

Theoretical Organic Chemistry

C. Párkányi
editor

ELSEVIER

Theoretical Organic Chemistry

SERIES EDITORS

Professor P. Politzer
Department of Chemistry
University of New Orleans
New Orleans, LA 70418, U.S.A.

Professor Z.B. Maksić
Ruder Bošković Institute
P.O. Box 1016,
10001 Zagreb, Croatia

VOLUME 1

Quantative Treatments of Solute/Solvent Interactions

P. Politzer and J.S. Murray (Editors)

VOLUME 2

Modern Density Functional Theory: A Tool for Chemistry

J.M. Seminario and P. Politzer (Editors)

VOLUME 3

Molecular Electrostatic Potentials: Concepts and Applications

J.S. Murray and K. Sen (Editors)

VOLUME 4

Recent Developments and Applications of Modern Density Functional Theory

J.M. Seminario (Editor)

VOLUME 5

Theoretical Organic Chemistry

C. Párkányi (Editor)

Theoretical Organic Chemistry

Edited by

Cyril Párkányi

Department of Chemistry and Biochemistry
Florida Atlantic University
Boca Raton, FL 33431-0991, USA



1998

ELSEVIER

Amsterdam – Lausanne – New York – Oxford – Shannon – Singapore – Tokyo

ELSEVIER SCIENCE B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

ISBN: 0 444 82660 2

© 1998 Elsevier Science B.V. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher, Elsevier Science B.V., Copyright & Permissions Department, P.O. Box 521, 1000 AM Amsterdam, The Netherlands.

Special regulations for readers in the U.S.A. – This publication has been registered with the Copyright Clearance Center Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923. Information can be obtained from the CCC about conditions under which photocopies of parts of this publication may be made in the U.S.A. All other copyright questions, including photocopying outside of the U.S.A., should be referred to the copyright owner, Elsevier Science B.V., unless otherwise specified.

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

This book is printed on acid-free paper.

Printed in The Netherlands.

FOREWORD

This volume is devoted to the various aspects of theoretical organic chemistry. In the nineteenth century, organic chemistry was primarily an experimental, empirical science. Throughout the twentieth century, the emphasis has been continually shifting to a more theoretical approach. Today, theoretical organic chemistry is a distinct area of research, with strong links to theoretical physical chemistry, quantum chemistry, computational chemistry, and physical organic chemistry.

Our objective in this volume has been to provide a cross-section of a number of interesting topics in theoretical organic chemistry, starting with a detailed account of the historical development of this discipline and including topics devoted to quantum chemistry, physical properties of organic compounds, their reactivity, their biological activity, and their excited-state properties. In these chapters, a close relationship and overlaps between theoretical organic chemistry and the other areas mentioned above are quite obvious.

Cyril Párkányi
Boca Raton, FL

ACKNOWLEDGMENTS

I greatly appreciate the help, advice, and support provided to me by Anita H. Buckel, Dr. Jane S. Murray, and Dr. Peter Politzer. I am also very grateful to my wife Marie for her endless patience, understanding, and encouragement.

TABLE OF CONTENTS

Chapter 1.	Theoretical Organic Chemistry: Looking Back in Wonder, Jan J.C. Mulder	1
1.	Personal Preface	1
2.	Introduction	3
3.	The First Period (1850-1875)	4
4.	Interlude 1	6
5.	The Second Period (1910-1935)	8
6.	Interlude 2	12
7.	The Third Period	14
8.	Epilogue	20
Chapter 2.	Inter-Relations between VB & MO Theories for Organic π -Networks, Douglas J. Klein	33
1.	Broad Motivation and Aim - Graph Theory	33
2.	VB and MO Models	35
3.	MO-Based Elaborations and Cross-Derivations	38
4.	Hückel Rule	41
5.	Polymers and Excitations	44
6.	Prospects	47
Chapter 3.	The Use of the Electrostatic Potential for Analysis and Prediction of Intermolecular Interactions, Tore Brinck	51
1.	Introduction	51
2.	Methodological Background	51
2.1.	Definition and physical significance	51
2.2.	Spatial minima in the electrostatic potential	52
2.3.	Surface electrostatic potential	55
2.4.	Geometries of weak complexes	58
2.5.	Polarization corrections to the interaction energy	60
2.6.	Charge transfer and the average local ionization energy	61
2.7.	Characters of the different interaction quantities	62
3.	Analysis of Site-Specific Interactions	65
3.1.	Hydrogen bonding	65
3.2.	Frequency shifts	71
3.3.	Protonation	71
4.	Analysis of Substituent Effects on Chemical Reactivity	73
4.1.	Background	73
4.2.	Acidities of aromatic systems	73
4.3.	O-H bond dissociation energies in phenols	77

