoresis is used primarily to analyze mixtures of peptides and proteins, rather than individual amino acids, but analogous principles apply. Because they incorporate different numbers of amino acids and because their side chains are different, two peptides will have slightly different acid–base properties and slightly different net charges at a particular pH. Thus, their mobilities in an electric field will be different, and electrophoresis can be used to separate them. The medium used to separate peptides and proteins is typically a polyacrylamide gel, leading to the term **gel electrophoresis** for this technique.

A second factor that governs the rate of migration during electrophoresis is the size (length and shape) of the peptide or protein. Larger molecules move through the polyacrylamide gel more slowly than smaller ones. In current practice, the experiment is modified to exploit differences in size more than differences in net charge, especially in the **SDS gel electrophoresis** of proteins. Approximately 1.5 g of the detergent *sodium dodecyl sulfate* (SDS, page 745) per gram of protein is added to the aqueous buffer. SDS binds to the protein, causing the protein to unfold so that it is roughly rod-shaped with the $CH_3(CH_2)_{10}CH_2$— groups of SDS associated with the lipophilic portions of the protein. The negatively charged sulfate groups are exposed to the water. The SDS molecules that they carry ensure that all the protein molecules are negatively charged and migrate toward the positive electrode. Furthermore, all the proteins in the mixture now have similar shapes and tend to travel at rates proportional to their chain length. Thus, when carried out on a preparative scale, SDS gel electrophoresis permits proteins in a mixture to be separated according to their molecular weight. On an analytical scale, it is used to estimate the molecular weight of a protein by comparing its electrophoretic mobility with that of proteins of known molecular weight.

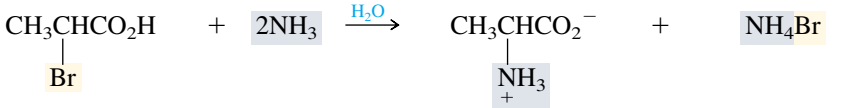
Later, in Section 27.29, we will see how gel electrophoresis is used in nucleic acid chemistry.

PROBLEM 27.4 Write structural formulas for the principal species present when the pH of a solution containing lysine is raised from 1 to 9 and again to 13.

The acid–base properties of their side chains are one way in which individual amino acids differ. This is important in peptides and proteins, where the properties of the substance depend on its amino acid constituents, especially on the nature of the side chains. It is also important in analyses in which a complex mixture of amino acids is separated into its components by taking advantage of the differences in their proton-donating and proton-accepting abilities.

27.4 SYNTHESIS OF AMINO ACIDS

One of the oldest methods for the synthesis of amino acids dates back to the nineteenth century and is simply a nucleophilic substitution in which ammonia reacts with an α-halo carboxylic acid.



2-Bromopropanoic acid Ammonia Alanine (65–70%) Ammonium bromide

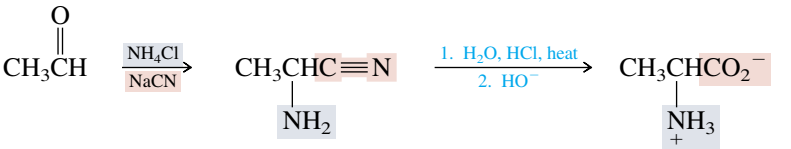
The α-halo acid is normally prepared by the Hell–Volhard–Zelinsky reaction (see Section 19.16).

PROBLEM 27.5 Outline the steps in a synthesis of valine from 3-methylbutanoic acid.

In the **Strecker synthesis** an aldehyde is converted to an α-amino acid with one more carbon atom by a two-stage procedure in which an α-amino nitrile is an intermediate.

The α-amino nitrile is formed by reaction of the aldehyde with ammonia or an ammonium salt and a source of cyanide ion. Hydrolysis of the nitrile group to a carboxylic acid function completes the synthesis.

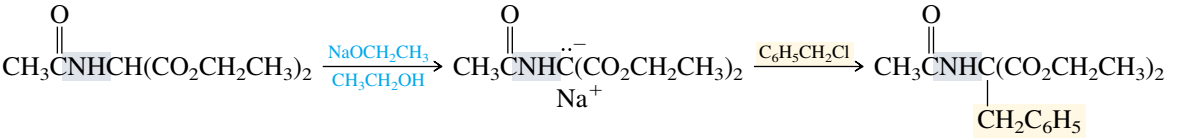
The synthesis of alanine was described by Adolf Strecker of the University of Würzburg (Germany) in a paper published in 1850.



Acetaldehyde — 2-Aminopropanenitrile — Alanine (52–60%)

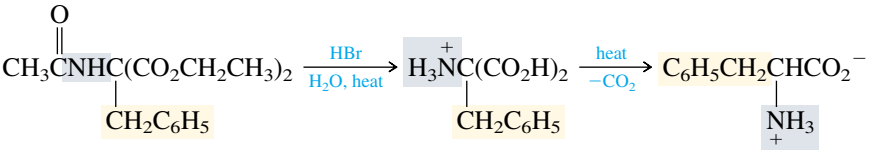
PROBLEM 27.6 Outline the steps in the preparation of valine by the Strecker synthesis.

The most widely used method for the laboratory synthesis of α-amino acids is a modification of the malonic ester synthesis (Section 21.7). The key reagent is *diethyl acetamidomalonate,* a derivative of malonic ester that already has the critical nitrogen substituent in place at the α-carbon atom. The side chain is introduced by alkylating diethyl acetamidomalonate in the same way as diethyl malonate itself is alkylated.



Diethyl acetamidomalonate — Sodium salt of diethyl acetamidomalonate — Diethyl acetamidobenzylmalonate (90%)

Hydrolysis removes the acetyl group from nitrogen and converts the two ester functions to carboxyl groups. Decarboxylation gives the desired product.



Diethyl acetamidobenzylmalonate — (not isolated) — Phenylalanine (65%)

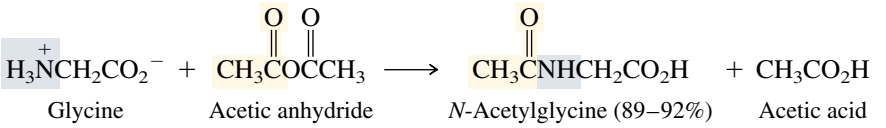
PROBLEM 27.7 Outline the steps in the synthesis of valine from diethyl acetamidomalonate. The overall yield of valine by this method is reported to be rather low (31%). Can you think of a reason why this synthesis is not very efficient?

Unless a resolution step is included, the α-amino acids prepared by the synthetic methods just described are racemic. Optically active amino acids, when desired, may be obtained by resolving a racemic mixture or by **enantioselective synthesis.** A synthesis is described as enantioselective if it produces one enantiomer of a chiral compound in an amount greater than its mirror image. Recall from Section 7.9 that optically inactive reactants cannot give optically active products. Enantioselective syntheses of amino acids therefore require an enantiomerically enriched chiral reagent or catalyst at some point in

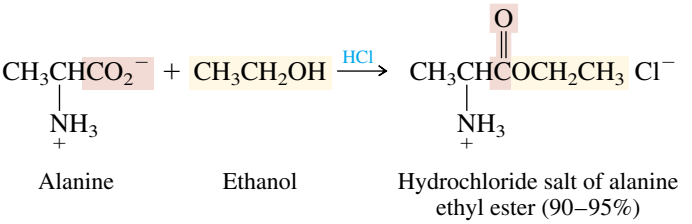
the process. If the chiral reagent or catalyst is a single enantiomer and if the reaction sequence is completely enantioselective, an optically pure amino acid is obtained. Chemists have succeeded in preparing α-amino acids by techniques that are more than 95% enantioselective. Although this is an impressive feat, we must not lose sight of the fact that the reactions that produce amino acids in living systems do so with 100% enantioselectivity.

27.5 REACTIONS OF AMINO ACIDS

Amino acids undergo reactions characteristic of both their amine and carboxylic acid functional groups. Acylation is a typical reaction of the amino group.

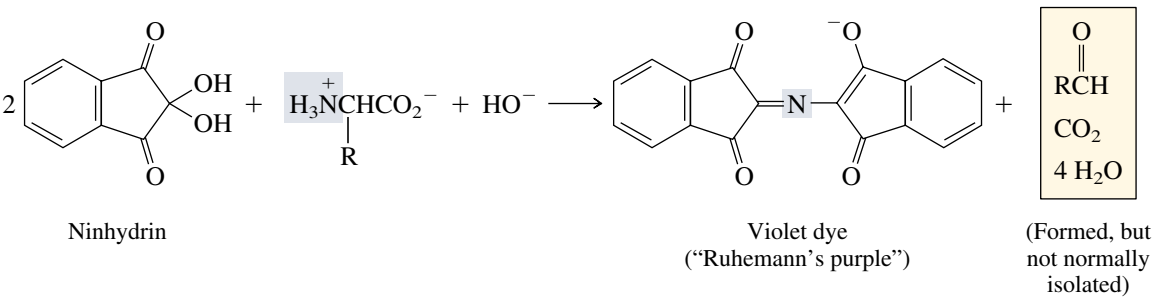


Ester formation is a typical reaction of the carboxyl group.



The presence of amino acids can be detected by the formation of a purple color on treatment with *ninhydrin*. The same compound responsible for the purple color is formed from all amino acids in which the α-amino group is primary.

Ninhydrin is used to detect fingerprints.



Proline, in which the α-amino group is secondary, gives an orange compound on reaction with ninhydrin.

PROBLEM 27.8 Suggest a reasonable mechanism for the reaction of an α-amino acid with ninhydrin.

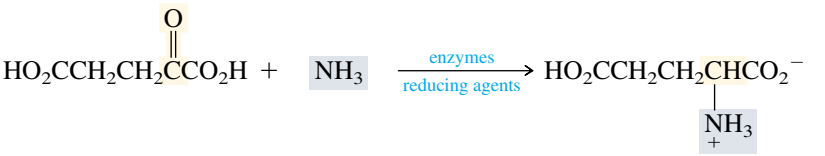
27.6 SOME BIOCHEMICAL REACTIONS OF AMINO ACIDS

The 20 amino acids listed in Table 27.1 are biosynthesized by a number of different pathways, and we will touch on only a few of them in an introductory way. We will examine the biosynthesis of glutamic acid first, since it illustrates a biochemical process

analogous to a reaction we have discussed earlier in the context of amine synthesis, *reductive amination* (Section 22.11).

Glutamic acid is formed in most organisms from ammonia and α-ketoglutaric acid. α-Ketoglutaric acid is one of the intermediates in the **tricarboxylic acid cycle** (also called the **Krebs cycle**) and arises via metabolic breakdown of food sources—carbohydrates, fats, and proteins.

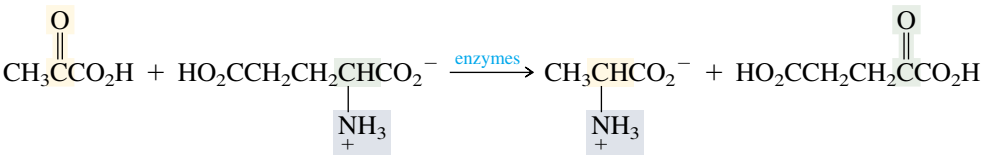
The August 1986 issue of the *Journal of Chemical Education* (pp. 673–677) contains a review of the Krebs cycle.



α-Ketoglutaric acid Ammonia L-Glutamic acid

Ammonia reacts with the ketone carbonyl group to give an imine (C═NH), which is then reduced to the amine function of the α-amino acid. Both imine formation and reduction are enzyme-catalyzed. The reduced form of nicotinamide adenine diphosphonucleotide (NADPH) is a coenzyme and acts as a reducing agent. The step in which the imine is reduced is the one in which the stereogenic center is introduced and gives only L-glutamic acid.

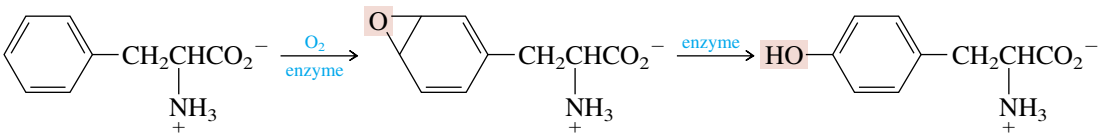
L-Glutamic acid is not an essential amino acid. It need not be present in the diet, since animals can biosynthesize it from sources of α-ketoglutaric acid. It is, however, a key intermediate in the biosynthesis of other amino acids by a process known as **transamination.** L-Alanine, for example, is formed from pyruvic acid by transamination from L-glutamic acid.



Pyruvic acid L-Glutamic acid L-Alanine α-Ketoglutaric acid

In transamination an amine group is transferred from L-glutamic acid to pyruvic acid. An outline of the mechanism of transamination is presented in Figure 27.4.

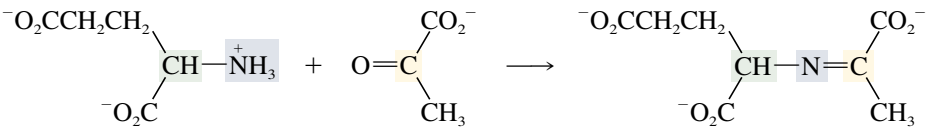
One amino acid often serves as the biological precursor to another. L-Phenylalanine is classified as an essential amino acid, whereas its *p*-hydroxy derivative, L-tyrosine, is not. This is because animals can convert L-phenylalanine to L-tyrosine by hydroxylation of the aromatic ring. An *arene oxide* (Section 24.7) is an intermediate.



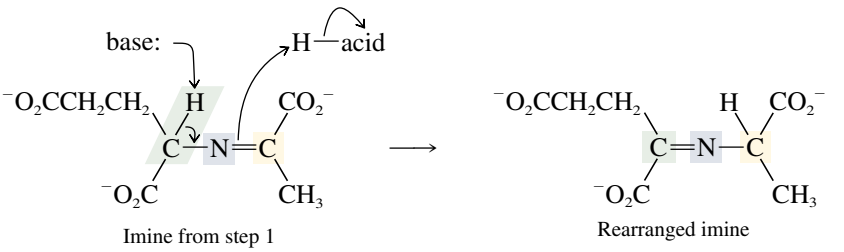
L-Phenylalanine Arene oxide intermediate L-Tyrosine

Some people lack the enzymes necessary to convert L-phenylalanine to L-tyrosine. Any L-phenylalanine that they obtain from their diet is diverted along a different metabolic pathway, giving phenylpyruvic acid:

Step 1: The amine function of L-glutamate reacts with the ketone carbonyl of pyruvate to form an imine.



Step 2: Enzyme-catalyzed proton-transfer steps cause migration of the double bond, converting the imine formed in step 1 to an isomeric imine.



Step 3: Hydrolysis of the rearranged imine gives L-alanine and α-ketoglutarate.

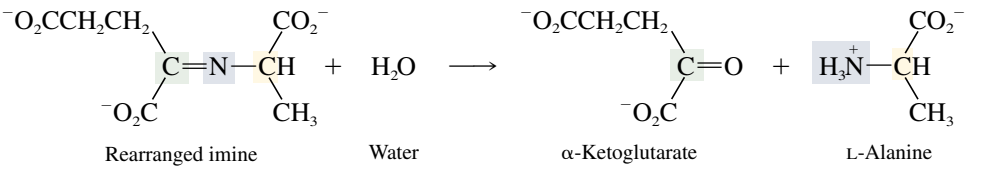
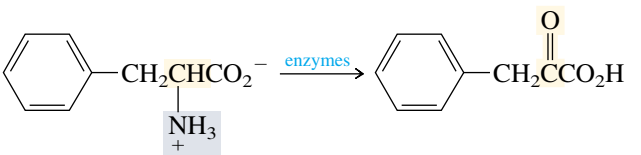
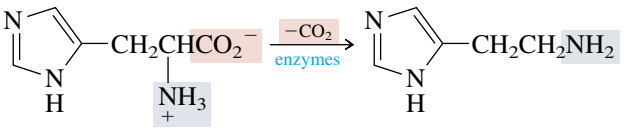


FIGURE 27.4 The mechanism of transamination. All the steps are enzyme-catalyzed.



Phenylpyruvic acid can cause mental retardation in infants who are deficient in the enzymes necessary to convert L-phenylalanine to L-tyrosine. This disorder is called **phenylketonuria,** or **PKU disease.** PKU disease can be detected by a simple test routinely administered to newborns. It cannot be cured, but is controlled by restricting the dietary intake of L-phenylalanine. In practice this means avoiding foods such as meat that are rich in L-phenylalanine.

Among the biochemical reactions that amino acids undergo is *decarboxylation* to amines. Decarboxylation of histidine, for example, gives histamine, a powerful vasodilator normally present in tissue and formed in excessive amounts under conditions of traumatic shock.



Histidine Histamine

Histamine is responsible for many of the symptoms associated with hay fever and other allergies. An antihistamine relieves these symptoms by blocking the action of histamine.

PROBLEM 27.9 One of the amino acids in Table 27.1 is the biological precursor to γ-aminobutyric acid (4-aminobutanoic acid), which it forms by a decarboxylation reaction. Which amino acid is this?

The chemistry of the brain and central nervous system is affected by a group of substances called **neurotransmitters.** Several of these neurotransmitters arise from L-tyrosine by structural modification and decarboxylation, as outlined in Figure 27.5.

For a review of neurotransmitters, see the February 1988 issue of the *Journal of Chemical Education* (pp. 108–111).

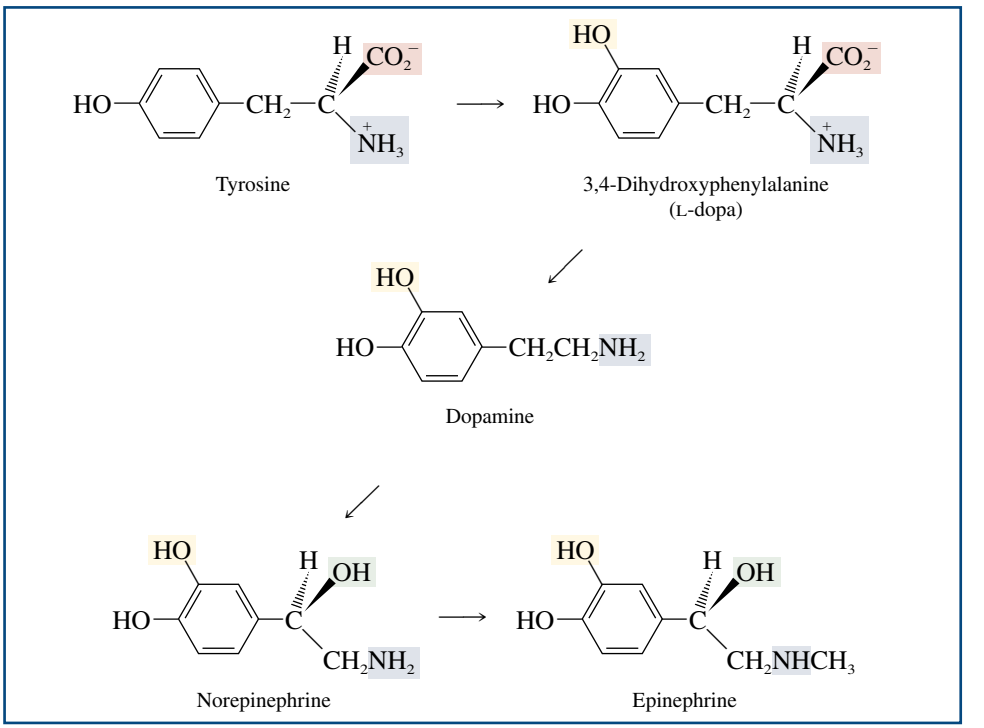
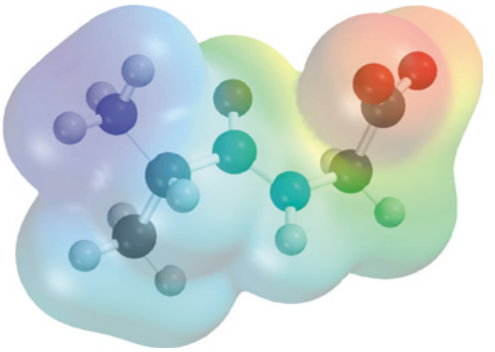
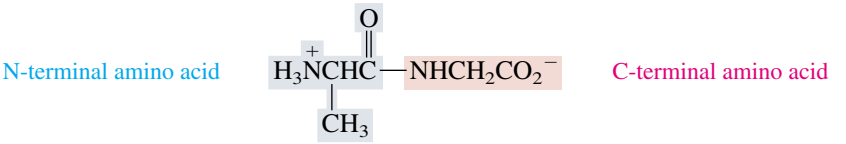


FIGURE 27.5 Tyrosine is the biosynthetic precursor to a number of neurotransmitters. Each transformation is enzyme-catalyzed. Hydroxylation of the aromatic ring of tyrosine converts it to 3,4-dihydroxyphenylalanine (L-dopa), decarboxylation of which gives dopamine. Hydroxylation of the benzylic carbon of dopamine converts it to norepinephrine (noradrenaline), and methylation of the amino group of norepinephrine yields epinephrine (adrenaline).