5. Statistically-Based Interaction Indices	81
5.1. Background	81
5.2. Definitions	82
5.3. Predictions of octanol/water partition coefficients	83
6. Summary	87
Chapter 4. Exploring Reaction Outcomes through the Reactivity-Selectivity Principle Estimated by Density Functional Theory Studies, Branko S. Jursic	95
1. Introduction	95
2. Computational Methodology	96
3. Basics for the Reactivity-Selectivity Approach	96
4. The Diels-Alder Reaction	101
4.1. Diels-Alder reaction of cyclopropene with butadiene	102
4.2. Diels-Alder reaction of cyclopropene with furan	105
5. Ring-Opening Reactions	108
5.1. Cyclobutene ring opening	109
5.2. Influence of substituents upon the reactivity of cyclobutene ring opening	111
6. Radical Reactions	117
6.1. Trichloromethyl radical proton abstraction reaction	117
6.2. Intramolecular radical addition to carbon-carbon double bond	119
7. Reactivity and Stability of Carbocations	123
7.1. Hydride affinity as a measure of carbocation reactivity	123
7.2. Strain energies as a measure of reactivity	126
8. Conclusion	127
Chapter 5. A Hardness and Softness Theory of Bond Energies and Chemical Reactivity, José L. Gázquez	135
1. Introduction	135
2. Reactivity Parameters	136
2.1. The density functional theory framework	136
2.2. Fundamental concepts	137
3. Energy and Hardness Differences	140
3.1. Bond energies	143
3.2. Activation energies	146
4. Catalyzed Reactions and Reactions in Solution	148
5. Concluding Remarks	150
Chapter 6. Molecular Geometry as a Source of Chemical Information for π -Electron Compounds, Tadeusz M. Krygowski and Michal K. Cyrański	153
Abstract	153
Introduction	154
1. Heat of Formation Derived from the Molecular Geometry: The Bond Energy Derived from CC Bond Lengths	155

1.1. Energy content of individual phenyl rings in various topological and chemical embedding	156
1.2. Ring energy content of benzene rings in benzenoid hydrocarbons	157
1.3. Ring energy content in the ring of TCNQ moieties involved in electron-donor-acceptor (EDA) complexes and salts	160
1.4. Ring energy content depending on the intermolecular H-bonding: the case of <i>p</i> -nitrosophenolate anion	161
1.5. Ring energy content as a quantitative measure of fulfilling the Hückel $4n + 2$ rule for derivatives of fulvene and heptafulvene	162
1.6. Estimation of H...O and H...N energy of interactions in H-bonds	163
2. Canonical Structure Weights Derived from the Molecular Geometry	165
2.1. Principles of the HOSE model	166
2.2. Substituent effect illustrated by use of the HOSE model	168
2.3. Structural evidence against the classical through resonance concept in <i>p</i> -nitroaniline and its derivatives	170
2.4. Does the nitro group interact mesomerically with the ring in nitrobenzene?	172
2.5. Angular group induced bond alternation - a new substituent effect detected by molecular geometry	174
3. Substituent Effect on the Molecular Geometry	177
4. Aromatic Character Derived from Molecular Geometry	180
5. Conclusions	183
Chapter 7. Average Local Ionization Energies: Significance and Applications, Jane S. Murray and Peter Politzer	189
1. Introduction	189
2. Average Local Ionization Energies of Atoms	190
3. Average Local Ionization Energies of Molecules	191
3.1. Applications to reactivity	191
3.2. Characterization of bonds	198
4. Summary	199
Chapter 8. Intrinsic Proton Affinity of Substituted Aromatics, Zvonimir B. Maksić and Mirjana Eckert-Maksić	203
1. Introduction	203
2. Absolute Proton Affinities	203
2.1. Experimental basicity scales	203
2.2. Theoretical models for calculating absolute <i>PAs</i>	204
2.3. Proton affinities in monosubstituted benzenes	206
2.4. Proton affinities in polysubstituted benzenes - the additivity rule	211
2.4.1. Increments	211
2.4.2. Disubstituted benzenes - the independent substituent approximation	214
2.4.3. Polysubstituted benzenes	215
2.4.4. The ipso protonation	217

2.4.5. Limitations of the MP2(I) model - the aniline story	222
2.4.6. Proton affinities of larger aromatics - naphthalenes	223
3. Miscellaneous Applications of the Additivity Rule	225
4. Conclusion	228
Chapter 9. Dipole Moments of Aromatic Heterocycles, Cyril Párkányi and Jean-Jacques Aaron	233
1. Introduction	233
2. Experimental Ground-State Dipole Moments	235
2.1. Dielectric constant methods	235
2.2. Microwave methods	238
2.3. The Stark effect method	239
2.4. Molecular beam method	239
2.5. Electric resonance method	239
2.6. Raman spectroscopy	239
2.7. Sign and direction of the dipole moment	239
3. Calculated Ground-State Dipole Moments	241
3.1. Empirical methods	241
3.2. Semiempirical methods	244
3.3. <i>Ab initio</i> methods	245
3.4. Semiempirical and <i>ab initio</i> methods - a comparison	245
4. Experimental Excited-State Dipole Moments	245
5. Calculated Excited-State Dipole Moments	249
6. Conclusion	251
Chapter 10. New Developments in the Analysis of Vibrational Spectra. On the Use of Adiabatic Internal Vibrational Modes, Dieter Cremer, J. Andreas Larsson, and Elfi Kraka	259
1. Introduction	259
2. The Concept of Localized Internal Vibrational Modes	260
3. The Basic Equations of Vibrational Spectroscopy	263
4. Previous Attempts of Defining Internal Vibrational Modes	266
5. Definition of Adiabatic Internal Modes	267
6. Definition of Adiabatic Internal Force Constant, Mass, and Frequency	271
7. Characterization of Normal Modes in Terms of Internal Vibrational Modes ...	273
8. Definition of Internal Mode Amplitudes Δ	277
9. Analysis of Vibrational Spectra in Terms of Adiabatic Internal Modes	281
10. Correlation of Vibrational Spectra of Different Molecules	288
11. Derivation of Bond Information from Vibrational Spectra	297
12. Adiabatic Internal Modes from Experimental Frequencies	302
13. A Generalization of Badger's Rule	308
14. Intensities of Adiabatic Internal Modes	312
15. Investigation of Reaction Mechanism with the Help of the CNM Analysis ...	316
16. Conclusions	324

Chapter 11. Atomistic Modeling of Enantioselection: Applications in Chiral Chromatography, Kenny B. Lipkowitz	329
Introduction	329
1. Stereochemistry	330
2. Chromatography	332
3. Molecular Modeling	335
4. Chiral Stationary Phase Systems	335
5. Modeling Enantioselective Binding	336
6. Type I CSPS	336
6.1. Motif based searches	337
6.2. Automated search strategies	341
7. Type II CSPS	354
8. Type III CSPS	363
9. Type IV CSPS	370
10. Type V CSPS	371
Summary	375
 Chapter 12. Theoretical Investigation of Carbon Nets and Molecules, Alexandru T. Balaban	 381
1. Introduction	381
2. Infinite Planar Nets of sp^2 -Hybridized Carbon Atoms	381
2.1. Graphite: two-dimensional infinite sheets	381
2.2. Other planar lattices with sp^2 -hybridized carbon	382
2.3. Tridimensional infinite lattices with sp^2 -hybridized carbon atoms	384
2.4. Graphitic cones with sp^2 -hybridized carbon atoms	384
3. Infinite Nets of sp^3 -Hybridized Carbon Atoms	385
3.1. Diamond: three-dimensional infinite network	385
3.2. Other systems with sp^3 -hybridized carbon atoms	386
3.3. Holes bordered by heteroatoms within the diamond lattice	386
4. Infinite Nets with Both sp^2 - and sp^3 -Hybridized Carbon Atoms	387
4.1. Local defects in the graphite lattice	387
4.2. Local defects in the diamond lattice	389
4.3. Block-copolymers of graphite and diamond (diamond-graphite hybrids)	390
4.4. Systems with regularly alternating sp^2/sp^3 -hybridized carbon atoms	390
5. Infinite Chains of sp -Hybridized Carbon Atoms	391
5.1. Chains of sp -hybridized carbon atoms: one-dimensional system	391
5.2. Heteroatom substitution inside polyacetylenic chains	391
6. Molecules with sp^2 -Hybridized Carbon Atoms	391
6.1. Fullerenes	391
6.2. Nanotubes and capsules	393
6.3. Carbon cages and nanotubes including oxygen, nitrogen or boron heteroatoms	395
7. Molecules with sp - and sp^2 -Hybridized Carbon Atoms	398
7.1. Cages with sp - and sp^2 -hybridized carbon atoms	398