27.7 PEPTIDES

A key biochemical reaction of amino acids is their conversion to peptides, polypeptides, and proteins. In all these substances amino acids are linked together by amide bonds. The amide bond between the amino group of one amino acid and the carboxyl of another is called a **peptide bond.** Alanylglycine is a representative dipeptide.



Alanylglycine
(Ala-Gly)

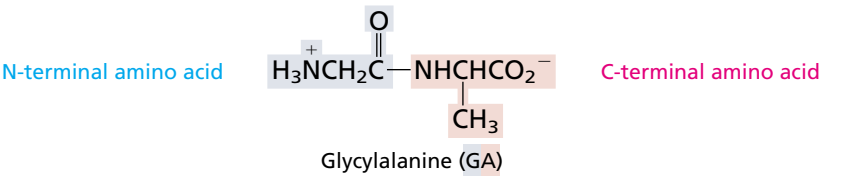
By agreement, peptide structures are written so that the amino group (as $H_3\overset{+}{N}$— or H_2N—) is at the left and the carboxyl group (as CO_2^- or CO_2H) is at the right. The left and right ends of the peptide are referred to as the **N terminus** (or amino terminus) and the **C terminus** (or carboxyl terminus), respectively. Alanine is the N-terminal amino acid in alanylglycine; glycine is the C-terminal amino acid. A dipeptide is named as an acyl derivative of the C-terminal amino acid. We call the precise order of bonding in a peptide its amino acid **sequence.** The amino acid sequence is conveniently specified by using the three-letter amino acid abbreviations for the respective amino acids and connecting them by hyphens. Individual amino acid components of peptides are often referred to as amino acid **residues.**

It is understood that α-amino acids occur as their L stereoisomers unless otherwise indicated. The D notation is explicitly shown when a D amino acid is present, and a racemic amino acid is identified by the prefix DL.

PROBLEM 27.10 Write structural formulas showing the constitution of each of the following dipeptides. Rewrite each sequence using one-letter abbreviations for the amino acids.

(a) Gly-Ala
(b) Ala-Phe
(c) Phe-Ala
(d) Gly-Glu
(e) Lys-Gly
(f) D-Ala-D-Ala

SAMPLE SOLUTION (a) Gly-Ala is a constitutional isomer of Ala-Gly. Glycine is the N-terminal amino acid in Gly-Ala; alanine is the C-terminal amino acid.

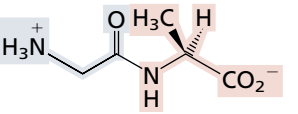


Glycylalanine (GA)

Figure 27.6 shows the structure of Ala-Gly as determined by X-ray crystallography. An important feature is the planar geometry associated with the peptide bond, and the most stable conformation with respect to this bond has the two α-carbon atoms anti to each other. Rotation about the amide linkage is slow because delocalization of the unshared electron pair of nitrogen into the carbonyl group gives partial double-bond character to the carbon–nitrogen bond.

PROBLEM 27.11 Expand your answer to Problem 27.10 by showing the structural formula for each dipeptide in a manner that reveals the stereochemistry at the α-carbon atom.

SAMPLE SOLUTION (a) Glycine is achiral, and so Gly-Ala has only one stereogenic center, the α-carbon atom of the L-alanine residue. When the carbon chain is drawn in an extended zigzag fashion and L-alanine is the C terminus, its structure is as shown:



Glycyl-L-alanine (Gly-Ala)

The structures of higher peptides follow in an analogous fashion. Figure 27.7 gives the structural formula and amino acid sequence of a naturally occurring pentapeptide known as *leucine enkephalin.* Enkephalins are pentapeptide components of **endorphins,** polypeptides present in the brain that act as the body's own painkillers. A second substance, known as *methionine enkephalin,* is also present in endorphins. Methionine enkephalin differs from leucine enkephalin only in having methionine instead of leucine as its C-terminal amino acid.

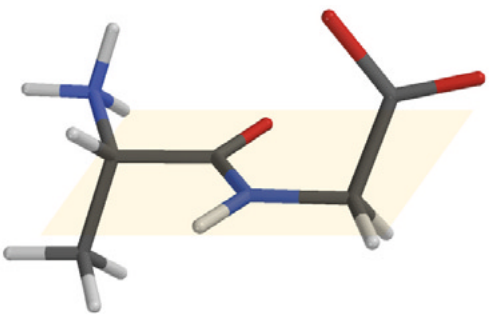


FIGURE 27.6 Structural features of the dipeptide L-alanylglycine as determined by X-ray crystallography.

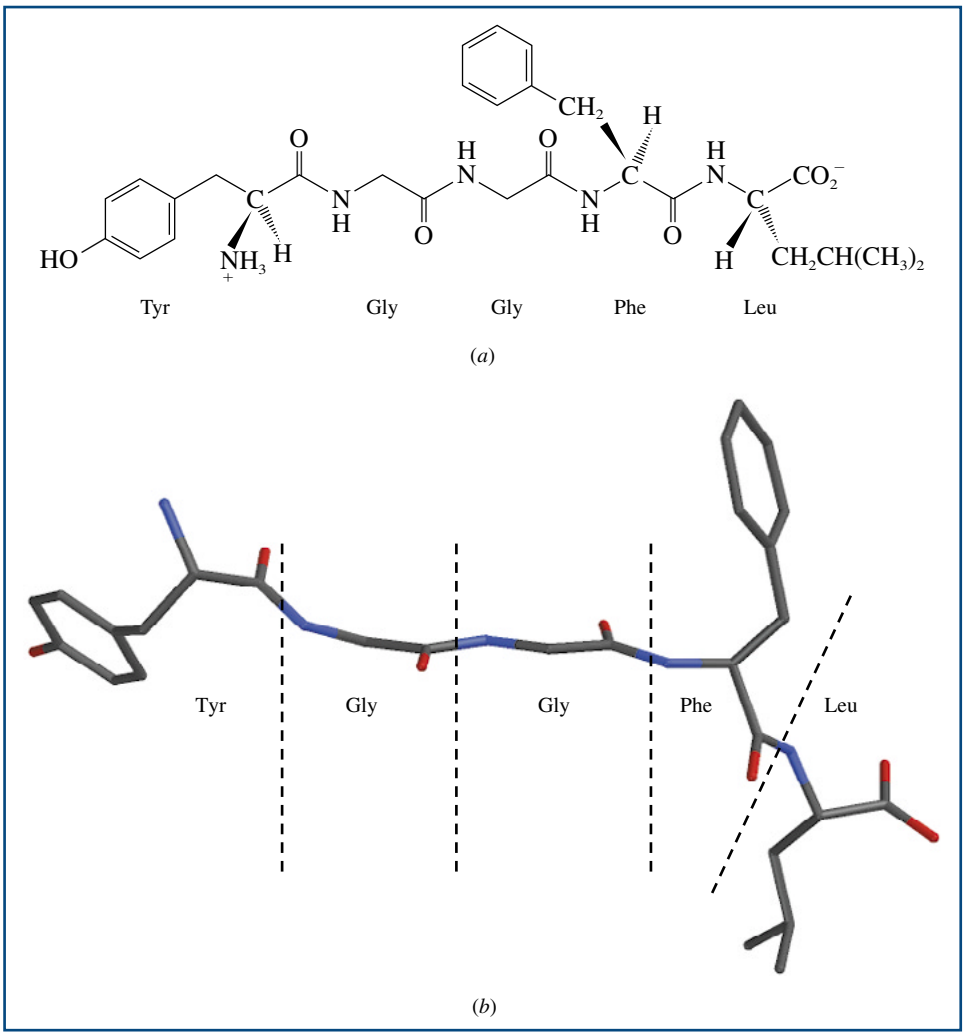


FIGURE 27.7 The structure of the pentapeptide leucine enkephalin shown as (*a*) a structural drawing and (*b*) as a molecular model. The shape of the molecular model was determined by X-ray crystallography. Hydrogens have been omitted for clarity.

PROBLEM 27.12 What is the amino acid sequence (using three-letter abbreviations) of methionine enkephalin? Show it using one-letter abbreviations.

Peptides having structures slightly different from those described to this point are known. One such variation is seen in the nonapeptide *oxytocin,* shown in Figure 27.8. Oxytocin is a hormone secreted by the pituitary gland that stimulates uterine contractions during childbirth. Rather than terminating in a carboxyl group, the terminal glycine residue in oxytocin has been modified so that it exists as the corresponding amide. Two cysteine units, one of them the N-terminal amino acid, are joined by the sulfur–sulfur bond of a large-ring cyclic disulfide unit. This is a common structural modification in polypeptides and proteins that contain cysteine residues. It provides a covalent bond between regions of peptide chains that may be many amino acid residues removed from each other.

Recall from Section 15.14 that compounds of the type RSH are readily oxidized to RSSR.

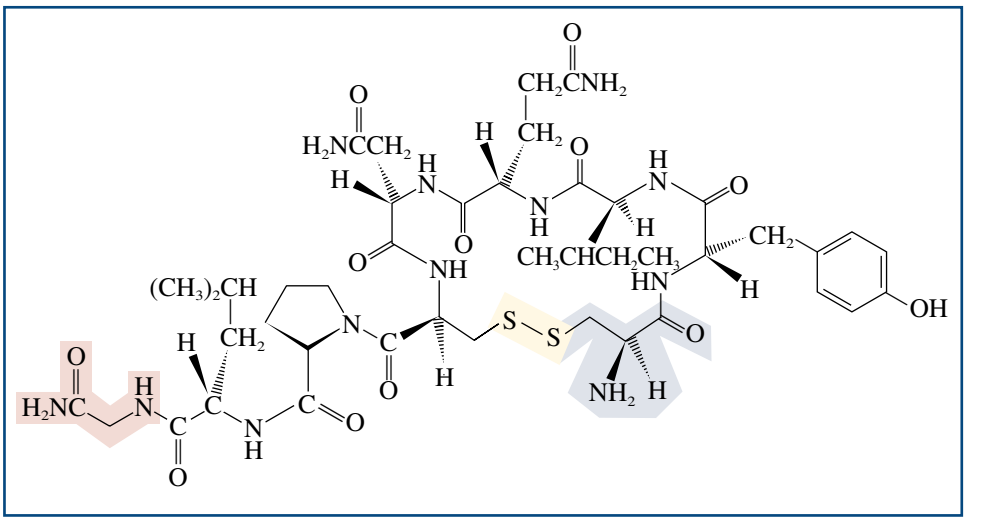


FIGURE 27.8 The structure of oxytocin, a nonapeptide containing a disulfide bond between two cysteine residues. One of these cysteines is the N-terminal amino acid and is highlighted in blue. The C-terminal amino acid is the amide of glycine and is highlighted in red. There are no free carboxyl groups in the molecule; all exist in the form of carboxamides.

27.8 INTRODUCTION TO PEPTIDE STRUCTURE DETERMINATION

There are several levels of peptide structure. The **primary structure** is the amino acid sequence plus any disulfide links. With the 20 amino acids of Table 27.1 as building blocks, 20^2 dipeptides, 20^3 tripeptides, 20^4 tetrapeptides, and so on, are possible. Given a peptide of unknown structure, how do we determine its amino acid sequence?

We'll describe peptide structure determination by first looking at one of the great achievements of biochemistry, the determination of the amino acid sequence of insulin by Frederick Sanger of Cambridge University (England). Sanger was awarded the 1958 Nobel Prize in chemistry for this work, which he began in 1944 and completed 10 years later. The methods used by Sanger and his coworkers are, of course, dated by now, but the overall strategy hasn't changed very much. We'll use Sanger's insulin work to orient us with respect to strategy, then show how current methods of protein sequencing have evolved from it.

Sanger was a corecipient of a second Nobel Prize in 1980 for devising methods for sequencing nucleic acids. Sanger's strategy for nucleic acid sequencing will be described in Section 27.29.

Sanger's strategy can be outlined as follows:

1. Determine what amino acids are present and their molar ratios.
2. Cleave the peptide into smaller fragments, separate these fragments, and determine the amino acid composition of the fragments.
3. Identify the N-terminal and the C-terminal amino acid in the original peptide and in each fragment.
4. Organize the information so that the amino acid sequences of small fragments can be overlapped to reveal the full sequence.

27.9 AMINO ACID ANALYSIS

The chemistry behind amino acid analysis is nothing more than acid-catalyzed hydrolysis of amide (peptide) bonds. The peptide is hydrolyzed by heating in 6 M hydrochloric acid for about 24 h to give a solution that contains all the amino acids. This mixture is then separated by **ion-exchange chromatography,** which separates the amino acids mainly according to their acid–base properties. As the amino acids leave the chromatography column, they are mixed with ninhydrin and the intensity of the ninhydrin

color monitored electronically. The amino acids are identified by comparing their chromatographic behavior with authentic samples, and their relative amounts from peak areas as recorded on a strip chart.

The entire operation is carried out automatically using an **amino acid analyzer** and is so sensitive that as little as 10^{-5}–10^{-7} g of the peptide is required.

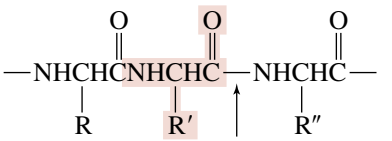
PROBLEM 27.13 Amino acid analysis of a certain tetrapeptide gave alanine, glycine, phenylalanine, and valine in equimolar amounts. What amino acid sequences are possible for this tetrapeptide?

27.10 PARTIAL HYDROLYSIS OF PEPTIDES

Whereas acid-catalyzed hydrolysis of peptides cleaves amide bonds indiscriminately and eventually breaks all of them, enzymatic hydrolysis is much more selective and is the method used to convert a peptide into smaller fragments.

The enzymes that catalyze the hydrolysis of peptides are called **peptidases, proteases,** or **proteolytic enzymes.** One group of pancreatic enzymes, known as *carboxypeptidases,* catalyzes only the hydrolysis of the peptide bond to the C-terminal amino acid, for example. *Trypsin,* a digestive enzyme present in the intestine, catalyzes only the hydrolysis of peptide bonds involving the carboxyl group of a lysine or arginine residue. *Chymotrypsin,* another digestive enzyme, is selective for peptide bonds involving the carboxyl group of amino acids with aromatic side chains (phenylalanine, tryrosine, tryptophan). In addition to these, many other digestive enzymes are known and their selectivity exploited in the selective hydrolysis of peptides.

Papain, the active component of most meat tenderizers, is a proteolytic enzyme.



Site of chymotrypsin-catalyzed hydrolysis when R′ is an aromatic side chain

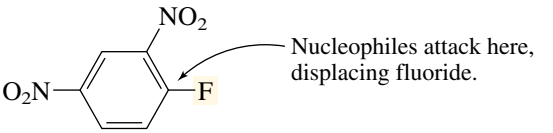
PROBLEM 27.14 Digestion of the tetrapeptide of Problem 27.13 with chymotrypsin gave a dipeptide that on amino acid analysis gave phenylalanine and valine in equimolar amounts. What amino acid sequences are possible for the tetrapeptide?

27.11 END GROUP ANALYSIS

An amino acid sequence is ambiguous unless we know the direction in which to read it—left to right, or right to left. We need to know which end is the N terminus and which is the C terminus. As we saw in the preceding section, carboxypeptidase-catalyzed hydrolysis cleaves the C-terminal amino acid and so can be used to identify it. What about the N terminus?

Several chemical methods have been devised for identifying the N-terminal amino acid. They all take advantage of the fact that the N-terminal amino group is free and can act as a nucleophile. The α-amino groups of all the other amino acids are part of amide linkages, are not free, and are much less nucleophilic. Sanger's method for N-terminal residue analysis involves treating a peptide with 1-fluoro-4-nitrobenzene, which is very reactive toward nucleophilic aromatic substitution.

1-Fluoro-4-nitrobenzene is commonly referred to as *Sanger's reagent.*

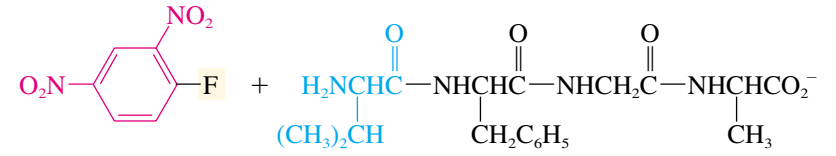


1-Fluoro-2,4-dinitrobenzene

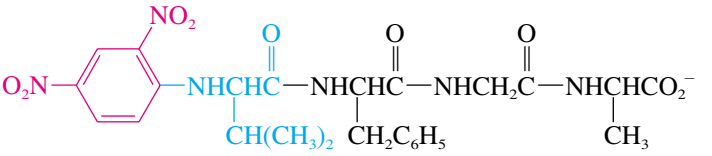
The amino group of the N-terminal amino acid displaces fluoride from 1-fluoro-2,4-dinitrobenzene and gives a peptide in which the N-terminal nitrogen is labeled with a 2,4-dinitrophenyl (DNP) group. This is shown for the case of Val-Phe-Gly-Ala in Figure 27.9. The 2,4-dinitrophenyl-labeled peptide DNP-Val-Phe-Gly-Ala is isolated and subjected to hydrolysis, after which the 2,4-dinitrophenyl derivative of the N-terminal amino acid is isolated and identified as DNP-Val by comparing its chromatographic behavior with that of standard samples of 2,4-dinitrophenyl-labeled amino acids. None of the other amino acid residues bear a 2,4-dinitrophenyl group; they appear in the hydrolysis product as the free amino acids.

FIGURE 27.9 Use of 1-fluoro-2,4-dinitrobenzene to identify the N-terminal amino acid of a peptide.

The reaction is carried out by mixing the peptide and 1-fluoro-2,4-dinitrobenzene in the presence of a weak base such as sodium carbonate. In the first step the base abstracts a proton from the terminal $H_3\overset{+}{N}$ group to give a free amino function. The nucleophilic amino group attacks 1-fluoro-2,4-dinitrobenzene, displacing fluoride.

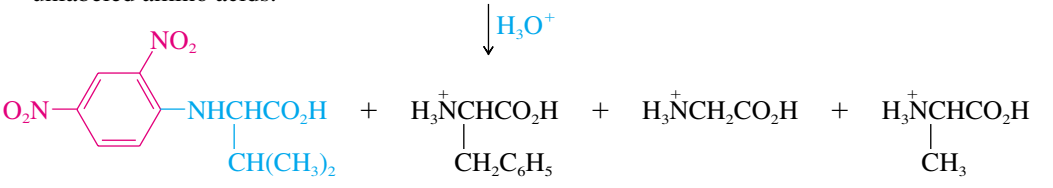


1-Fluoro-2,4-dinitrobenzene Val-Phe-Gly-Ala



DNP-Val-Phe-Gly-Ala

Acid hydrolysis cleaves the amide bonds of the 2,4-dinitrophenyl-labeled peptide, giving the 2,4-dinitrophenyl-labeled N-terminal amino acid and a mixture of unlabeled amino acids.



DNP-Val Phe Gly Ala

Labeling the N-terminal amino acid as its DNP derivative is mainly of historical interest and has been replaced by other methods. We'll discuss one of these—the Edman degradation—in Section 27.13. First, though, we'll complete our review of the general strategy for peptide sequencing by seeing how Sanger tied all of the information together into a structure for insulin.