7.2. Molecules with <i>sp</i> -hybridized carbon atoms	398
7.3. Covalently-bonded nested cages with <i>sp</i> - and/or <i>sp</i> ³ -hybridized carbon, or carbon and silicon atoms	399
8. Conclusions: from Radioastronomy to Remedying Dangling Bonds Carbon Nets	400
Chapter 13. Protein Transmembrane Structure: Recognition and Prediction by Using Hydrophobicity Scales through Preference Functions, Davor Juretić, Bono Lučić, Damir Zucić, and Nenad Trinajstić	405
1. Introduction	405
2. Methods	407
2.1. Selecting protein data bases for training and for testing	407
2.2. Main performance parameters used to judge the prediction quality ..	409
2.3. Hydrophobic moment profile	410
2.3.1. The training procedure for the preference functions method	411
2.3.2. The testing procedure	411
2.3.3. Decision constants choice	411
2.3.4. Collection of environments and smoothing procedure	412
2.3.5. Filtering procedure	412
2.3.6. Predicting transmembrane β -strands (TMBS)	413
2.3.7. Adopted cross-validation technique	414
3. Results	414
3.1. Conformational preference for transmembrane α -helix is strongly dependent on sequence hydrophobic environment for most amino acid types	414
3.2. Expected and predicted length distribution for transmembrane helical segments	416
3.3. What is the optimal choice of the sliding window size?	418
3.4. How do the results depend on different devices used in the SPLIT algorithm?	418
3.5. What are the best scales of amino acid attributes?	420
3.6. The prediction results with Kyte-Doolittle preference functions	422
3.7. Testing for false positive predictions in membrane and soluble proteins of crystallographically known structure	424
3.8. Cross-validation, overtraining and sensitivity to the choice of protein data base	427
3.9. Comparisons with other methods	429
3.10. Using prediction profiles with both α and β motifs	432
4. Discussion	434
Chapter 14. Polycyclic Aromatic Hydrocarbon Carcinogenicity: Theoretical Modelling and Experimental Facts, László von Szentpály and Ratna Ghosh	447
1. Introduction to Chemical Carcinogenesis	447
2. PAH Carcinogenicity and Theoretical Models	450

2.1. The bay-region theory	453
2.2. The MCS model	454
2.2.1. Metabolic factor	455
2.2.2. Carbocation formation	456
2.2.3. Size factor	458
2.2.4. Performance and limitations	458
3. DNA Binding of Carcinogenic Hydrocarbon Metabolites	461
4. Hydrolysis and PAH Carcinogenicity	472
5. Molecular Modelling of Intercalated PAH Triol Carbocations	477
5.1. <i>Ab initio</i> calculations on PAHTC conformations	478
5.2. AMBER modelling of intercalated PAHTC-DNA complexes	481
6. Conclusion	487
Chapter 15. Cycloaddition Reactions Involving Heterocyclic Compounds as Synthons in the Preparation of Valuable Organic Compounds. An Effective Com- bination of a Computational Study and Synthetic Applications of Hetero- cycle Transformations, Branko S. Jursic	501
1. Introduction	501
2. Computational Methodology	502
3. Diels-Alder Reactions with Five-Membered Heterocycles with One Heteroatom	502
3.1. Furan, pyrrole, and thiophene as dienophiles in reaction with acetylene, ethylene, and cyclopentadiene	502
3.2. Addition of benzyne to furan, pyrrole, and thiophene	513
3.3. Cycloaddition reactions with pyrrole as diene for Diels-Alder reaction	518
3.4. Diels-Alder reactions with benzo[<i>b</i>]- and benzo[<i>c</i>]-fused hetero- cycles	529
4. Diels-Alder Reactions with Five-Membered Heterocycles with Two Hetero- atoms	539
4.1. Addition of acetylene, ethylene, and cyclopropene to heterocycles with heteroatoms in the 1 and 2 positions	542
4.2. Addition of acetylene, ethylene, and cyclopropene to heterocycles with heteroatoms in the 1 and 3 positions	546
5. Diels-Alder Reactions with Five-Membered Heterocycles with Three Hetero- atoms	549
5.1. Addition of acetylene, ethylene, and cyclopropene to heterocycles with heteroatoms in 1, 2, and 3 positions	552
5.2. Addition of cyclopropene to heterocycles with heteroatoms in the 1, 2, and 5 positions	554
5.3. Addition of acetylene, ethylene, and cyclopropene to heterocycles with heteroatoms in the 1, 2, and 4 positions	555
5.4. Further investigation of the role of 1,3,4-oxadiazole as a diene in Diels-Alder reactions	558