27.12 INSULIN

Insulin has 51 amino acids, divided between two chains. One of these, the A chain, has 21 amino acids; the other, the B chain, has 30. The A and B chains are joined by disulfide bonds between cysteine residues (Cys—Cys). Figure 27.10 shows some of the information that defines the amino acid sequence of the B chain.

- Reaction of the B chain peptide with 1-fluoro-4-nitrobenzene established that phenylalanine is the N terminus.
- Pepsin-catalyzed hydrolysis gave the four peptides shown in blue in Figure 27.10. (Their sequences were determined in separate experiments.) These four peptides contain 27 of the 30 amino acids in the B chain, but there are no points of overlap between them.
- The sequences of the four tetrapeptides shown in red in Figure 27.10 bridge the gaps between three of the four "blue" peptides to give an unbroken sequence from 1 through 24.
- The peptide shown in yellow was isolated by trypsin-catalyzed hydrolysis and has an amino acid sequence that completes the remaining overlaps.

Sanger also determined the sequence of the A chain and identified the cysteine residues involved in disulfide bonds between the A and B chains as well as in the

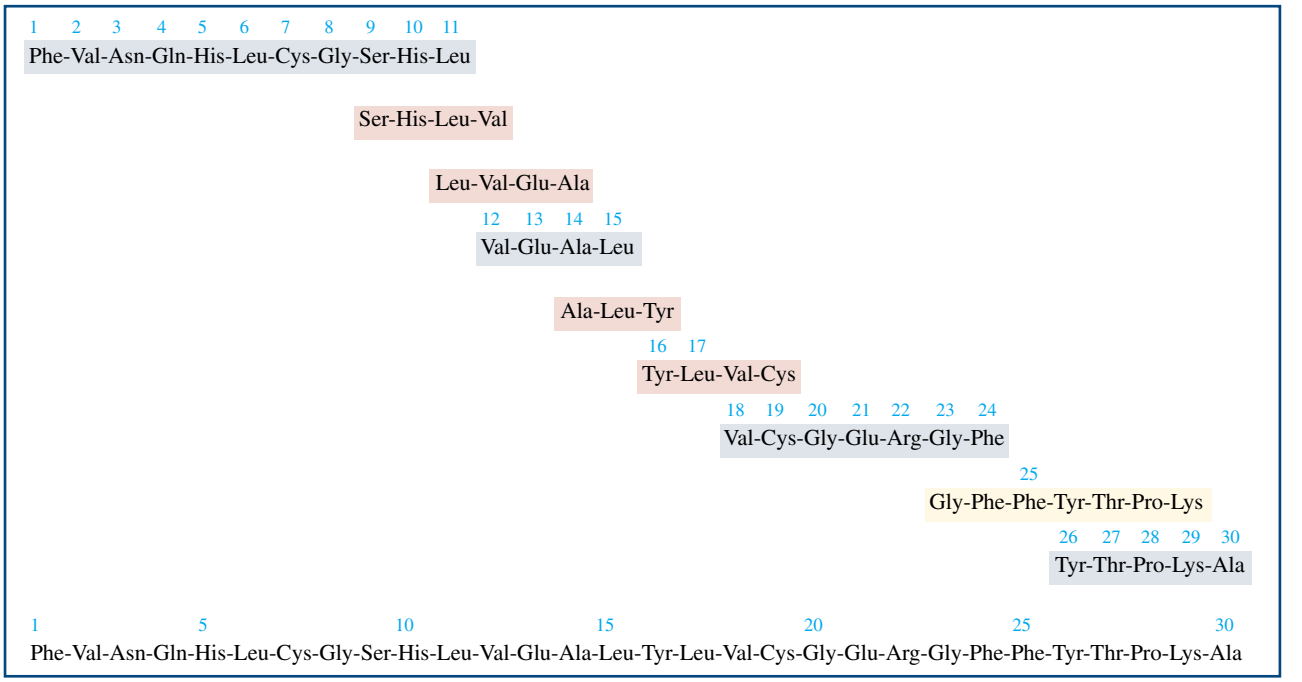


FIGURE 27.10 Diagram showing how the amino acid sequence of the B chain of bovine insulin can be determined by overlap of peptide fragments. Pepsin-catalyzed hydrolysis produced the fragments shown in blue, trypsin produced the one shown in yellow, and acid-catalyzed hydrolysis gave many fragments, including the four shown in red.

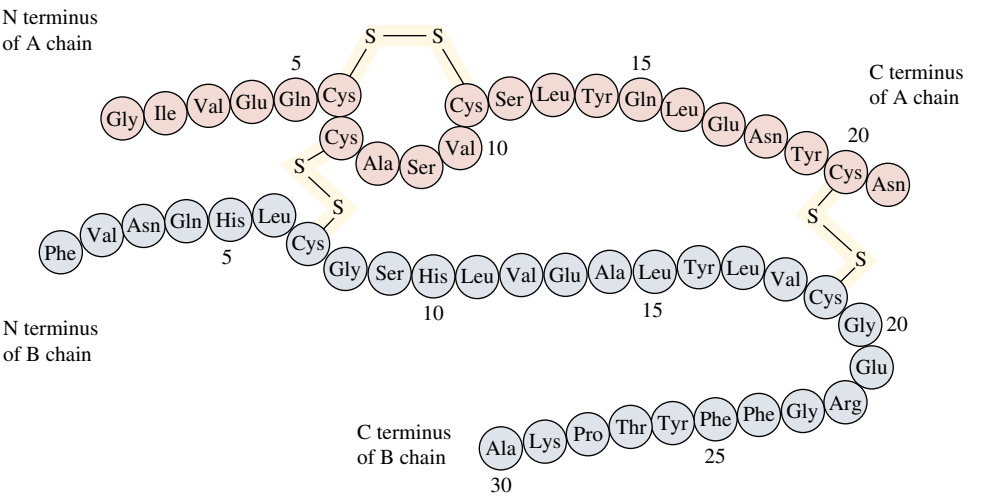


FIGURE 27.11 The amino acid sequence in bovine insulin. The A chain is shown in red and the B chain in blue. The A chain is joined to the B chain by two disulfide units (yellow). There is also a disulfide bond linking cysteines 6 and 11 in the A chain. Human insulin has threonine and isoleucine at residues 8 and 10, respectively, in the A chain and threonine as the C-terminal amino acid in the B chain.

disulfide linkage within the A chain. The complete insulin structure is shown in Figure 27.11. The structure shown is that of bovine insulin (from cattle). The A chains of human insulin and bovine insulin differ in only two amino acid residues; their B chains are identical except for the amino acid at the C terminus.

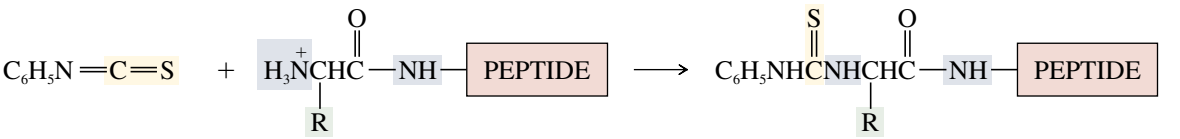
27.13 THE EDMAN DEGRADATION AND AUTOMATED SEQUENCING OF PEPTIDES

The years that have passed since Sanger determined the structure of insulin have seen refinements in technique while retaining the same overall strategy. Enzyme-catalyzed hydrolysis to convert a large peptide to smaller fragments remains an important component, as does searching for overlaps among these smaller fragments. The method for N-terminal residue analysis, however, has been improved so that much smaller amounts of peptide are required, and the analysis has been automated.

When Sanger's method for N-terminal residue analysis was discussed, you may have wondered why it was not done sequentially. Simply start at the N terminus and work steadily back to the C terminus identifying one amino acid after another. The idea is fine, but it just doesn't work well in practice, at least with 1-fluoro-4-nitrobenzene.

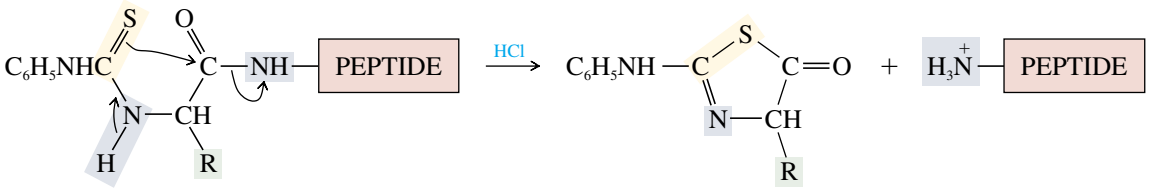
A major advance was devised by Pehr Edman (University of Lund, Sweden) that has become the standard method for N-terminal residue analysis. The **Edman degradation** is based on the chemistry shown in Figure 27.12. A peptide reacts with phenyl isothiocyanate to give a *phenylthiocarbamoyl* (PTC) derivative, as shown in the first step. This PTC derivative is then treated with an acid in an *anhydrous* medium (Edman used nitromethane saturated with hydrogen chloride) to cleave the amide bond between the N-terminal amino acid and the remainder of the peptide. No other peptide bonds are cleaved in this step as amide bond hydrolysis requires water. When the PTC derivative is treated with acid in an anhydrous medium, the sulfur atom of the C═S unit acts as an internal nucleophile, and the only amide bond cleaved under these conditions is the one to the N-terminal amino acid. The product of this cleavage, called a *thiazolone,* is unstable under the conditions of its formation and rearranges to a *phenylthiohydantoin* (PTH), which is isolated and identified by comparing it with standard samples of PTH derivatives of known amino acids. This is normally done by chromatographic methods, but mass spectrometry has also been used.

Step 1: A peptide is treated with phenyl isothiocyanate to give a phenylthiocarbamoyl (PTC) derivative.



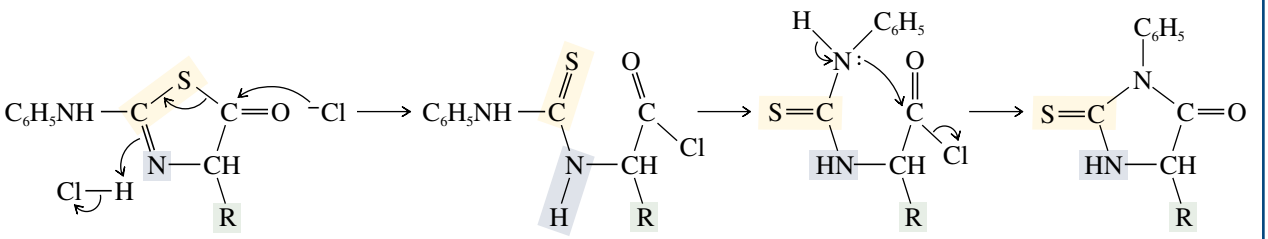
Phenyl isothiocyanate PTC derivative

Step 2: On reaction with hydrogen chloride in an anhydrous solvent, the thiocarbonyl sulfur of the PTC derivative attacks the carbonyl carbon of the N-terminal amino acid. The N-terminal amino acid is cleaved as a thiazolone derivative from the remainder of the peptide.



PTC derivative Thiazolone Remainder of peptide

Step 3: Once formed, the thiazolone derivative isomerizes to a more stable phenylthiohydantoin (PTH) derivative, which is isolated and characterized, thereby providing identification of the N-terminal amino acid. The remainder of the peptide (formed in step 2) can be isolated and subjected to a second Edman degradation.



Thiazolone PTH derivative

FIGURE 27.12 Identification of the N-terminal amino acid of a peptide by Edman degradation.

Only the N-terminal amide bond is broken in the Edman degradation; the rest of the peptide chain remains intact. It can be isolated and subjected to a second Edman procedure to determine its new N terminus. We can proceed along a peptide chain by beginning with the N terminus and determining each amino acid in order. The sequence is given directly by the structure of the PTH derivative formed in each successive degradation.

PROBLEM 27.15 Give the structure of the PTH derivative isolated in the second Edman cycle of the tetrapeptide Val-Phe-Gly-Ala.

Ideally, one could determine the primary structure of even the largest protein by repeating the Edman procedure. Because anything less than 100% conversion in any single Edman degradation gives a mixture containing some of the original peptide along with the degraded one, two different PTH derivatives are formed in the next Edman cycle, and the ideal is not realized in practice. Nevertheless, some impressive results

have been achieved. It is a fairly routine matter to sequence the first 20 amino acids from the N terminus by repetitive Edman cycles, and even 60 residues have been determined on a single sample of the protein myoglobin. The entire procedure has been automated and incorporated into a device called an **Edman sequenator,** which carries out all the operations under computer control.

The amount of sample required is quite small; as little as 10^{-10} mol is typical. So many peptides and proteins have been sequenced now that it is impossible to give an accurate count. What was Nobel Prize-winning work in 1958 is routine today. Nor has the story ended. Sequencing of nucleic acids has advanced so dramatically that it is possible to clone the gene that codes for a particular protein, sequence its DNA, and deduce the structure of the protein from the nucleotide sequence of the DNA. We'll have more to say about DNA sequencing later in the chapter.

27.14 THE STRATEGY OF PEPTIDE SYNTHESIS

One way to confirm the structure proposed for a peptide is to synthesize a peptide having a specific sequence of amino acids and compare the two. This was done, for example, in the case of *bradykinin,* a peptide present in blood that acts to lower blood pressure. Excess bradykinin, formed as a response to the sting of wasps and other insects containing substances in their venom that stimulate bradykinin release, causes severe local pain. Bradykinin was originally believed to be an octapeptide containing two proline residues; however, a nonapeptide containing three prolines in the following sequence was synthesized and determined to be identical with natural bradykinin in every respect, including biological activity:

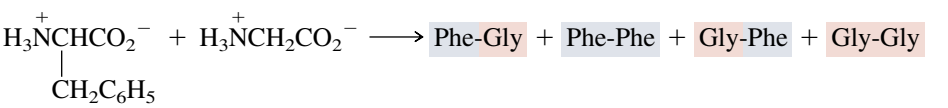
Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg

Bradykinin

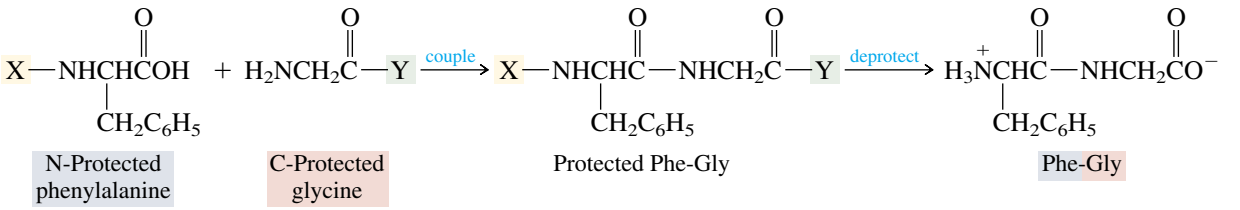
A reevaluation of the original sequence data established that natural bradykinin was indeed the nonapeptide shown. Here the synthesis of a peptide did more than confirm structure; synthesis was instrumental in determining structure.

Chemists and biochemists also synthesize peptides in order to better understand how they act. By systematically altering the sequence, it's sometimes possible to find out which amino acids are intimately involved in the reactions that involve a particular peptide. Many synthetic peptides have been prepared in searching for new drugs.

The objective in peptide synthesis may be simply stated: to connect amino acids in a prescribed sequence by amide bond formation between them. A number of very effective methods and reagents have been designed for peptide bond formation, so that the joining together of amino acids by amide linkages is not difficult. The real difficulty lies in ensuring that the correct sequence is obtained. This can be illustrated by considering the synthesis of a representative dipeptide, Phe-Gly. Random peptide bond formation in a mixture containing phenylalanine and glycine would be expected to lead to four dipeptides:



In order to direct the synthesis so that only Phe-Gly is formed, the amino group of phenylalanine and the carboxyl group of glycine must be protected so that they cannot react under the conditions of peptide bond formation. We can represent the peptide bond formation step by the following equation, where X and Y are amine- and carboxyl-protecting groups, respectively:



Thus, the synthesis of a dipeptide of prescribed sequence requires at least three operations:

1. *Protect* the amino group of the N-terminal amino acid and the carboxyl group of the C-terminal amino acid.
2. *Couple* the two protected amino acids by amide bond formation between them.
3. *Deprotect* the amino group at the N terminus and the carboxyl group at the C terminus.

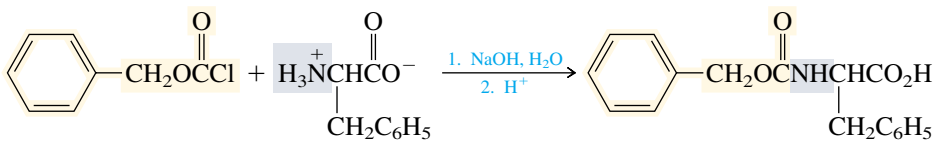
Higher peptides are prepared in an analogous way by a direct extension of the logic just outlined for the synthesis of dipeptides.

Sections 27.15 through 27.18 describe the chemistry associated with the protection and deprotection of amino and carboxyl functions, along with methods for peptide bond formation.

27.15 AMINO GROUP PROTECTION

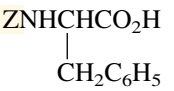
The reactivity of an amino group is suppressed by converting it to an amide, and amino groups are most often protected by acylation. The benzyloxycarbonyl group ($C_6H_5CH_2OC(=O)-$) is one of the most often used amino-protecting groups. It is attached by acylation of an amino acid with benzyloxycarbonyl chloride.

Another name for the benzyloxycarbonyl group is *carbobenzoxy*. This name, and its abbreviation *Cbz*, are often found in the older literature, but are no longer a part of IUPAC nomenclature.



PROBLEM 27.16 Lysine reacts with two equivalents of benzyloxycarbonyl chloride to give a derivative containing two benzyloxycarbonyl groups. What is the structure of this compound?

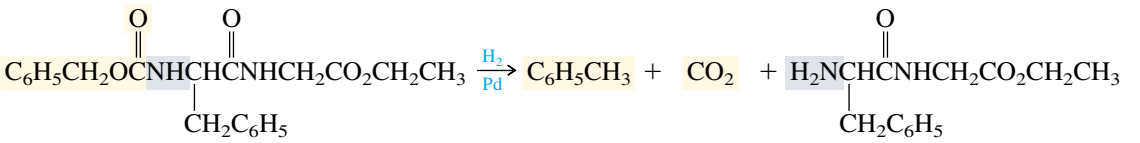
Just as it is customary to identify individual amino acids by abbreviations, so too with protected amino acids. The approved abbreviation for a benzyloxycarbonyl group is the letter *Z*. Thus, *N*-benzyloxycarbonylphenylalanine is represented as



or more simply as Z-Phe

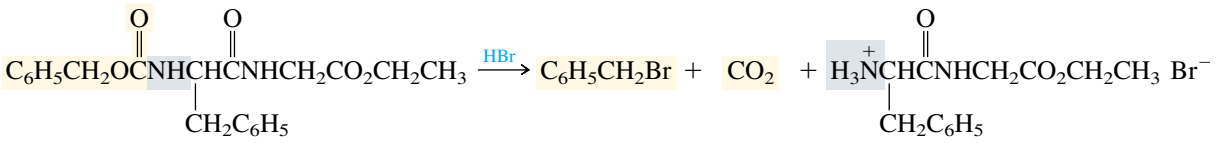
The value of the benzyloxycarbonyl protecting group is that it is easily removed by reactions other than hydrolysis. In peptide synthesis, amide bonds are formed. We protect the N terminus as an amide but need to remove the protecting group without cleaving the very amide bonds we labored so hard to construct. Removing the protecting group by hydrolysis would surely bring about cleavage of peptide bonds as well. One advantage that the benzyloxycarbonyl protecting group enjoys over more familiar acyl groups such as acetyl is that it can be removed by *hydrogenolysis* in the presence of palladium. The following equation illustrates this for the removal of the benzyloxycarbonyl protecting group from the ethyl ester of Z-Phe-Gly:

Hydrogenolysis refers to the cleavage of a molecule under conditions of catalytic hydrogenation.



N-Benzyloxycarbonylphenylalanylglycine ethyl ester — Toluene — Carbon dioxide — Phenylalanylglycine ethyl ester (100%)

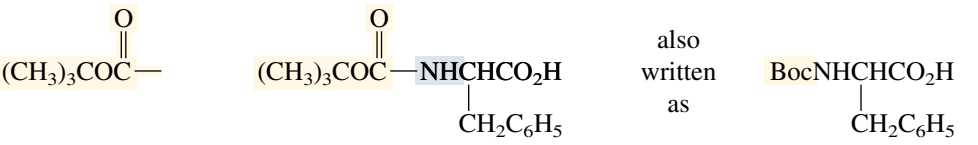
Alternatively, the benzyloxycarbonyl protecting group may be removed by treatment with hydrogen bromide in acetic acid:



N-Benzyloxycarbonylphenylalanylglycine ethyl ester — Benzyl bromide — Carbon dioxide — Phenylalanylglycine ethyl ester hydrobromide (82%)

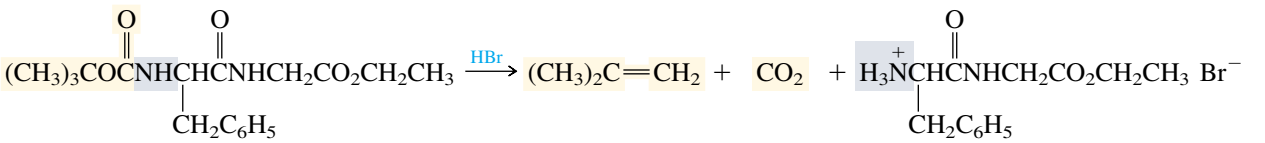
Deprotection by this method rests on the ease with which benzyl esters are cleaved by nucleophilic attack at the benzylic carbon in the presence of strong acids. Bromide ion is the nucleophile.

A related N-terminal-protecting group is *tert*-butoxycarbonyl, abbreviated *Boc:*



tert-Butoxycarbonyl (Boc-) — *N-tert*-Butoxycarbonylphenylalanine — also written as — Boc-Phe

Like the benzyloxycarbonyl protecting group, the Boc group may be removed by treatment with hydrogen bromide (it is stable to hydrogenolysis, however):



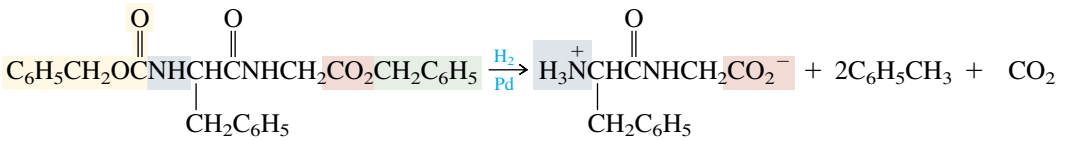
N-tert-Butoxycarbonylphenylalanylglycine ethyl ester | 2-Methylpropene | Carbon dioxide | Phenylalanylglycine ethyl ester hydrobromide (86%)

The *tert*-butyl group is cleaved as the corresponding carbocation. Loss of a proton from *tert*-butyl cation converts it to 2-methylpropene. Because of the ease with which a *tert*-butyl group is cleaved as a carbocation, other acidic reagents, such as trifluoroacetic acid, may also be used.

An experiment using Boc protection in the synthesis of a dipeptide can be found in the November 1989 issue of the *Journal of Chemical Education,* pp. 965–967.

27.16 CARBOXYL GROUP PROTECTION

Carboxyl groups of amino acids and peptides are normally protected as esters. Methyl and ethyl esters are prepared by Fischer esterification. Deprotection of methyl and ethyl esters is accomplished by hydrolysis in base. Benzyl esters are a popular choice because they can be removed by hydrogenolysis. Thus a synthetic peptide, protected at both its N terminus with a Z group and at its C terminus as a benzyl ester, can be completely deprotected in a single operation.



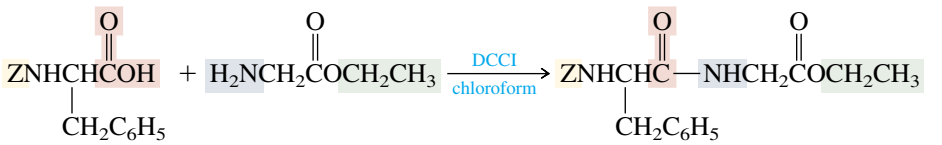
N-Benzyloxycarbonylphenylalanylglycine benzyl ester | Phenylalanylglycine (87%) | Toluene | Carbon dioxide

Several of the amino acids listed in Table 27.1 bear side-chain functional groups, which must also be protected during peptide synthesis. In most cases, protecting groups are available that can be removed by hydrogenolysis.

27.17 PEPTIDE BOND FORMATION

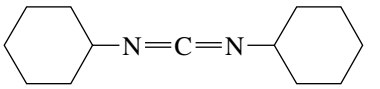
To form a peptide bond between two suitably protected amino acids, the free carboxyl group of one of them must be *activated* so that it is a reactive acylating agent. The most familiar acylating agents are acyl chlorides, and they were once extensively used to couple amino acids. Certain drawbacks to this approach, however, led chemists to seek alternative methods.

In one method, treatment of a solution containing the N-protected and the C-protected amino acids with *N,N′*-dicyclohexylcarbodiimide (DCCI) leads directly to peptide bond formation:



Z-Protected phenylalanine | Glycine ethyl ester | Z-Protected Phe-Gly ethyl ester (83%)

N,N′-Dicyclohexylcarbodiimide has the structure shown:



N,N′-Dicyclohexylcarbodiimide (DCCI)

The mechanism by which DCCI promotes the condensation of an amine and a carboxylic acid to give an amide is outlined in Figure 27.13.

PROBLEM 27.17 Show the steps involved in the synthesis of Ala-Leu from alanine and leucine using benzyloxycarbonyl and benzyl ester protecting groups and DCCI-promoted peptide bond formation.

In the second major method of peptide synthesis the carboxyl group is activated by converting it to an *active ester,* usually a *p*-nitrophenyl ester. Recall from Section 20.11 that esters react with ammonia and amines to give amides. *p*-Nitrophenyl esters are much more reactive than methyl and ethyl esters in these reactions because *p*-nitrophenoxide is a better (less basic) leaving group than methoxide and ethoxide. Simply allowing the active ester and a C-protected amino acid to stand in a suitable solvent is sufficient to bring about peptide bond formation by nucleophilic acyl substitution.

Z-Protected phenylalanine *p*-nitrophenyl ester

Glycine ethyl ester

Z-Protected Phe-Gly ethyl ester (78%)

p-Nitrophenol

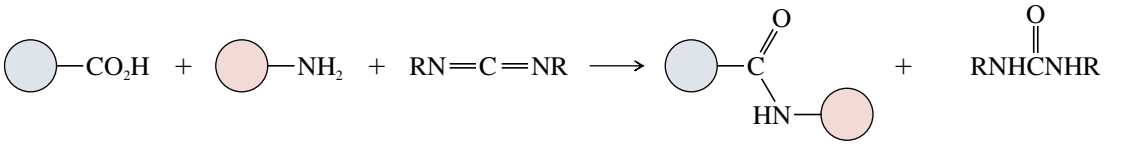
The *p*-nitrophenol formed as a byproduct in this reaction is easily removed by extraction with dilute aqueous base. Unlike free amino acids and peptides, protected peptides are not zwitterionic and are more soluble in organic solvents than in water.

PROBLEM 27.18 *p*-Nitrophenyl esters are made from Z-protected amino acids by reaction with *p*-nitrophenol in the presence of *N,N′*-dicyclohexylcarbodiimide. Suggest a reasonable mechanism for this reaction.

PROBLEM 27.19 Show how you could convert the ethyl ester of Z-Phe-Gly to Leu-Phe-Gly (as its ethyl ester) by the active ester method.

Higher peptides are prepared either by stepwise extension of peptide chains, one amino acid at a time, or by coupling of fragments containing several residues (the **fragment condensation** approach). Human pituitary adrenocorticotropic hormone (ACTH), for example, has 39 amino acids and was synthesized by coupling of smaller peptides containing residues 1–10, 11–16, 17–24, and 25–39. An attractive feature of this approach is that the various protected peptide fragments may be individually purified, which simplifies the purification of the final product. Among the substances that have been synthesized by fragment condensation are insulin (51 amino acids) and the protein ribonuclease A (124 amino acids). In the stepwise extension approach, the starting

Overall reaction:

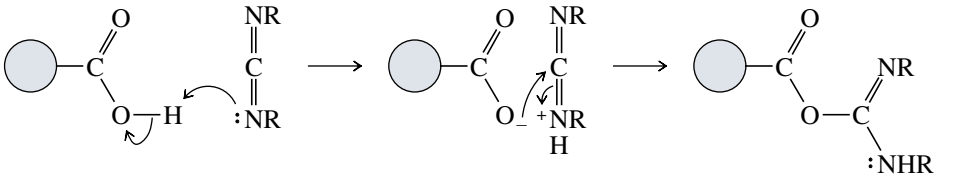


Carboxylic acid | Amine | DCCI | Amide | *N,N′*-Dicyclohexylurea

DCCI = *N,N′*-dicyclohexylcarbodiimide; R = cyclohexyl

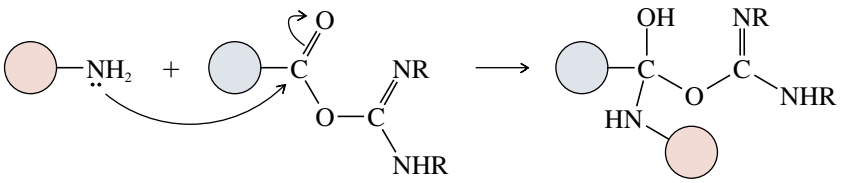
Mechanism:

Step 1: In the first stage of the reaction, the carboxylic acid adds to one of the double bonds of DCCI to give an *O*-acylisourea.



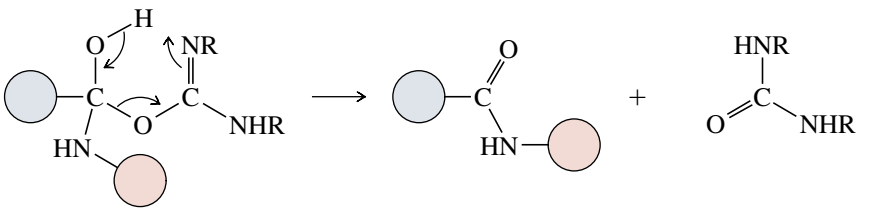
Carboxylic acid | DCCI | *O*-Acylisourea

Step 2: Structurally, *O*-acylisoureas resemble carboxylic acid anhydrides and are powerful acylating agents. In the reaction's second stage the amine adds to the carbonyl group of the *O*-acylisourea to give a tetrahedral intermediate.



Amine | *O*-Acylisourea | Tetrahedral intermediate

Step 3: The tetrahedral intermediate dissociates to an amide and *N,N′*-dicyclohexylurea.



Tetrahedral intermediate | Amide | *N,N′*-Dicyclohexylurea

FIGURE 27.13 The mechanism of amide bond formation by *N,N′*-dicyclohexylcarbodiimide-promoted condensation of a carboxylic acid and an amine.

peptide in a particular step differs from the coupling product by only one amino acid residue and the properties of the two peptides may be so similar as to make purification by conventional techniques all but impossible. The following section describes a method by which many of the difficulties involved in the purification of intermediates have been overcome.

27.18 SOLID-PHASE PEPTIDE SYNTHESIS: THE MERRIFIELD METHOD

Merrifield was awarded the 1984 Nobel Prize in chemistry for developing the solid-phase method of peptide synthesis.

In 1962, R. Bruce Merrifield of Rockefeller University reported the synthesis of the nonapeptide bradykinin (see Section 27.14) by a novel method. In Merrifield's method, peptide coupling and deprotection are carried out not in homogeneous solution but at the surface of an insoluble polymer, or *solid support.* Beads of a copolymer prepared from styrene containing about 2% divinylbenzene are treated with chloromethyl methyl ether and tin(IV) chloride to give a resin in which about 10% of the aromatic rings bear —CH_2Cl groups (Figure 27.14). The growing peptide is anchored to this polymer, and excess reagents, impurities, and byproducts are removed by thorough washing after each operation. This greatly simplifies the purification of intermediates.

The actual process of solid-phase peptide synthesis, outlined in Figure 27.15, begins with the attachment of the C-terminal amino acid to the chloromethylated polymer in step 1. Nucleophilic substitution by the carboxylate anion of an *N*-Boc-protected C-terminal amino acid displaces chloride from the chloromethyl group of the polymer to form an ester, protecting the C terminus while anchoring it to a solid support. Next, the Boc group is removed by treatment with acid (step 2), and the polymer containing the unmasked N terminus is washed with a series of organic solvents. Byproducts are removed, and only the polymer and its attached C-terminal amino acid residue remain. Next (step 3), a peptide bond to an *N*-Boc-protected amino acid is formed by condensation in the presence of *N,N′*-dicyclohexylcarbodiimide. Again, the polymer is washed thoroughly. The Boc-protecting group is then removed by acid treatment (step 4), and after washing, the polymer is now ready for the addition of another amino acid residue by a repetition of the cycle. When all the amino acids have been added, the synthetic peptide is removed from the polymeric support by treatment with hydrogen bromide in trifluoroacetic acid.

By successively adding amino acid residues to the C-terminal amino acid, it took Merrifield only 8 days to synthesize bradykinin in 68% yield. The biological activity of synthetic bradykinin was identical with that of natural material.

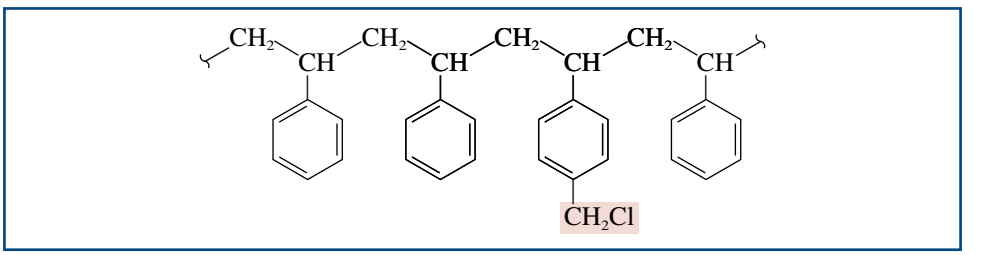


FIGURE 27.14 A section of polystyrene showing one of the benzene rings modified by chloromethylation. Individual polystyrene chains in the resin used in solid-phase peptide synthesis are connected to one another at various points (cross-linked) by adding a small amount of *p*-divinylbenzene to the styrene monomer. The chloromethylation step is carried out under conditions such that only about 10% of the benzene rings bear —CH_2Cl groups.

Step 1: The Boc-protected amino acid is anchored to the resin. Nucleophilic substitution of the benzylic chloride by the carboxylate anion gives an ester.

Step 2: The Boc protecting group is removed by treatment with hydrochloric acid in dilute acetic acid. After the resin has been washed, the C-terminal amino acid is ready for coupling.

Step 3: The resin-bound C-terminal amino acid is coupled to an N-protected amino acid by using *N,N′*-dicyclohexylcarbodiimide. Excess reagent and *N,N′*-dicyclohexylurea are washed away from the resin after coupling is complete.

Step 4: The Boc protecting group is removed as in step 2. If desired, steps 3 and 4 may be repeated to introduce as many amino acid residues as desired.

Step *n*: When the peptide is completely assembled, it is removed from the resin by treatment with hydrogen bromide in trifluoroacetic acid.

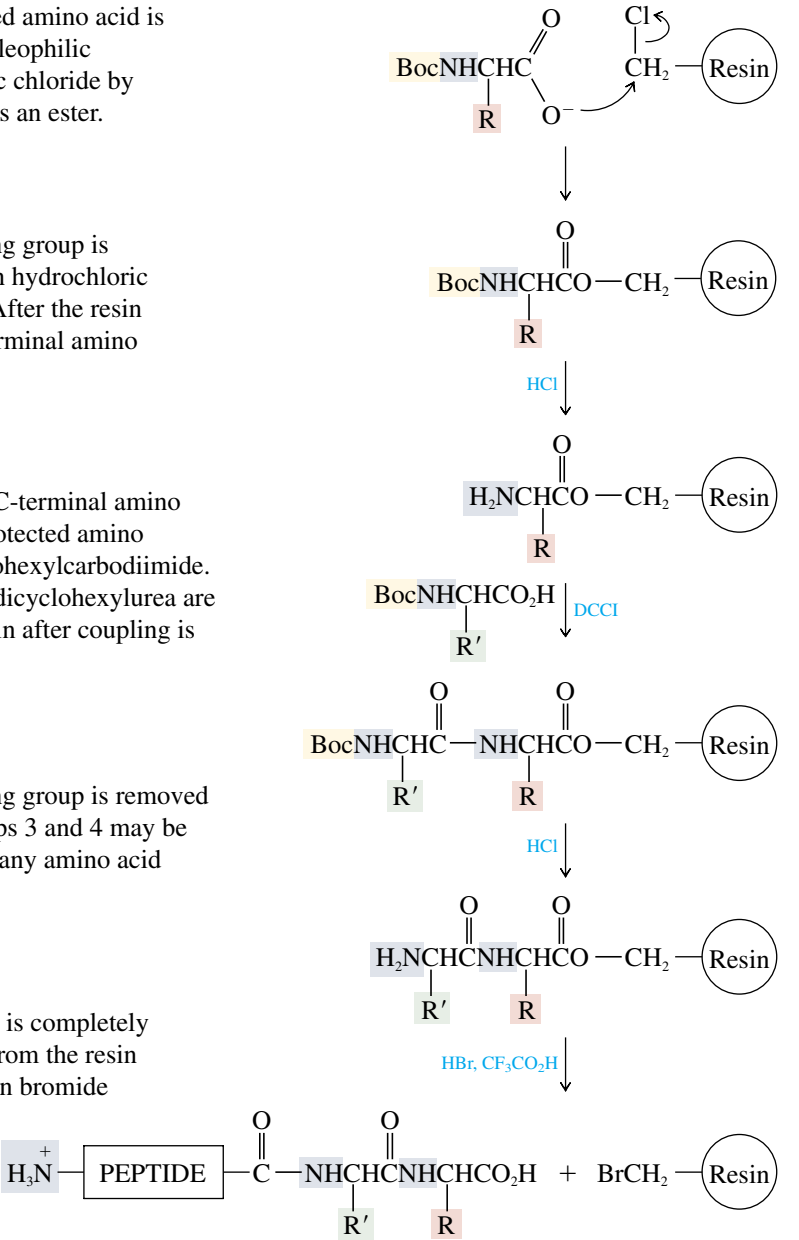


FIGURE 27.15 Peptide synthesis by the solid-phase method of Merrifield. Amino acid residues are attached sequentially beginning at the C terminus.

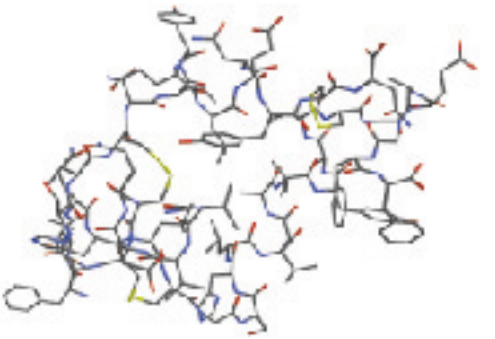
PROBLEM 27.20 Starting with phenylalanine and glycine, outline the steps in the preparation of Phe-Gly by the Merrifield method.

Merrifield successfully automated all the steps in solid-phase peptide synthesis, and computer-controlled equipment is now commercially available to perform this synthesis. Using an early version of his "peptide synthesizer," in collaboration with coworker Bernd Gutte, Merrifield reported the synthesis of the enzyme ribonuclease in 1969. It took them

only 6 weeks to perform the 369 reactions and 11,391 steps necessary to assemble the sequence of 124 amino acids of ribonuclease.