6. Cycloaddition Reactions with Activated Heterocycles That Have Two or Three Heteroatoms	563
6.1. Activation of 1,2-diazole as a diene for Diels-Alder reaction	563
6.2. Transformation of cyclic malonohydrazides into the Diels-Alder reactive 1,3-diazole	567
6.3. Quaternization of nitrogen atom as a way to activate 1,3-diazole, and 1,3,4-triazole as a diene for the Diels-Alder reaction	569
6.4. Oxidation of a sulfur atom: a way to activate 1,3-thiazole and 1,3,4-thiadiazole as dienes for the Diels-Alder reaction	571
7. Conclusion	574
Chapter 16. Triplet Photoreactions; Structural Dependence of Spin-Orbit Coupling and Intersystem Crossing in Organic Biradicals, Martin Klessinger	581
1. Introduction	581
2. Basic Theory	582
2.1. Wave functions and operators	582
2.2. Matrix elements between bonded functions	584
2.3. Evaluation of spin-orbit integrals	586
3. Spin-Orbit Coupling and Intersystem Crossing in Biradicals	587
3.1. Carbene	588
3.2. Ethylene	590
3.3. Trimethylene	592
3.4. 1,2-Dimethyltrimethylene	595
3.5. Tetramethylene	596
3.6. Oxatetramethylene	599
4. Models for Spin-Orbit Coupling	600
4.1. The 2-in-2 model	600
4.2. Symmetry considerations	603
4.3. The "through-space" vector model	603
5. Conclusions	606
Index	611

Theoretical Organic Chemistry: Looking Back in Wonder

Jan J.C.Mulder, Gorlaeus Laboratories, P.O.Box 9502,
Leiden University, 2300 RA, Leiden, The Netherlands

1. Personal preface

In 1958 the Chemical Society organized the "Kekulé Symposium" in London. The papers presented at the meeting were published under the auspices of the International Union of Pure and Applied Chemistry, Section of Organic Chemistry, under the title "Theoretical Organic Chemistry" [1]. Indeed Kekulé regarded his contribution [2] as theoretical, and as it was concerned with "the chemical nature of carbon" it was certainly organic.

In 1958 I started studying with L.J.Oosterhoff¹ who had been professor of theoretical organic chemistry in Leiden since 1950. It was an auspicious moment to enter the field. Computers started to make an impact in the beginning of a period measured between 1955 and 1980 that marked the heyday of theoretical and physical organic chemistry.

In the course of this introductory chapter on the history of theoretical organic chemistry I will have occasion to comment on demarcation lines that separate these disciplines, and on the relation between chemistry and physics. These questions have been discussed by Walker [3] and Theobald [4] and recently at length by Nye [5] and van der Vet [6]. Nye's book especially has been a valuable source of information.

¹ This manuscript is dedicated to the memory of an unforgettable teacher.

On a lighter tone, I cannot resist the temptation to mention two characterizations that concern the difference between physics and chemistry. The first one, due to Oosterhoff [7], states that "the difference between chemistry and physics is in essence the difference between chemists and physicists". The second one [8] is expressed as follows: "The theoretical physicist moves like a swallow with elegant swerves through the thin air of abstract thought, whilst the theoretical chemist on most occasions rummages in the earth like a dung-beetle, that only exceptionally is able to raise itself above the ground, albeit with a loud whirr".

The idea of revolutionary progress in certain periods as developed by Kuhn [9] and very recently by McAllister [10], has some bearing on what I will discuss. It will be argued that theoretical organic chemistry has known three periods of dramatic change. The first of these periods (1850-1875) witnessed the birth of the structural formula and its development from formal representation to a reflection of physical reality. The second (1910-1935) saw the advent of quantum mechanics and the concepts of the electron pair, resonance and mesomerism, and hybridisation. In the third one (1955-1980), already mentioned, it is perhaps the successful application of molecular orbital theory to chemical reactions, made possible by a very fruitful interplay of calculations and concepts, which is most significant.

A word of warning before starting my exposition is in order. Although this is a chapter on the history of theoretical organic chemistry, and as such has a beginning and an end, this certainly is not the case in the literature. Having been in and with this subject for almost 40 years, it is unavoidable that my appreciation for the contributions of many of my colleagues has become idiosyncratic. Only the future can decide whether my nostalgia coincides in a more than trivial way with truly historic developments.