Solid-phase peptide synthesis does not solve all purification problems, however. Even if every coupling step in the ribonuclease synthesis proceeded in 99% yield, the product would be contaminated with many different peptides containing 123 amino acids, 122 amino acids, and so on. Thus, Merrifield and Gutte's 6 weeks of synthesis was followed by 4 months spent in purifying the final product. The technique has since been refined to the point that yields at the 99% level and greater are achieved with current instrumentation, and thousands of peptides and peptide analogs have been prepared by the solid-phase method.

Merrifield's concept of a solid-phase method for peptide synthesis and his development of methods for carrying it out set the stage for an entirely new way to do chemical reactions. Solid-phase synthesis has been extended to include numerous other classes of compounds and has helped spawn a whole new field called **combinatorial chemistry.** Combinatorial synthesis allows a chemist, using solid-phase techniques, to prepare hundreds of related compounds (called *libraries*) at a time. It is one of the most active areas of organic synthesis, especially in the pharmaceutical industry.



27.19 SECONDARY STRUCTURES OF PEPTIDES AND PROTEINS

The primary structure of a peptide is its amino acid sequence. We also speak of the **secondary structure** of a peptide, that is, the conformational relationship of nearest neighbor amino acids with respect to each other. On the basis of X-ray crystallographic studies and careful examination of molecular models, Linus Pauling and Robert B. Corey of the California Institute of Technology showed that certain peptide conformations were more stable than others. Two arrangements, the **α helix** and the **pleated β sheet,** stand out as secondary structural units that are both particularly stable and commonly encountered. Both of these incorporate two important features:

1. The geometry of the peptide bond is planar and the main chain is arranged in an anti conformation (Section 27.7).
2. Hydrogen bonding can occur when the N—H group of one amino acid unit and the C═O group of another are close in space; conformations that maximize the number of these hydrogen bonds are stabilized by them.

Figure 27.16 illustrates a β sheet structure for a protein composed of alternating glycine and alanine residues. There are hydrogen bonds between the C═O and H—N groups of adjacent antiparallel chains. Van der Waals repulsions between the α hydrogens

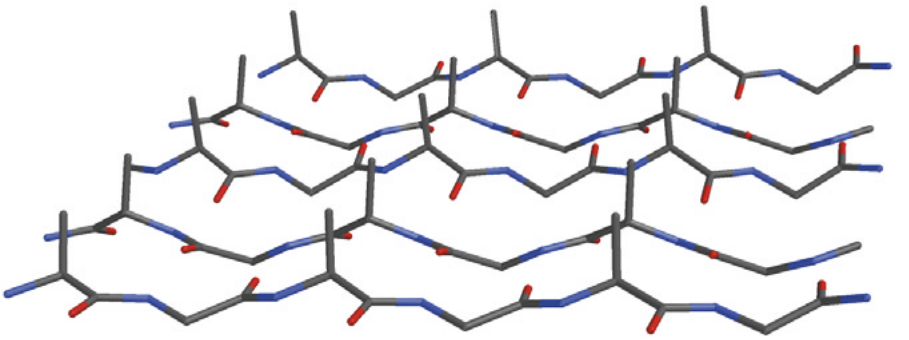


FIGURE 27.16 The β-sheet secondary structure of a protein, composed of alternating glycine and alanine residues. Hydrogen bonding occurs between the amide N—H of one chain and the carbonyl oxygen of another. Van der Waals repulsions between substituents at the α-carbon atoms, shown here as vertical methyl groups, introduces creases in the sheet. The structure of the pleated β sheet is seen more clearly by examining the molecular model on *Learning By Modeling* and rotating it in three dimensions.

of glycine and the methyl groups of alanine cause the chains to rotate with respect to one another to give a rippled effect. Hence the name *pleated β sheet.* The pleated β sheet is an important secondary structure, especially in proteins that are rich in amino acids with small side chains, such as H (glycine), CH_3 (alanine), and CH_2OH (serine). *Fibroin,* the major protein of most silk fibers, is almost entirely pleated β sheet, and over 80% of it is a repeating sequence of the six-residue unit -Gly-Ser-Gly-Ala-Gly-Ala-. The pleated β sheet is flexible, but since the peptide chains are nearly in an extended conformation, it resists stretching.

Unlike the pleated β sheet, in which hydrogen bonds are formed *between* two chains, the α *helix* is stabilized by hydrogen bonds *within* a single chain. Figure 27.17 illustrates a section of peptide α helix constructed from L-alanine. A right-handed helical conformation with about 3.6 amino acids per turn permits each carbonyl oxygen to be hydrogen-bonded to an amide proton and vice versa. The α helix is found in many proteins; the principal protein components of muscle *(myosin)* and wool *(α-keratin)*, for example, contain high percentages of α helix. When wool fibers are stretched, these helical regions are elongated by the breaking of hydrogen bonds. Disulfide bonds between cysteine residues of neighboring α-keratin chains are too strong to be broken during stretching, however, and they limit the extent of distortion. After the stretching force is removed, the hydrogen bonds reform spontaneously, and the wool fiber returns to its original shape. Wool has properties that are different from those of silk because the secondary structures of the two fibers are different, and their secondary structures are different because the primary structures are different.

Proline is the only amino acid in Table 27.1 that is a secondary amine, and its presence in a peptide chain introduces an amide nitrogen that has no hydrogen available for hydrogen bonding. This disrupts the network of hydrogen bonds and divides the peptide into two separate regions of α helix. The presence of proline is often associated with a bend in the peptide chain.

Proteins, or sections of proteins, sometimes exist as **random coils,** an arrangement that lacks the regularity of the α helix or pleated β sheet.

FIGURE 27.17 An α helix of a portion of a protein in which all of the amino acids are alanine. The helix is stabilized by hydrogen bonds between the N—H proton of one amide group and the carbonyl oxygen of another. The methyl groups at the α carbon project away from the outer surface of the helix. When viewed along the helical axis, the chain turns in a clockwise direction (a right-handed helix). The structure of the α helix is seen more clearly by examining the molecular model on *Learning By Modeling* and rotating it in three dimensions.

27.20 TERTIARY STRUCTURE OF PEPTIDES AND PROTEINS

The **tertiary structure** of a peptide or protein refers to the folding of the chain. The way the chain is folded affects both the physical properties of a protein and its biological function. Structural proteins, such as those present in skin, hair, tendons, wool, and silk, may have either helical or pleated-sheet secondary structures, but in general are elongated in shape, with a chain length many times the chain diameter. They are classed as *fibrous* proteins and, as befits their structural role, tend to be insoluble in water. Many other proteins, including most enzymes, operate in aqueous media; some are soluble, but most are dispersed as colloids. Proteins of this type are called *globular* proteins. Globular proteins are approximately spherical. Figure 27.18 shows carboxypeptidase A (Section 27.10), a globular protein containing 307 amino acids. A typical protein such as carboxypeptidase A incorporates elements of a number of secondary structures: some segments are helical; others, pleated sheet; and still others correspond to no simple description.

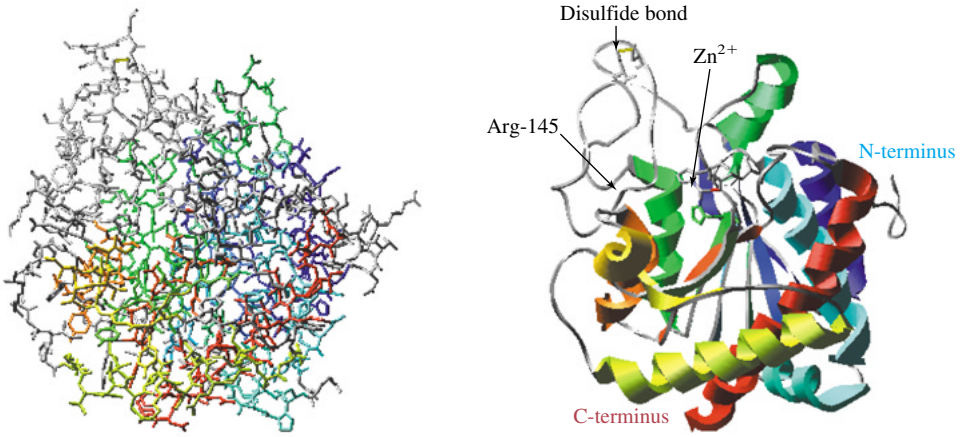


FIGURE 27.18 The structure of carboxypeptidase A displayed as (*a*) a tube model and (*b*) a ribbon diagram. The tube model shows all of the amino acids and their side chains. The most evident feature illustrated by (*a*) is the globular shape of the enzyme. The ribbon diagram emphasizes the folding of the chain and the helical regions. As can be seen in (*b*), a substantial portion of the protein, the sections colored gray, is not helical but is random coil. The orientation of the protein and the color-coding are the same in both views.

The shape of a large protein is influenced by many factors, including, of course, its primary and secondary structure. The disulfide bond shown in Figure 27.18 links Cys-138 of carboxypeptidase A to Cys-161 and contributes to the tertiary structure. Carboxypeptidase A contains a Zn^{2+} ion, which is essential to the catalytic activity of the enzyme, and its presence influences the tertiary structure. The Zn^{2+} ion lies near the center of the enzyme, where it is coordinated to the imidazole nitrogens of two histidine residues (His-69, His-196) and to the carboxylate side chain of Glu-72.

Protein tertiary structure is also influenced by the environment. In water a globular protein usually adopts a shape that places its lipophilic groups toward the interior, with its polar groups on the surface, where they are solvated by water molecules. About 65% of the mass of most cells is water, and the proteins present in cells are said to be in their *native state*—the tertiary structure in which they express their biological activity. When the tertiary structure of a protein is disrupted by adding substances that cause the protein chain to unfold, the protein becomes *denatured* and loses most, if not all, of its activity. Evidence that supports the view that the tertiary structure is dictated by the primary structure includes experiments in which proteins are denatured and allowed to stand, whereupon they are observed to spontaneously readopt their native-state conformation with full recovery of biological activity.

Most protein tertiary structures are determined by X-ray crystallography. The first, myoglobin, the oxygen storage protein of muscle, was determined in 1957. Since then thousands more have been determined. In the form of crystallographic coordinates, the data are deposited in the **Protein Data Bank** and are freely available. The three-dimensional structure of carboxypeptidase in Figure 27.18, for example, was produced by downloading the coordinates from the Protein Data Bank and converting them to a molecular model. At present, the Protein Data Bank averages about one new protein structure per day.

For their work on myoglobin and hemoglobin, respectively, John C. Kendrew and Max F. Perutz were awarded the 1962 Nobel Prize in chemistry.

Knowing how the protein chain is folded is a key ingredient in understanding the mechanism by which an enzyme catalyzes a reaction. Take carboxypeptidase for example. This enzyme catalyzes the hydrolysis of the peptide bond at the C terminus. It is believed that an ionic bond between the positively charged side chain of an arginine residue (Arg-145) of the enzyme and the negatively charged carboxylate group of the substrate's terminal amino acid binds the peptide at the **active site,** the region of the enzyme's interior where the catalytically important functional groups are located. There,

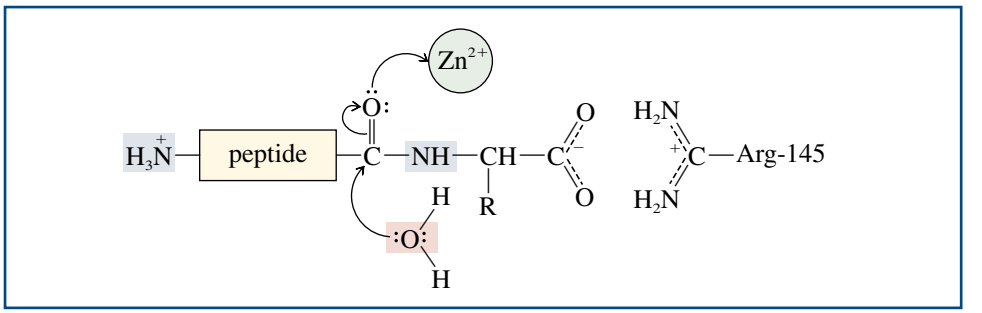


FIGURE 27.19 Proposed mechanism of hydrolysis of a peptide catalyzed by carboxypeptidase A. The peptide is bound at the active site by an ionic bond between its C-terminal amino acid and the positively charged side chain of arginine-145. Coordination of Zn^{2+} to oxygen makes the carbon of the carbonyl group more positive and increases the rate of nucleophilic attack by water.

the Zn^{2+} ion acts as a Lewis acid toward the carbonyl oxygen of the peptide substrate, increasing its susceptibility to attack by a water molecule (Figure 27.19).

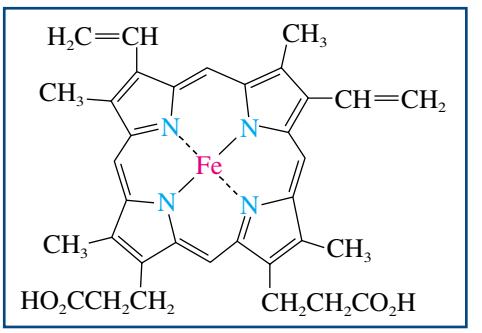
Living systems contain thousands of different enzymes. As we have seen, all are structurally quite complex, and there are no sweeping generalizations that can be made to include all aspects of enzymic catalysis. The case of carboxypeptidase A illustrates one mode of enzyme action, the bringing together of reactants and catalytically active functions at the active site.

Almost, but not all enzymes are proteins. For identifying certain RNA-catalyzed biological processes Sidney Altman (Yale University) and Thomas R. Cech (University of Colorado) shared the 1989 Nobel Prize in chemistry.

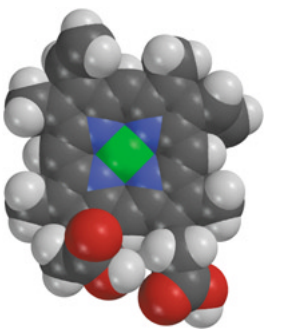
27.21 COENZYMES

The number of chemical processes that protein side chains can engage in is rather limited. Most prominent among them are proton donation, proton abstraction, and nucleophilic addition to carbonyl groups. In many biological processes a richer variety of reactivity is required, and proteins often act in combination with nonprotein organic molecules to bring about the necessary chemistry. These "helper molecules," referred to as **coenzymes, cofactors,** or **prosthetic groups,** interact with both the enzyme and the substrate to produce the necessary chemical change. Acting alone, for example, proteins lack the necessary functionality to be effective oxidizing or reducing agents. They can catalyze biological oxidations and reductions, however, in the presence of a suitable coenzyme. In earlier sections we saw numerous examples of these reactions in which the coenzyme NAD^+ acted as an oxidizing agent, and others in which NADH acted as a reducing agent.

Heme (Figure 27.20) is an important prosthetic group in which iron(II) is coordinated with the four nitrogen atoms of a type of tetracyclic aromatic substance known as



(*a*)



(*b*)

FIGURE 27.20 Heme shown as (*a*) a structural drawing and as (*b*) a space-filling model. The space-filling model shows the coplanar arrangement of the groups surrounding iron.

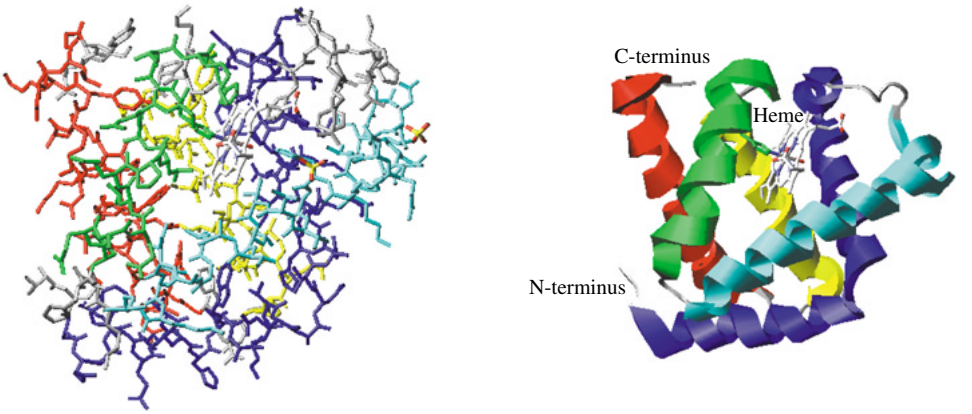


FIGURE 27.21 The structure of sperm-whale myoglobin displayed as (*a*) a tube model and (*b*) a ribbon diagram. The tube model shows all of the amino acids in the chain; the ribbon diagram shows the folding of the chain. There are five separate regions of α-helix in myoglobin which are shown in different colors to show them more clearly. The heme portion is included in both drawings, but is easier to locate in the ribbon diagram, as is the histidine side chain that is attached to the iron of heme.

a *porphyrin.* The oxygen-storing protein of muscle, myoglobin, represented schematically in Figure 27.21, consists of a heme group surrounded by a protein of 153 amino acids. Four of the six available coordination sites of Fe^{2+} are taken up by the nitrogens of the porphyrin, one by a histidine residue of the protein, and the last by a water molecule. Myoglobin stores oxygen obtained from the blood by formation of an $Fe—O_2$ complex. The oxygen displaces water as the sixth ligand on iron and is held there until needed. The protein serves as a container for the heme and prevents oxidation of Fe^{2+} to Fe^{3+}, an oxidation state in which iron lacks the ability to bind oxygen. Separately, neither heme nor the protein binds oxygen in aqueous solution; together, they do it very well.

27.22 PROTEIN QUATERNARY STRUCTURE: HEMOGLOBIN

Rather than existing as a single polypeptide chain, some proteins are assemblies of two or more chains. The manner in which these subunits are organized is called the **quaternary structure** of the protein.

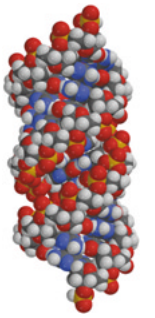
Hemoglobin is the oxygen-carrying protein of blood. It binds oxygen at the lungs and transports it to the muscles, where it is stored by myoglobin. Hemoglobin binds oxygen in very much the same way as myoglobin, using heme as the prosthetic group. Hemoglobin is much larger than myoglobin, however, having a molecular weight of 64,500, whereas that of myoglobin is 17,500; hemoglobin contains four heme units, myoglobin only one. Hemoglobin is an assembly of four hemes and four protein chains, including two identical chains called the *alpha chains* and two identical chains called the *beta chains.*

An article entitled "Hemoglobin: Its Occurrence, Structure, and Adaptation" appeared in the March 1982 issue of the *Journal of Chemical Education* (pp. 173–178).

Some substances, such as CO, form strong bonds to the iron of heme, strong enough to displace O_2 from it. Carbon monoxide binds 30–50 times more effectively than oxygen to myoglobin and hundreds of times better than oxygen to hemoglobin. Strong binding of CO at the active site interferes with the ability of heme to perform its biological task of transporting and storing oxygen, with potentially lethal results.

How function depends on structure can be seen in the case of the genetic disorder *sickle cell anemia.* This is a debilitating, sometimes fatal, disease in which red blood cells become distorted ("sickle-shaped") and interfere with the flow of blood through the capillaries. This condition results from the presence of an abnormal hemoglobin in affected people. The primary structures of the beta chain of normal and sickle cell hemoglobin differ by a single amino acid out of 149; sickle cell hemoglobin has valine in

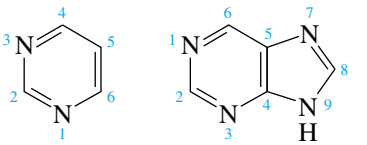
place of glutamic acid as the sixth residue from the N terminus. A tiny change in amino acid sequence can produce a life-threatening result! This modification is genetically controlled and probably became established in the gene pool because bearers of the trait have an increased resistance to malaria.



27.23 PYRIMIDINES AND PURINES

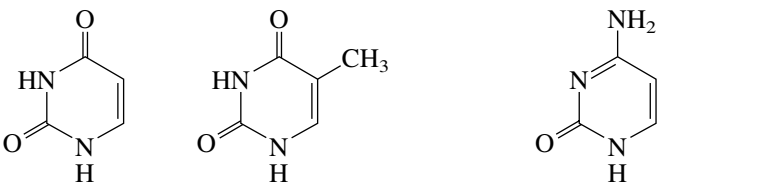
One of the major achievements in all of science has been the identification, at the molecular level, of the chemical interactions that are involved in the transfer of genetic information and the control of protein biosynthesis. The substances involved are biological macromolecules called **nucleic acids.** Nucleic acids were isolated over 100 years ago, and, as their name implies, they are acidic substances present in the nuclei of cells. There are two major kinds of nucleic acids: ribonucleic acid (RNA) and deoxyribonucleic acid (DNA). To understand the complex structure of nucleic acids, we first need to examine some simpler substances, nitrogen-containing aromatic heterocycles called *pyrimidines* and *purines.* The parent substance of each class and the numbering system used are shown:

Recall that heterocyclic aromatic compounds were introduced in Section 11.21.



Pyrimidine Purine

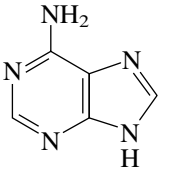
The pyrimidines that occur in DNA are cytosine and thymine. Cytosine is also a structural unit in RNA, which, however, contains uracil instead of thymine. Other pyrimidine derivatives are sometimes present but in small amounts.



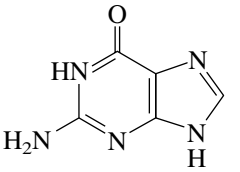
Uracil (occurs in RNA) Thymine (occurs in DNA) Cytosine (occurs in both RNA and DNA)

PROBLEM 27.21 5-Fluorouracil is a drug used in cancer chemotherapy. What is its structure?

Adenine and guanine are the principal purines of both DNA and RNA.



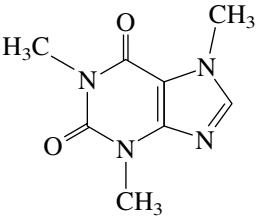
Adenine



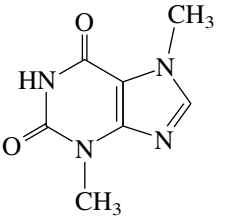
Guanine

The rings of purines and pyrimidines are aromatic and planar. You will see how important this flat shape is when we consider the structure of nucleic acids.

Pyrimidines and purines occur naturally in substances other than nucleic acids. Coffee, for example, is a familiar source of caffeine. Tea contains both caffeine and theobromine.



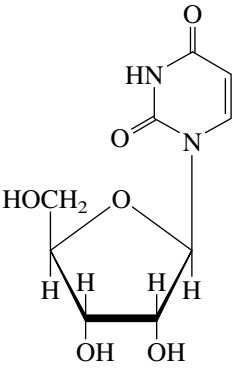
Caffeine



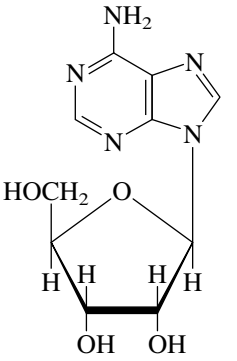
Theobromine

27.24 NUCLEOSIDES

The term **nucleoside** was once restricted to pyrimidine and purine *N*-glycosides of D-ribofuranose and 2-deoxy-D-ribofuranose, because these are the substances present in nucleic acids. The term is used more liberally now with respect to the carbohydrate portion, but is still usually limited to pyrimidine and purine substituents at the anomeric carbon. *Uridine* is a representative pyrimidine nucleoside; it bears a D-ribofuranose group at N-1. *Adenosine* is a representative purine nucleoside; its carbohydrate unit is attached at N-9.



Uridine
(1-β-D-ribofuranosyluracil)



Adenosine
(9-β-D-ribofuranosyladenine)

It is customary to refer to the noncarbohydrate portion of a nucleoside as a purine or pyrimidine *base*.

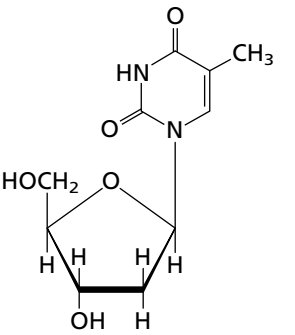
PROBLEM 27.22 The names of the principal nucleosides obtained from RNA and DNA are listed. Write a structural formula for each one.

(a) Thymidine (thymine-derived nucleoside in DNA)

(b) Cytidine (cytosine-derived nucleoside in RNA)

(c) Guanosine (guanine-derived nucleoside in RNA)

SAMPLE SOLUTION (a) Thymine is a pyrimidine base present in DNA; its carbohydrate substituent is 2-deoxyribofuranose, which is attached to N-1 of thymine.

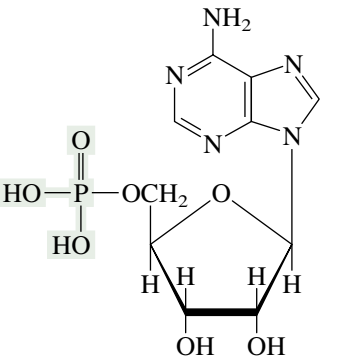


Thymidine

Nucleosides of 2-deoxyribose are named in the same way. Carbons in the carbohydrate portion of the molecule are identified as 1′, 2′, 3′, 4′, and 5′ to distinguish them from atoms in the purine or pyrimidine base. Thus, the adenine nucleoside of 2-deoxyribose is called 2′-deoxyadenosine or 9-β-2′-deoxyribofuranosyladenine.

27.25 NUCLEOTIDES

Nucleotides are phosphoric acid esters of nucleosides. The 5′-monophosphate of adenosine is called *5′-adenylic acid* or *adenosine 5′-monophosphate* (AMP).

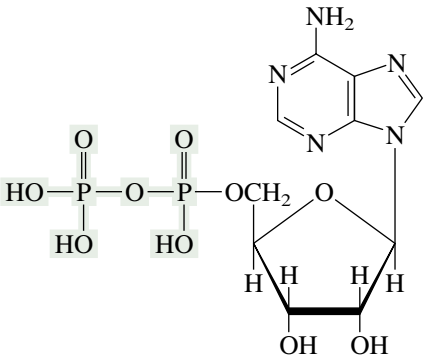


5′-Adenylic acid (AMP)

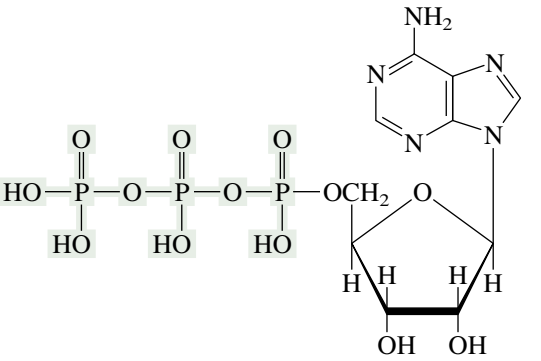
As its name implies, 5′-adenylic acid is an acidic substance; it is a diprotic acid with pK_a's for ionization of 3.8 and 6.2, respectively. In aqueous solution at pH 7, both OH groups of the $P(O)(OH)_2$ unit are ionized.

The analogous D-ribonucleotides of the other purines and pyrimidines are *uridylic acid, guanylic acid,* and *cytidylic acid. Thymidylic acid* is the 5′-monophosphate of thymidine (the carbohydrate is 2-deoxyribose in this case).

Other important 5′-nucleotides of adenosine include *adenosine diphosphate* (ADP) and *adenosine triphosphate* (ATP):



Adenosine diphosphate (ADP)



Adenosine triphosphate (ATP)

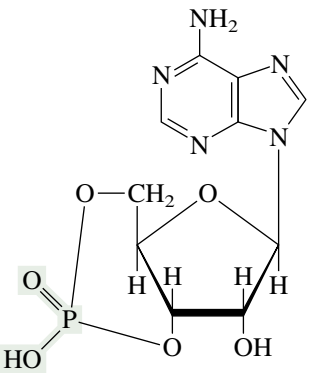
Each phosphorylation step in the sequence shown is endothermic:

$$\text{Adenosine} \xrightarrow[\text{enzymes}]{PO_4^{3-}} \text{AMP} \xrightarrow[\text{enzymes}]{PO_4^{3-}} \text{ADP} \xrightarrow[\text{enzymes}]{PO_4^{3-}} \text{ATP}$$

The energy to drive each step comes from carbohydrates by the process of glycolysis. It is convenient to view ATP as the storage vessel for the energy released during conversion of carbohydrates to carbon dioxide and water. That energy becomes available to the cells when ATP undergoes hydrolysis. The hydrolysis of ATP to ADP and phosphate has a $\Delta G°$ value of $-$ 35 kJ/mol ($-$ 8.4 kcal/mol).

For a discussion of glycolysis, see the July 1986 issue of the *Journal of Chemical Education* (pp. 566–570).

Adenosine 3′-5′-cyclic monophosphate *(cyclicAMP* or *cAMP)* is an important regulator of a large number of biological processes. It is a cyclic ester of phosphoric acid and adenosine involving the hydroxyl groups at C-3′ and C-5′.



Adenosine 3′-5′-cyclic monophosphate (cAMP)

27.26 NUCLEIC ACIDS

Nucleic acids are **polynucleotides** in which a phosphate ester unit links the 5′ oxygen of one nucleotide to the 3′ oxygen of another. Figure 27.22 is a generalized depiction of the structure of a nucleic acid. Nucleic acids are classified as ribonucleic acids (RNA) or deoxyribonucleic acids (DNA) depending on the carbohydrate present.

Research on nucleic acids progressed slowly until it became evident during the 1940s that they played a role in the transfer of genetic information. It was known that

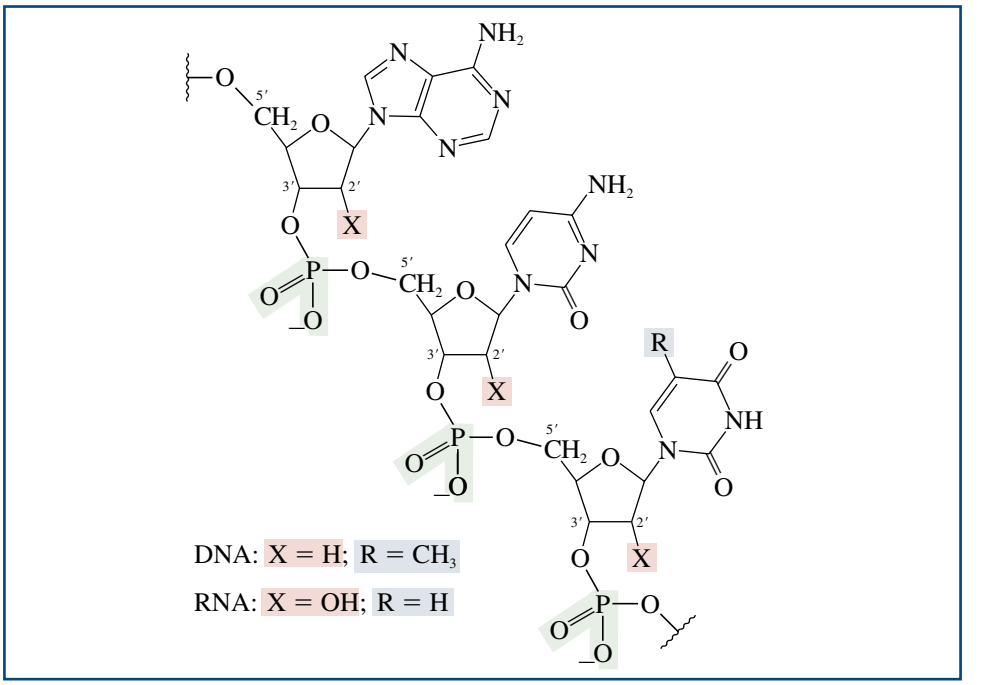


FIGURE 27.22 A portion of a polynucleotide chain.

the genetic information of an organism resides in the chromosomes present in each of its cells and that individual chromosomes are made up of smaller units called *genes.* When it became apparent that genes are DNA, interest in nucleic acids intensified. There was a feeling that once the structure of DNA was established, the precise way in which it carried out its designated role would become more evident. In some respects the problems are similar to those of protein chemistry. Knowing that DNA is a polynucleotide is comparable with knowing that proteins are polyamides. What is the nucleotide sequence (primary structure)? What is the precise shape of the polynucleotide chain (secondary and tertiary structure)? Is the genetic material a single strand of DNA, or is it an assembly of two or more strands? The complexity of the problem can be indicated by noting that a typical strand of human DNA contains approximately 10^8 nucleotides; if uncoiled it would be several centimeters long, yet it and many others like it reside in cells too small to see with the naked eye.

Watson and Crick have each written accounts of their work, and both are well worth reading. Watson's is entitled *The Double Helix.* Crick's is *What Mad Pursuit: A Personal View of Scientific Discovery.*

In 1953 James D. Watson and Francis H. C. Crick pulled together data from biology, biochemistry, chemistry, and X-ray crystallography, along with the insight they gained from molecular models, to propose a structure for DNA and a mechanism for its replication. Their two brief papers paved the way for an explosive growth in our understanding of life processes at the molecular level, the field we now call *molecular biology.* Along with Maurice Wilkins, who was responsible for the X-ray crystallographic work, Watson and Crick shared the 1962 Nobel Prize in physiology or medicine.

27.27 STRUCTURE AND REPLICATION OF DNA: THE DOUBLE HELIX

Watson and Crick were aided in their search for the structure of DNA by a discovery made by Erwin Chargaff (Columbia University). Chargaff found that there was a consistent pattern in the composition of DNAs from various sources. Although there was a wide variation in the distribution of the bases among species, half the bases in all samples

of DNA were purines and the other half were pyrimidines. Furthermore, the ratio of the purine adenine (A) to the pyrimidine thymine (T) was always close to 1:1. Likewise, the ratio of the purine guanine (G) to the pyrimidine cytosine (C) was also close to 1:1. Analysis of human DNA, for example, revealed it to have the following composition:

Purine	Pyrimidine	Base ratio
Adenine (A) 30.3%	Thymine (T) 30.3%	A/T = 1.00
Guanine (G) 19.5%	Cytosine (C) 19.9%	G/C = 0.98
Total purines 49.8%	Total pyrimidines 50.1%	

Feeling that the constancy in the A/T and G/C ratios was no accident, Watson and Crick proposed that it resulted from a structural complementarity between A and T and between G and C. Consideration of various hydrogen bonding arrangements revealed that A and T could form the hydrogen-bonded *base pair* shown in Figure 27.23*a* and that G and C could associate as in Figure 27.23*b*. Specific base pairing of A to T and of G to C by hydrogen bonds is a key element in the Watson–Crick model for the structure of DNA. We shall see that it is also a key element in the replication of DNA.

Because each hydrogen-bonded base pair contains one purine and one pyrimidine, A---T and G---C are approximately the same size. Thus, two nucleic acid chains may be aligned side by side with their bases in the middle, as illustrated in Figure 27.24. The two chains are joined by the network of hydrogen bonds between the paired bases A---T and G---C. Since X-ray crystallographic data indicated a helical structure, Watson and Crick proposed that the two strands are intertwined as a **double helix** (Figure 27.25).

The Watson–Crick base pairing model for DNA structure holds the key to understanding the process of DNA **replication.** During cell division a cell's DNA is duplicated, that in the new cell being identical with that in the original cell. At one stage of cell division the DNA double helix begins to unwind, separating the two chains. As portrayed in Figure 27.26, each strand serves as the template on which a new DNA strand is constructed. Each new strand is exactly like the original partner because the A---T, G---C base pairing requirement ensures that the new strand is the precise complement of the template, just as the old strand was. As the double helix unravels, each strand becomes one half of a new and identical DNA double helix.

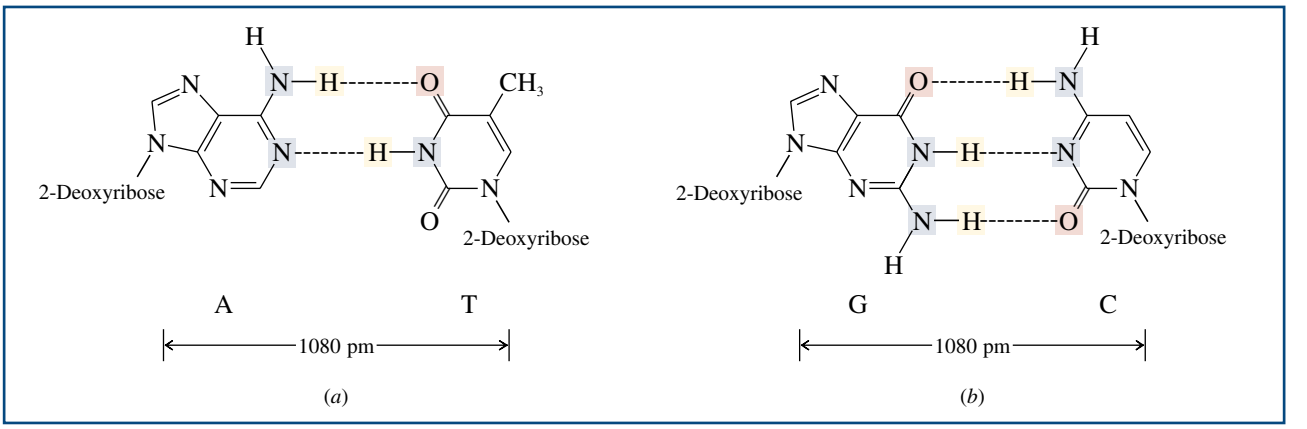


FIGURE 27.23 Base pairing between (*a*) adenine and thymine and (*b*) guanine and cytosine.