2. Introduction

The objective of theoretical organic chemistry has always been to correlate systematic variation in physical, chemical and (eventually) biological properties of organic molecules, with systematic variation in their molecular structure. This, of course, is only a relative correlation that was already possible long before the advent of quantum mechanics. The formulation of structure-colour correlation rules by Witt [11], Dilthey [12] and Wizinger [13], forms an impressive example. In inorganic chemistry the periodic system provided correlation via "isocolumn" substitution. An absolute correlation of properties with structure becomes possible, at least in principle, within the application of quantum mechanics to chemical problems. Probably this has been called quantum chemistry for the very first time in a curious application of the general theory of relativity to molecular systems by de Donder [14]. In 1929 a little book by Haas appeared [15], which may have been the first with the title Quantum Chemistry.

Theoretical organic chemistry is principally concerned with the structure of molecules and - in reactions - of transition states. In contrast physical chemistry and theoretical chemistry are also tackling the bulk properties and, more importantly, are bridging the gap that separates them from the molecular properties. Physical organic chemistry, invented by Hammett [16], occupies an intermediate position and has in time become the experimental partner of theoretical organic chemistry. As far as the covalent bond, ubiquitous in organic molecules, can only be understood using quantum mechanics, it follows that for instance in the textbooks by Streitwieser [17] and Dewar [18], theoretical organic chemistry becomes almost synonymous with quantum chemistry,

exemplified in the application of molecular orbital theory to organic molecules. The early books by Henrich [19] and Branch and Calvin [20] had little or no quantum mechanics whatsoever. Wheland [21] and Walter Hückel [22] wrote influential texts, that contained almost no molecular orbital theory. Pullman and Pullman's book [23] has a balanced treatment of valence bond and molecular orbital methods. Hermans [24], Staab [25], Liberles [26] and Lowry and Richardson [27] are really all physical organic chemistry texts. The book by Sandorfy [28] quite rightly had a huge success and was translated into German and English. Very recently a hybrid of physical and theoretical organic chemistry appeared [29], written by Shaik, Schlegel and Wolfe, and especially concerned with the valence bond configuration mixing model for SN_2 -reactions, which in itself had a forerunner in Salem's [30] beautiful little book. These last two references, together with new trends in organic photochemistry that will be discussed, constitute an important core area of theoretical organic chemistry at the present time. Naturally the qualitative explanations will always lean heavily on the quantitative calculations.

With the foregoing, theoretical organic chemistry has been positioned in its scientific environment and the analysis of its development can now be undertaken.

3. The First Period (1850-1875)

The understanding of the behaviour of organic molecules which follows upon Couper's introduction of the structural formula [31] can hardly be overrated. At first - as emphasized by Frankland - the line between two atoms only meant the mutual saturation

of valencies [32], but soon, due to Crum Brown [33], the graphical display of the physical positions of the atoms with respect to one another crept into play. This sequence of discoveries culminated with the realisation by van't Hoff [34] and Le Bel [35] that molecules exist in a three-dimensional space. Of course the history of these events has been described many times. Mackle [36] and Rouvray [37] have given brief reviews with emphasis on the concepts of valence and bond symbolism, that are enlightening. In a very short time the transition was made from the constitutional formula to the structural formula, elucidating the constitution *via* the mutual saturation of valencies. From there the step towards the true meaning of the structure, *i.e.* the demonstration of physical connectivity, was made, and finally the so constructed network became an edifice in three dimensions. The debate amongst the leading organic chemists of the day, exemplified by the extremely critical Kolbe [38] forms ample testimony of the significance of this revolutionary change in concepts. An important aspect of the existing problems was the confusion about atomic weights and the concept of equivalents. It was only through Cannizzaro's [39] introduction of the true atomic (and molecular) weights, due to the earlier work by Avogadro, that the situation was clarified.

Then, in the middle of this period, it is discovered - again by Kekulé - that there are cases in which a single structural formula does not account for the chemical properties of the molecule [40]. His explanation has been called the "oscillation hypothesis" and has been quoted by Staab [41] in another 100 years remembrance. Had it been the NH_3 structure going through the D_{3h} planar form, that Kekulé was discussing, his reasoning would have been entirely correct. The real state of affairs will become the central issue in the second period (5.). In fact it was only during the third period (7.), that the structures for C_6H_6 originally proposed by Dewar [42], Hückel [43], and Ladenburg

[44], were shown by van Tamelen [45], Wilzbach [46], and Katz [47], with their colleagues, to be different molecules, perfectly capable of existence. Incidentally, benzvalene was mentioned by Hückel only as a possible non-canonical bond eigenfunction and not as a real structural formula. The theoretical development in organic chemistry might have taken a different and possibly faster route, had all this been known at the time. The elusive Claus' [48] structure of benzene may also be called "octahedral" benzene and would - with one of the diagonal bonds uncoupled - be a candidate for existence in the triplet state if suitable precursors for a photochemical transformation could be found.