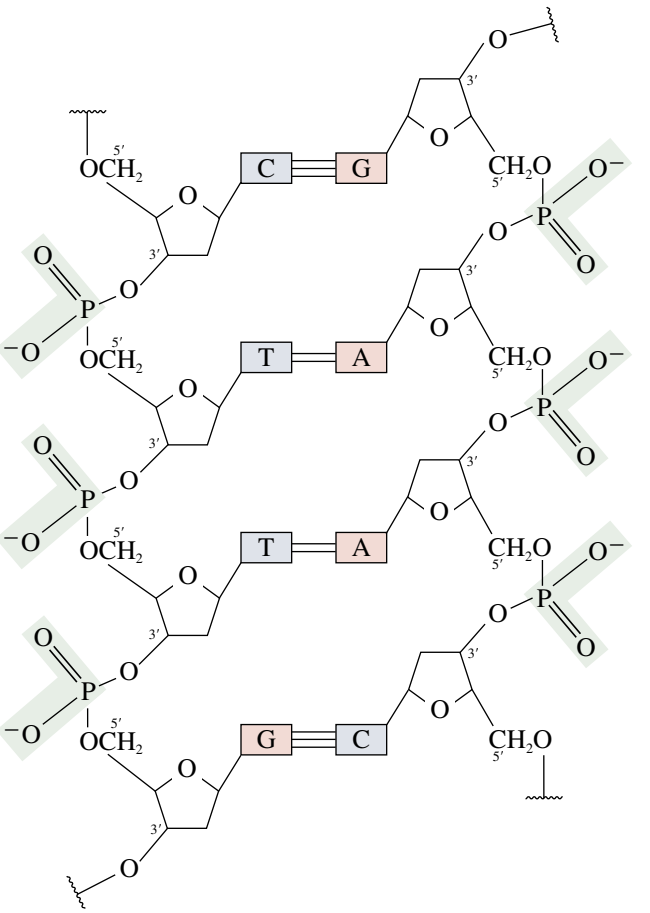


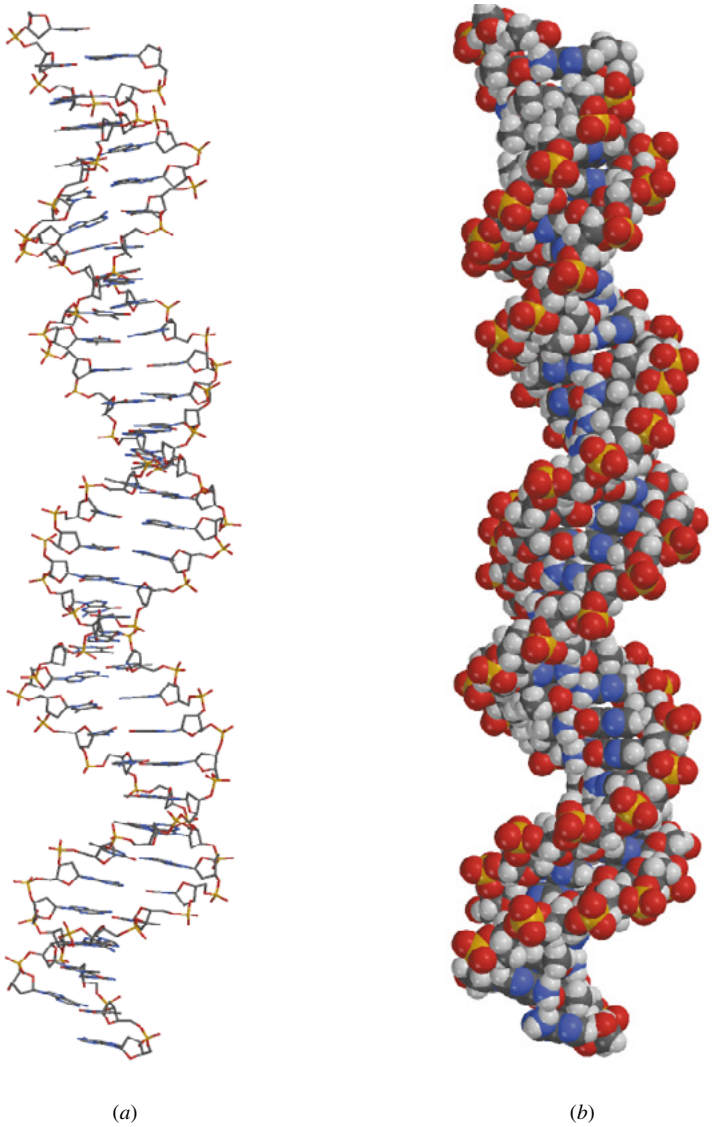
FIGURE 27.24 Hydrogen bonds between complementary bases (A and T, and G and C) permit pairing of two DNA strands. The strands are antiparallel; the 5′ end of the left strand is at the top, while the 5′ end of the right strand is at the bottom.

The structural requirements for the pairing of nucleic acid bases are also critical for utilizing genetic information, and in living systems this means protein biosynthesis.

27.28 DNA-DIRECTED PROTEIN BIOSYNTHESIS

Protein biosynthesis is directed by DNA through the agency of several types of ribonucleic acid called *messenger RNA (mRNA), transfer RNA (tRNA),* and *ribosomal RNA (rRNA).* There are two main stages in protein biosynthesis: **transcription** and **translation.**

In the transcription stage a molecule of mRNA having a nucleotide sequence complementary to one of the strands of a DNA double helix is constructed. A diagram illustrating transcription is presented in Figure 27.27 on page 1099. Transcription begins at the 5′ end of the DNA molecule, and ribonucleotides with bases complementary to the DNA bases are polymerized with the aid of the enzyme *RNA polymerase.* Thymine does not occur in RNA; the base that pairs with adenine in RNA is uracil. Unlike DNA, RNA is single-stranded.



(*a*) (*b*)

FIGURE 27.25 Tube (*a*) and space-filling (*b*) models of a DNA double helix. The carbohydrate–phosphate "backbone" is on the outside and can be roughly traced in (*b*) by the red oxygen atoms. The blue atoms belong to the purine and pyrimidine bases and lie on the inside. The base-pairing is more clearly seen in (*a*).

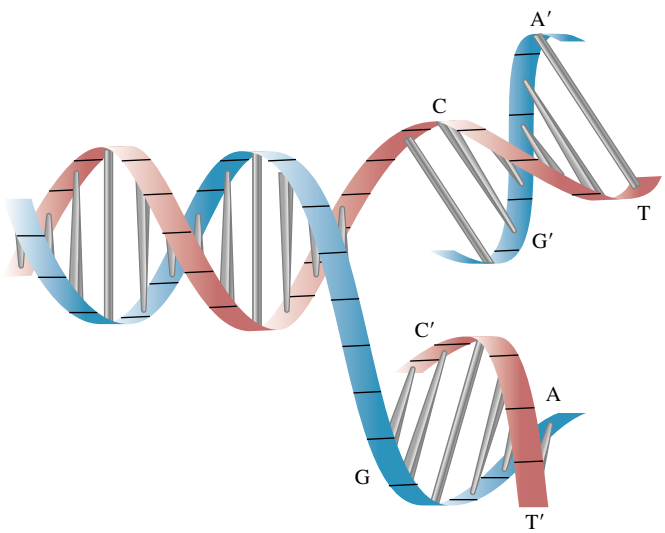


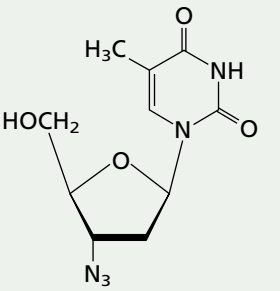
FIGURE 27.26 During DNA replication the double helix unwinds, and each of the original strands serves as a template for the synthesis of its complementary strand.

AIDS

The explosive growth of our knowledge of nucleic acid chemistry and its role in molecular biology in the 1980s happened to coincide with a challenge to human health that would have defied understanding a generation ago. That challenge is *acquired immune deficiency syndrome,* or **AIDS.** AIDS is a condition in which the body's immune system is devastated by a viral infection to the extent that it can no longer perform its vital function of identifying and destroying invading organisms. AIDS victims often die from "opportunistic" infections—diseases that are normally held in check by a healthy immune system but which can become deadly when the immune system is compromised. In the short time since its discovery, AIDS has claimed the lives of over 11 million people worldwide, and the most recent estimates place the number of those infected at more than 30 million.

The virus responsible for almost all the AIDS cases in the United States was identified by scientists at the Louis Pasteur Institute in Paris in 1983 and is known as *human immunodeficiency virus 1* (HIV-1). HIV-1 is believed to have originated in Africa, where a related virus, HIV-2, was discovered in 1986 by the Pasteur Institute group. Both HIV-1 and HIV-2 are classed as **retroviruses,** because their genetic material is RNA rather than DNA. HIVs require a host cell to reproduce, and the hosts in humans are the so-called T4 lymphocytes, which are the cells primarily responsible for inducing the immune system to respond when provoked. The HIV penetrates the cell wall of a T4 lymphocyte and deposits both its RNA and an enzyme called *reverse transcriptase* inside the T4 cell, where the reverse transcriptase catalyzes the formation of a DNA strand that is complementary to the viral RNA. The transcribed DNA then serves as the template for formation of double-helical DNA, which, with the information it carries for reproduction of the HIV, becomes incorporated into the T4 cell's own genetic material. The viral DNA induces the host lymphocyte to begin producing copies of the virus, which then leave the host to infect other T4 cells. In the course of HIV reproduction, the ability of the T4 lymphocyte to reproduce itself is hampered. As the number of T4 cells decrease, so does the body's ability to combat infections.

At this time, there is no known cure for AIDS, but progress is being made in delaying the onset of symptoms and prolonging the lives of those infected with HIV. The first advance in treatment came with drugs such as zidovudine, also known as azidothymine, or AZT. AZT interferes with the ability of HIV to reproduce by blocking the action of reverse transcriptase. As seen by its structure



Zidovudine (AZT)

AZT is a nucleoside. Several other nucleosides that are also reverse transcriptase inhibitors are in clinical use as well, sometimes in combination with AZT as "drug cocktails." A mixture makes it more difficult for a virus to develop resistance than a single drug does.

The most recent advance has been to simultaneously attack HIV on a second front using a protease inhibitor. Recall from Section 27.10 that proteases are enzymes that catalyze the hydrolysis of proteins at specific points. When HIV uses a cell's DNA to synthesize its own proteins, those proteins are in a form that must be modified by protease-catalyzed hydrolysis to become useful. Protease inhibitors prevent this modification and, in combination with reverse transcriptase inhibitors, slow the reproduction of HIV and have been found to dramatically reduce the "viral load" in HIV-infected patients.

The AIDS outbreak has been and continues to be a tragedy on a massive scale. Until a cure is discovered, or a vaccine developed, sustained efforts at preventing its transmission offer our best weapon against the spread of AIDS.

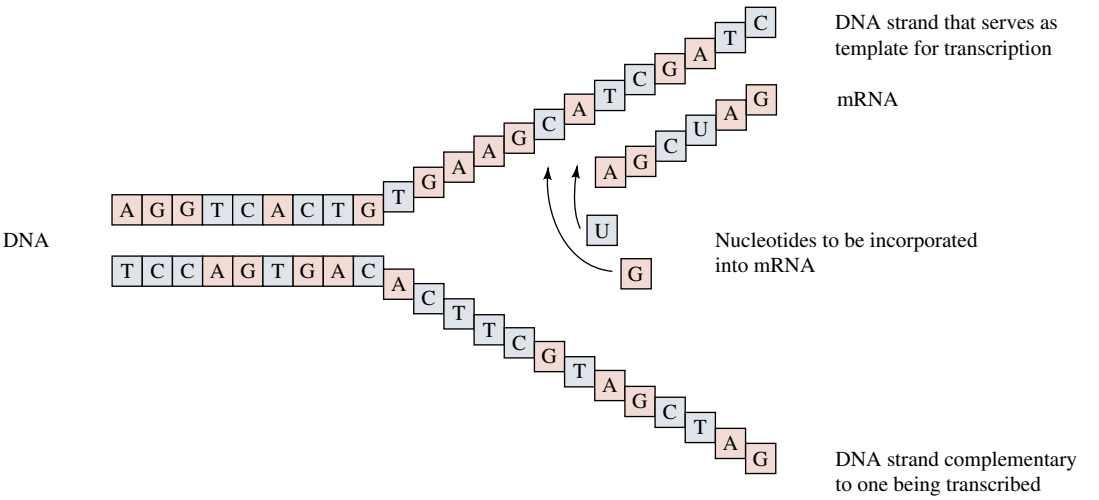


FIGURE 27.27 During transcription a molecule of mRNA is assembled by using DNA as a template.

In the translation stage, the nucleotide sequence of the mRNA is decoded and "read" as an amino acid sequence to be constructed. Since there are only four different bases in mRNA and 20 amino acids to be coded for, codes using either one nucleotide to one amino acid or two nucleotides to one amino acid are inadequate. If nucleotides are read in sets of three, however, the four mRNA bases (A, U, C, G) generate 64 possible "words," more than sufficient to code for 20 amino acids. It has been established that the *genetic code* is indeed made up of triplets of adjacent nucleotides called *codons*. The amino acids corresponding to each of the 64 possible codons of mRNA have been determined (Table 27.4).

TABLE 27.4 The Genetic Code (Messenger RNA Codons)*

Alanine GCU GCA GCC GCG	Arginine CGU CGA AGA CGC CGG AGG	Asparagine AAU AAC	Aspartic acid GAU GAC	Cysteine UGU UGC
Glutamic acid GAA GAG	Glutamine CAA CAG	Glycine GGU GGA GGC GGG	Histidine CAU CAC	Isoleucine AUU AUA AUC
Leucine UUA CUU CUA UUG CUC CUG	Lysine AAA AAG	Methionine AUG	Phenylalanine UUU UUC	Proline CCU CCA CCC CCG
Serine UCU UCA AGU UCC UCG AGC	Threonine ACU ACA ACC ACG	Tryptophan UGG	Tyrosine UAU UAC	Valine GUU GUA GUC GUG

*The first letter of each triplet corresponds to the nucleotide nearer the 5′ terminus, the last letter to the nucleotide nearer the 3′ terminus. UAA, UGA, and UAG are not included in the table; they are chain-terminating codons.

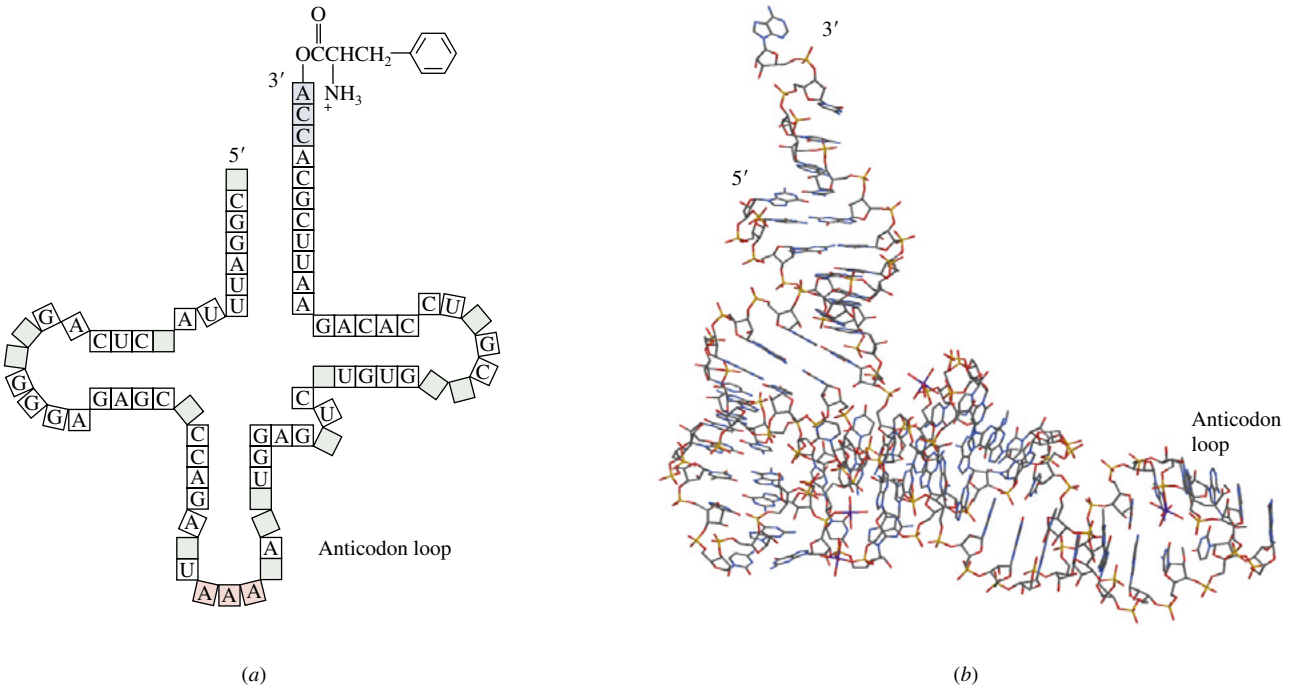


FIGURE 27.28 Phenylalanine tRNA. (*a*) A schematic drawing showing the sequence of bases. RNAs usually contain modified bases (green boxes), slightly different from those in other RNAs. The anticodon for phenylalanine is shown in red, and the CCA triplet which bears the phenylalanine is in blue. (*b*) The experimentally determined structure for yeast phenylalanine tRNA. Complementary base-pairing is present in some regions, but not in others.

PROBLEM 27.23 It was pointed out in Section 27.22 that sickle cell hemoglobin has valine in place of glutamic acid at one point in its protein chain. Compare the codons for valine and glutamic acid. How do they differ?

According to Crick, the so-called central dogma of molecular biology is "DNA makes RNA makes protein."

The mechanism of translation makes use of the same complementary base pairing principle used in replication and transcription. Each amino acid is associated with a particular tRNA. Transfer RNA is much smaller than DNA and mRNA. It is single-stranded and contains 70–90 ribonucleotides arranged in a "cloverleaf" pattern (Figure 27.28). Its characteristic shape results from the presence of paired bases in some regions and their absence in others. All tRNAs have a CCA triplet at their 3′ terminus, to which is attached, by an ester linkage, an amino acid unique to that particular tRNA. At one of the loops of the tRNA there is a nucleotide triplet called the *anticodon,* which is complementary to a codon of mRNA. The codons of mRNA are read by the anticodons of tRNA, and the proper amino acids are transferred in sequence to the growing protein.

27.29 DNA SEQUENCING

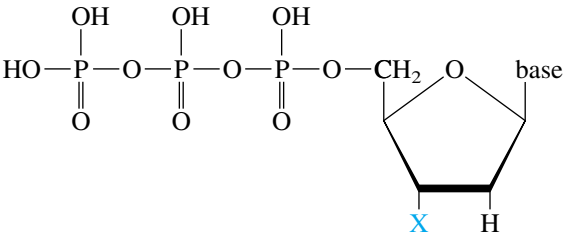
In 1988, the United States Congress authorized the first allocation of funds in what may be a $3 billion project dedicated to determining the sequence of bases that make up the human **genome.** (The genome is the aggregate of all the genes that determine what an organism becomes.) Given that the human genome contains approximately 3×10^9 base pairs, this expenditure amounts to $1 per base pair—a strikingly small cost when one considers both the complexity of the project and the increased understanding of human

biology that is sure to result. DNA sequencing, which lies at the heart of the human genome project, is a relatively new technique but one that has seen dramatic advances in efficiency in a very short time.

To explain how DNA sequencing works, we must first mention **restriction enzymes.** Like all organisms, bacteria are subject to infection by external invaders (e.g., viruses and other bacteria) and possess defenses in the form of restriction enzymes that destroy the intruder by cleaving its DNA. About 200 different restriction enzymes are known. They differ in respect to the nucleotide sequence they recognize, and each restriction enzyme cleaves DNA at a specific nucleotide site. Thus, one can take a large piece of DNA and, with the aid of restriction enzymes, cleave it into units small enough to be sequenced conveniently. These smaller DNA fragments are separated and purified by gel electrophoresis. At a pH of 7.4, each phosphate link between adjacent nucleotides is ionized, giving the DNA fragments a negative charge and causing them to migrate to the positively charged electrode. Separation is size-dependent. Larger polynucleotides move more slowly through the polyacrylamide gel than smaller ones. The technique is so sensitive that two polynucleotides differing in length by only a single nucleotide can be separated from each other on polyacrylamide gels.

Gel electrophoresis of proteins was described in the boxed essay accompanying Section 27.3.

Once the DNA is separated into smaller fragments, each fragment is sequenced independently. Again, gel electrophoresis is used, this time as an analytical tool. In the technique devised by Frederick Sanger, the two strands of a sample of a small fragment of DNA, 100–200 base pairs in length, are separated and one strand is used as a template to create complements of itself. The single-stranded sample is divided among four test tubes, each of which contains the materials necessary for DNA synthesis. These materials include the four nucleosides present in DNA, 2′-deoxyadenosine (dA), 2′-deoxythymidine (dT), 2′-deoxyguanosine (dG), and 2′-deoxycytidine (dC) as their triphosphates dATP, dTTP, dGTP, and dCTP.



X = OH	X = H
dATP	ddATP
dTTP	ddTTP
dGTP	ddGTP
dCTP	ddCTP

Also present in the first test tube is a synthetic analog of adenosine triphosphate in which both the 2′ and 3′ hydroxyl groups have been replaced by hydrogens. This compound is called 2′,3′-dideoxyadenosine triphosphate (ddATP). Similarly, ddTTP is added to the second tube, ddGTP to the third, and ddCTP to the fourth. Each tube also contains a "primer." The primer is a short section of the complementary DNA strand, which has been labeled with a radioactive isotope of phosphorus (^{32}P) that emits α particles. When the electrophoresis gel is examined at the end of the experiment, the positions of the DNAs formed by chain extension of the primer are located by detecting their α emission by a technique called *autoradiography.*

As DNA synthesis proceeds, nucleotides from the solution are added to the growing polynucleotide chain. Chain extension takes place without complication as long as the incorporated nucleotides are derived from dATP, dTTP, dGTP, and dCTP. If, however, the incorporated species is derived from a dideoxy analog, chain extension stops. Because the dideoxy species ddA, ddT, ddG, and ddC lack hydroxyl groups at 3′, they cannot engage in the 3′ → 5′ phosphodiester linkage necessary for chain extension. Thus,

the first tube—the one containing ddATP—contains a mixture of DNA fragments of different length, *all of which terminate in ddA*. Similarly, all the polynucleotides in the second tube terminate in ddT, those in the third tube terminate in ddG, and those in the fourth terminate in ddC.

The contents of each tube are then subjected to electrophoresis in separate lanes on the same sheet of polyacrylamide gel and the DNAs located by autoradiography. A typical electrophoresis gel of a DNA fragment containing 50 nucleotides will exhibit a pattern of 50 bands distributed among the four lanes with no overlaps. Each band corresponds to a polynucleotide that is one nucleotide longer than the one that precedes it (which may be in a different lane). One then simply "reads" the nucleotide sequence according to the lane in which each succeeding band appears.

The Sanger method for DNA sequencing is summarized in Figure 27.29.

This work produced a second Nobel Prize for Sanger. (His first was for protein sequencing in 1958.) Sanger shared the 1980 chemistry prize with Walter Gilbert of Harvard University, who developed a chemical method for DNA sequencing (the Maxam–Gilbert method), and with Paul Berg of Stanford University, who was responsible for many of the most important techniques in nucleic acid chemistry and biology.

In 1995, a team of U.S. scientists announced the complete sequencing of the 1.8 million base genome of a species of influenza bacterium.

A recent modification of Sanger's method has resulted in the commercial availability of automated **DNA sequenators** based on Sanger's use of dideoxy analogs of nucleotides. Instead, however, of tagging a primer with ^{32}P, the purine and pyrimidine base portions of the dideoxynucleotides are each modified to contain a side chain that bears a different fluorescent dye, and all the dideoxy analogs are present in the same reaction. After electrophoretic separation of the products in a single lane, the gel is read by argon–laser irradiation at four different wavelengths. One wavelength causes the modified ddA-containing polynucleotides to fluoresce, another causes modified-ddT fluores-

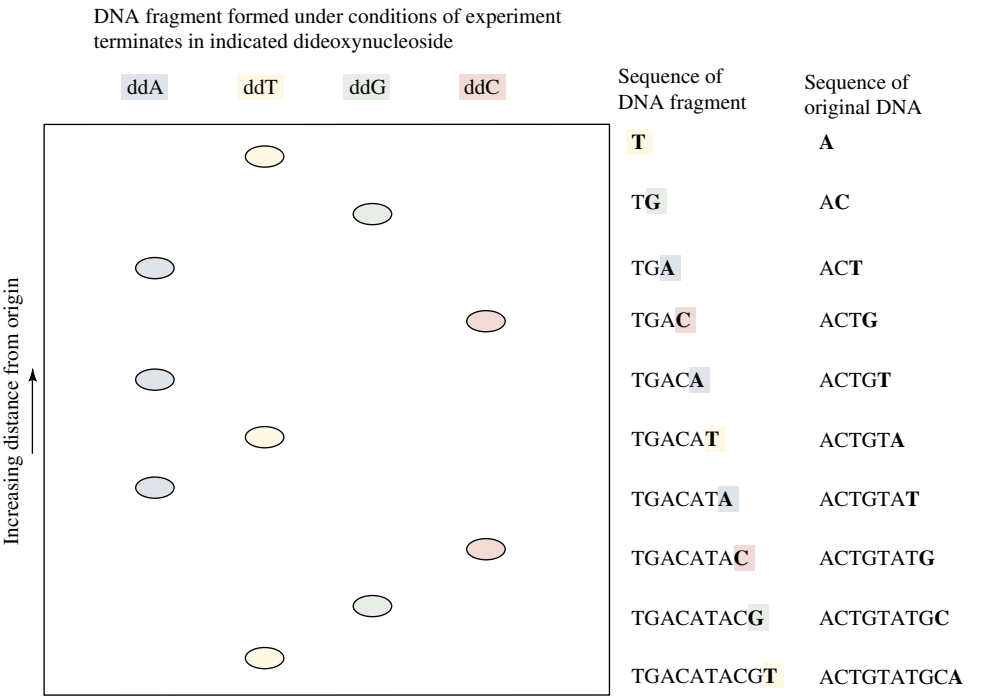


FIGURE 27.29 Sequencing of a short strand of DNA (10 bases) by Sanger's method using dideoxynucleotides to halt polynucleotide chain extension. Double-stranded DNA is separated, and one of the strands is used to produce complements of itself in four different tubes. All of the tubes contain a primer tagged with ^{32}P, dATP, dTTP, dGTP, and dCTP (see text for abbreviations). The first tube also contains ddATP; the second, ddTTP; the third, ddGTP; and the fourth, ddCTP. All of the DNA fragments in the first tube terminate in A, those in the second terminate in T, those in the third terminate in G, and those in the fourth terminate in C. Location of the zones by autoradiographic detection of ^{32}P identifies the terminal nucleoside. The original DNA strand is its complement.

cence, and so on. The data are stored and analyzed in a computer and printed out as the DNA sequence. It is claimed that a single instrument can sequence 10,000 nucleotides per day, making the hope of sequencing the 3 billion base pairs in the human genome a not-impossible goal. The present plan is to complete a draft of the DNA sequence of the human genome by 2001 and a refined version by 2003.

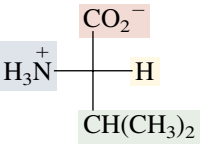
27.30 SUMMARY

This chapter revolves around **proteins.** The first third describes the building blocks of proteins, progressing through **amino acids** and **peptides.** The middle third deals with proteins themselves. The last third discusses **nucleic acids** and their role in the biosynthesis of proteins.

Section 27.1 A group of 20 amino acids, listed in Table 27.1, regularly appears as the hydrolysis products of proteins. All are α-amino acids.

Section 27.2 Except for glycine, which is achiral, all of the α-amino acids present in proteins are chiral and have the L configuration at the α carbon.

Section 27.3 The most stable structure of a neutral amino acid is a **zwitterion.** The pH of an aqueous solution at which the concentration of the zwitterion is a maximum is called the isoelectric point (pI).



Fischer projection of
L-valine in its zwitterionic form

Section 27.4 Amino acids are synthesized in the laboratory from

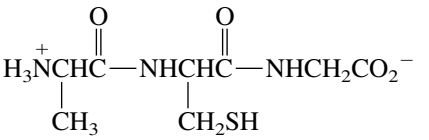
1. α-Halo acids by reaction with ammonia
2. Aldehydes by reaction with ammonia and cyanide ion (the Strecker synthesis)
3. Alkyl halides by reaction with the enolate anion derived from diethyl acetamidomalonate

The amino acids prepared by these methods are formed as racemic mixtures and are optically inactive.

Section 27.5 Amino acids undergo reactions characteristic of the amino group (e.g., amide formation) and the carboxyl group (e.g., esterification). Amino acid side chains undergo reactions characteristic of the functional groups they contain.

Section 27.6 The reactions that amino acids undergo in living systems include **transamination** and **decarboxylation.**

Section 27.7 An amide linkage between two α-amino acids is called a **peptide bond.** The **primary structure** of a peptide is given by its amino acid sequence plus any disulfide bonds between two cysteine residues. By convention, peptides are named and written beginning at the N terminus.



Alanylcysteinylglycine

Section 27.8 The primary structure of a peptide is determined by a systematic approach in which the protein is cleaved to smaller fragments, even individual amino acids. The smaller fragments are sequenced and the main sequence deduced by finding regions of overlap among the smaller peptides.

Section 27.9 Complete hydrolysis of a peptide gives a mixture of amino acids. An **amino acid analyzer** identifies the individual amino acids and determines their molar ratios.

Section 27.10 Incomplete hydrolysis can be accomplished by using enzymes to catalyze cleavage at specific peptide bonds.

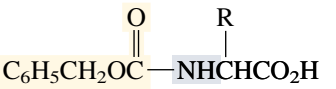
Section 27.11 Carboxypeptidase-catalyzed hydrolysis can be used to identify the C-terminal amino acid. The N terminus is determined by chemical means. One reagent used for this purpose is 1-fluoro-2,4-dinitrobenzene (see Figure 27.8).

Section 27.12 The procedure described in Sections 27.8–27.11 was used to determine the amino acid sequence of insulin.

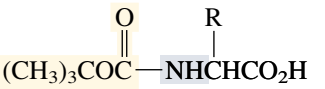
Section 27.13 Modern methods of peptide sequencing follow a strategy similar to that used to sequence insulin, but are automated and can be carried out on a small scale. A key feature is repetitive N-terminal identification using the **Edman degradation.**

Section 27.14 Synthesis of a peptide of prescribed sequence requires the use of protecting groups to minimize the number of possible reactions.

Section 27.15 Amino-protecting groups include *benzyloxycarbonyl* (Z) and *tert-butoxycarbonyl* (Boc).



Benzyloxycarbonyl-protected amino acid

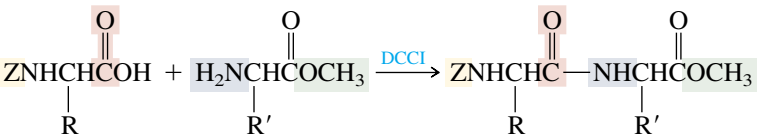


tert-Butoxycarbonyl-protected amino acid

Hydrogen bromide may be used to remove either the benzyloxycarbonyl or *tert*-butoxycarbonyl protecting group. The benzyloxycarbonyl protecting group may also be removed by catalytic hydrogenolysis.

Section 27.16 Carboxyl groups are normally protected as benzyl, methyl, or ethyl esters. Hydrolysis in dilute base is normally used to deprotect methyl and ethyl esters. Benzyl protecting groups are removed by hydrogenolysis.

Section 27.17 Peptide bond formation between a protected amino acid having a free carboxyl group and a protected amino acid having a free amino group can be accomplished with the aid of *N,N′*-dicyclohexylcarbodiimide (DCCI).



Section 27.18 In the Merrifield method the carboxyl group of an amino acid is anchored to a solid support and the chain extended one amino acid at a time. When all the amino acid residues have been added, the polypeptide is removed from the solid support.

Section 27.19 Two **secondary structures** of proteins are particularly prominent. The *pleated β sheet* is stabilized by hydrogen bonds between N—H and C═O groups of adjacent chains. The α *helix* is stabilized by hydrogen bonds within a single polypeptide chain.

Section 27.20 The folding of a peptide chain is its **tertiary structure.** The tertiary structure has a tremendous influence on the properties of the peptide and the biological role it plays. The tertiary structure is normally determined by X-ray crystallography.

Many globular proteins are enzymes. They accelerate the rates of chemical reactions in biological systems, but the kinds of reactions that take place are the fundamental reactions of organic chemistry. One way in which enzymes accelerate these reactions is by bringing reactive functions together in the presence of catalytically active functions of the protein.

Section 27.21 Often the catalytically active functions of an enzyme are nothing more than proton donors and proton acceptors. In many cases a protein acts in cooperation with a **coenzyme,** a small molecule having the proper functionality to carry out a chemical change not otherwise available to the protein itself.

Section 27.22 Many proteins consist of two or more chains, and the way in which the various units are assembled in the native state of the protein is called its **quaternary structure.**

Sections 27-23–27.26 Carbohydrate derivatives of purine and pyrimidine are among the most important compounds of biological chemistry. *N*-Glycosides of D-ribose and 2-deoxy-D-ribose in which the substituent at the anomeric position is a derivative of purine or pyrimidine are called **nucleosides. Nucleotides** are phosphate esters of nucleosides. **Nucleic acids** are polymers of nucleotides.

Section 27.27 Nucleic acids derived from 2-deoxy-D-ribose **(DNA)** are responsible for storing and transmitting genetic information. DNA exists as a double-stranded pair of helices in which hydrogen bonds are responsible for complementary base pairing between adenine (A) and thymine (T), and between guanine (G) and cytosine (C). During cell division the two strands of DNA unwind and are duplicated. Each strand acts as a template on which its complement is constructed.

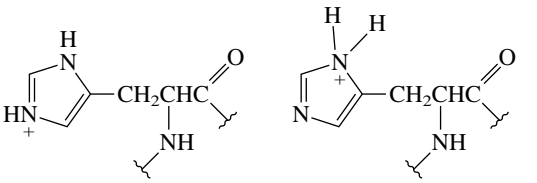
Section 27.28 In the **transcription** stage of protein biosynthesis a molecule of **messenger RNA** (mRNA) having a nucleotide sequence complementary to that of DNA is assembled. Transcription is followed by **translation,** in

which triplets of nucleotides of mRNA called **codons** are recognized by **transfer RNA** (tRNA) for a particular amino acid, and that amino acid is added to the growing peptide chain.

Section 27.29 The nucleotide sequence of DNA can be determined by a technique in which a short section of single-stranded DNA is allowed to produce its complement in the presence of dideoxy analogs of ATP, TTP, GTP, and CTP. DNA formation terminates when a dideoxy analog is incorporated into the growing polynucleotide chain. A mixture of polynucleotides differing from one another by an incremental nucleoside is produced and analyzed by electrophoresis. From the observed sequence of the complementary chain, the sequence of the original DNA is deduced.

PROBLEMS

27.24 The imidazole ring of the histidine side chain acts as a proton acceptor in certain enzyme-catalyzed reactions. Which is the more stable protonated form of the histidine residue, A or B? Why?



A B

27.25 Acrylonitrile ($CH_2{=}CHC{\equiv}N$) readily undergoes conjugate addition when treated with nucleophilic reagents. Describe a synthesis of β-alanine ($H_3\overset{+}{N}CH_2CH_2CO_2^-$) that takes advantage of this fact.

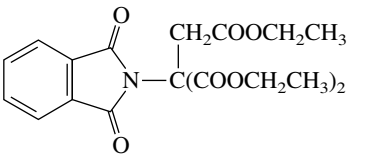
27.26 (a) Isoleucine has been prepared by the following sequence of reactions. Give the structure of compounds A through D isolated as intermediates in this synthesis.

$$CH_3CH_2\underset{\displaystyle Br}{\underset{|}{C}}HCH_3 \xrightarrow[\text{sodium ethoxide}]{\text{diethyl malonate}} A \xrightarrow[2.\ HCl]{1.\ KOH} B\ (C_7H_{12}O_4)$$

$$B \xrightarrow{Br_2} C\ (C_7H_{11}BrO_4) \xrightarrow{\text{heat}} D \xrightarrow[H_2O]{NH_3} \text{isoleucine (racemic)}$$

(b) An analogous procedure has been used to prepare phenylalanine. What alkyl halide would you choose as the starting material for this synthesis?

27.27 Hydrolysis of the following compound in concentrated hydrochloric acid for several hours at 100°C gives one of the amino acids in Table 27.1. Which one? Is it optically active?



27.28 If you synthesized the tripeptide Leu-Phe-Ser from amino acids prepared by the Strecker synthesis, how many stereoisomers would you expect to be formed?

27.29 How many peaks would you expect to see on the strip chart after amino acid analysis of bradykinin?

Arg-Pro-Pro-Gly-Phe-Ser-Pro-Phe-Arg

Bradykinin

27.30 Automated amino acid analysis of peptides containing asparagine (Asn) and glutamine (Gln) residues gives a peak corresponding to ammonia. Why?

27.31 What are the products of each of the following reactions? Your answer should account for all the amino acid residues in the starting peptides.

(a) Reaction of Leu-Gly-Ser with 1-fluoro-2,4-dinitrobenzene

(b) Hydrolysis of the compound in part (a) in concentrated hydrochloric acid (100°C)

(c) Treatment of Ile-Glu-Phe with $C_6H_5N{=}C{=}S$, followed by hydrogen bromide in nitromethane

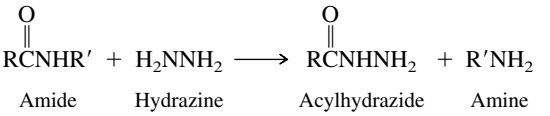
(d) Reaction of Asn-Ser-Ala with benzyloxycarbonyl chloride

(e) Reaction of the product of part (d) with *p*-nitrophenol and *N,N′*-dicyclohexylcarbodiimide

(f) Reaction of the product of part (e) with the ethyl ester of valine

(g) Hydrogenolysis of the product of part (f) over palladium

27.32 Hydrazine cleaves amide bonds to form *acylhydrazides* according to the general mechanism of nucleophilic acyl substitution discussed in Chapter 20:



This reaction forms the basis of one method of terminal residue analysis. A peptide is treated with excess hydrazine in order to cleave all the peptide linkages. One of the terminal amino acids is cleaved as the free amino acid and identified; all the other amino acid residues are converted to acylhydrazides. Which amino acid is identified by *hydrazinolysis,* the N terminus or the C terminus?

27.33 *Somatostatin* is a tetradecapeptide of the hypothalamus that inhibits the release of pituitary growth hormone. Its amino acid sequence has been determined by a combination of Edman degradations and enzymic hydrolysis experiments. On the basis of the following data, deduce the primary structure of somatostatin:

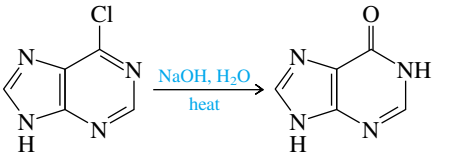
1. Edman degradation gave PTH-Ala.
2. Selective hydrolysis gave peptides having the following indicated sequences:
   Phe-Trp
   Thr-Ser-Cys
   Lys-Thr-Phe
   Thr-Phe-Thr-Ser-Cys
   Asn-Phe-Phe-Trp-Lys
   Ala-Gly-Cys-Lys-Asn-Phe
3. Somatostatin has a disulfide bridge.

27.34 What protected amino acid would you anchor to the solid support in the first step of a synthesis of oxytocin (see Figure 27.8) by the Merrifield method?

27.35 *Nebularine* is a toxic nucleoside isolated from a species of mushroom. Its systematic name is 9-β-D-ribofuranosylpurine. Write a structural formula for nebularine.

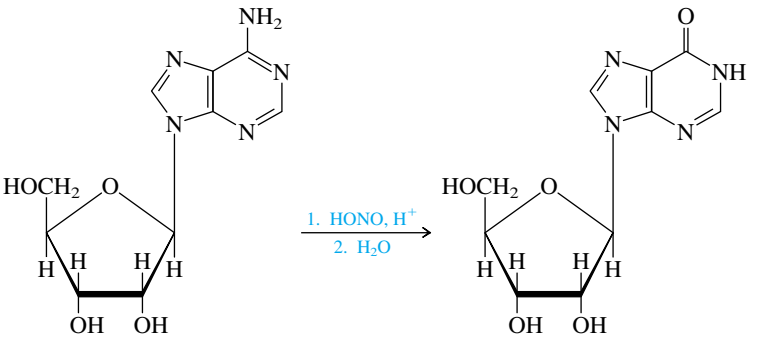
27.36 The nucleoside *vidarabine* (ara-A) shows promise as an antiviral agent. Its structure is identical with that of adenosine (Section 27.24) except the D-arabinose replaces D-ribose as the carbohydrate component. Write a structural formula for this substance.

27.37 When 6-chloropurine is heated with aqueous sodium hydroxide, it is quantitatively converted to *hypoxantine*. Suggest a reasonable mechanism for this reaction.



6-Chloropurine Hypoxanthine

27.38 Treatment of adenosine with nitrous acid gives a nucleoside known as *inosine:*



Adenosine Inosine

Suggest a reasonable mechanism for this reaction.

27.39 (a) The 5′-nucleotide of inosine, *inosinic acid* ($C_{10}H_{13}N_4O_8P$), is added to foods as a flavor enhancer. What is the structure of inosinic acid? (The structure of inosine is given in Problem 27.38.)

(b) The compound *2′,3′-dideoxyinosine (DDI)* holds promise as a drug for the treatment of AIDS. What is the structure of DDI?

27.40 In one of the early experiments designed to elucidate the genetic code, Marshall Nirenberg of the U.S. National Institutes of Health (Nobel Prize in physiology or medicine, 1968) prepared a synthetic mRNA in which all the bases were uracil. He added this poly(U) to a cell-free system containing all the necessary materials for protein biosynthesis. A polymer of a single amino acid was obtained. What amino acid was polymerized?