4. Interlude 1

The fruits of the eventful 25 years in which the molecule became tangible in organic chemistry had to be digested and explanations repeated. The precise contents of the structural formula, *i.e.* the meaning of the bar representing the bond, prompted the interest of others. As reviewed by Rouvray [49] hundred years later, Cayley [50] was the first to apply graph theory to isomer counting. The mathematicians Clifford [51] (Clifford algebra), Sylvester [52] and Gordan [53] (the Clebsch-Gordan series!) were concerned with invariant theory, and it is interesting that the analogy was discovered this early, because the subsequent development of valence bond theory in the hands of Weyl and Rumer [54] showed the connection to be not only formal in character, but a source for a viable theory of chemical bonding. Many years later, Clifford's contributions were rediscovered and exploited by Paldus and Sarma [55]. They showed the utility of $U(2^n)$ over $U(n)$ and the use of spinor invariants in chemistry. The geometry of molecules has been an essential element of theoretical organic chemistry

from the beginning. An important part has been played by the cycloalkanes. The strain theory developed by Bayer [56] may be viewed as the start of conformational analysis, mainly because Sachse [57] was able to show the flaw in the assumption that these molecules were planar. Later Mohr [58] completed the argument and the "chair" and "boat" forms of cyclohexane were born. The idea of easily interconvertible isomers is already present in substituted ethanes and the calculation of the barrier of rotation in the parent molecule is a foremost problem in quantum chemistry. Finally, in the third period, Kern and coauthors [59] were able to show that the main effect is the exchange repulsion between the C-H bonding pairs.

Reactivity of strongly bonded molecules was considered by Thiele [60], who introduced the concept of residual valence. This is tied in with Bayer's strain theory in the sense that a solution for the same question was sought. Whereas the stereochemistry of the cycloalkanes made Bayer's theory obsolete, the later introduction of resonance more or less confirmed Thiele's intuition.

Notwithstanding new attempts by Bamberger [61] and Armstrong [62] the structure of benzene remained a stumbling block. One of the problems that plagued the theoreticians when analyzing the effect of substituents, was the difference between polarity and polarizability, but Vorländer saw it clearly and early [63]. Here one discerns the seeds of the later inductive and mesomeric effects. Chemistry as a whole was dominated by the gradual filling of the periodic table and the debate on its (ir)regularities, as discussed in detail in van Spronsen [64]. Early ideas by Abegg [65] and Drude [66] called attention to the electronic character of valence using the positions of elements in the system.

5. The second period (1910-1935)

Nobody will argue the importance of the idea of the electron pair bond, introduced by Lewis [67], in chemistry. Together with the Bohr theory of the electronic structure of the atom [68] and its connection with the periodic system [69], one has the ingredients for a true chemical theory. The octet model introduced by Langmuir [70] soon demonstrated its immense explanative power for organic and inorganic structure alike.

The electronic character of the chemical bond opened the door to polarity, and this was exactly the concept needed for the understanding of chemical reactions. The great schools of the study of organic reaction mechanisms took off immediately and their development took place independently of the creation of quantum mechanics. One concept though, became a link between the two, and this of course was resonance. The way that this connection was made is interesting because of the different views of the participants [71]. There can be no question about the fact that organic chemists like Weitz [72] and Arndt [73] did discover the necessity of describing the structure of certain molecules as intermediate between extreme formulae before the resonance concept was introduced in quantum mechanics by Heisenberg [74]. The main difference between the chemical and the quantum-mechanical significance of resonance lies in the reactivity *versus* the stability argument. The general impression though, that it was Ingold [75] who invented mesomerism is wrong, as discovered by Eistert [76]. The difference between mesomerism and tautomerism took some time to be recognized but is in essence connected to the Born-Oppenheimer approximation [77].

In the second period the electronic structure of benzene - but not naphthalene! - was

finally understood due to Robinson [78], but interestingly it was not Pauling but Hückel, who first applied the valence bond method to benzene [79]. On the other hand, Wheland and Pauling [80] were the first to apply the Hückel method systematically. The regularities in the properties of substituted benzenes were known and interpreted for instance by Vorländer [81], but the empirical rules following from this knowledge met with frequent criticism, as exemplified by Lowry [82]. Much later Heilbronner and Grinter [83] succeeded in bringing physical and chemical properties together and explaining them correctly.

The story of resonance, which starts in the middle of this period, is an intriguing one. There are at least two but perhaps three directions to discuss. Pauling pushed the concept mainly as a qualitative method to gain insight into the stability and reactivity of molecules. This is the line followed in "The Nature of the Chemical Bond" [84]. Together with the curved arrow, introduced by Robinson, it became for many years the preferred way of thinking for organic chemists. At the same time Pauling and his collaborators created the qualitative valence bond calculations for π -electron systems [85]. This became the method of choice in the pre-computer era because the number of structures could be controlled, whereas in the Hückel molecular orbital method the number of π -atomic orbital centers automatically fixed the dimension of the secular equation. Both methods led into a dead alley for large systems, but the Hückel method was superior as soon as computers became available. Moreover, as it turned out, the molecular orbital method was easier to generalize into programs and large basis sets presented no special problems.

In between, the relationship of the two main quantum-chemical methods was established in the general sense by Slater [86] and later by Longuet-Higgins [87]. The fact that molecular orbital and valence bond methods must, if used with the same basis set and the

same approximations, but with full configuration interaction or inclusion of all structures, lead to the same results, was of little help if this process was impractical. Thus a number of examples arose in the literature where the methods gave different results. The first of these was the oxygen molecule where the MO method in the hands of Lennard-Jones [88] was superior in predicting the triplet ground state, with respect to the first order VB result as discussed by Wheland [89]. The second example is cyclobutadiene where again the MO method easily predicts the ground-state triplet, but the interaction of the two covalent structures in the VB model gives a singlet state. In this case extensive calculations [90] made even before computer technology was fully developed, indicated that the VB result is probably the right one. In fact, if one realizes that the oxygen molecule is isoelectronic with ethylene and also takes into account that orthogonal ethylene is equivalent to cyclobutadiene because of the isomorphism of the D_{2d} and D_{4h} symmetry groups [91], the two examples become almost identical. There is, however, one important difference between O_2 on the one hand and orthogonal ethylene and cyclobutadiene on the other. The last two can lower their symmetry and so remove the orbital degeneracy which is present, and which favors the triplet configuration. This is the pseudo-Jahn-Teller effect [92], to be distinguished from the Jahn-Teller effect [93], that describes the fate of a state degeneracy. It has become clear later that the Jahn-Teller situation, being a conical intersection, will in fact only be affected in two (or a combination of two) symmetry-lowering coordinates, but will persist in other degrees of freedom. The symmetry of the intersection geometry makes it easy to find but plays no further role. The relationship between symmetry and degeneracy is taught to students by means of the first example where it exhibits itself, the two-dimensional square well. Nevertheless, the same example also demonstrates the simplification that is involved, as

was nicely shown by Shaw [94]. This, until now, only found its way to the textbook by Berry, Rice and Ross [95].

It is also possible to find relations between the MO and VB approaches on an intermediate level, as shown by Heilbronner [96]. His rather extreme view was that resonance theory expressed molecular orbital results in a different language. The "classical" valence bond model would emerge again quite recently in applications by Durand and Malrieu [97] using the Heisenberg Hamiltonian, and Bernardi, Olivucci and Robb [98], modelling photochemical reactions.

Hybridisation was invented simultaneously by Pauling [99] and Slater [100] within the framework of valence bond theory. The concept took hold immediately and obtained a place in all textbooks. Although the idea is superfluous in a molecular orbital context, it has remained a point of departure in the discussion of the geometry of organic molecules. The interesting question whether (i) hybridisation "happens" or is only a model which may or may not be used, and (ii) if hybridisation "explains" the shape of molecules, has been tackled by Cook [101] much later. His answer was yes to the first question and no to the second one. The problem was taken up again by Ogilvie [102], but his analysis only takes care of the semantics. Kutzelnigg [103] established the importance of non-orthogonal hybrids that allow for the description of smaller angles, but he also discussed the relationship between hybridisation and "electron pair repulsion" in a very lucid manner. This also applies to the treatment of the chemical bond given in Kutzelnigg's book [104], where the role of the kinetic energy, that was first emphasized by Hellmann [105], gets its proper place. Ruedenberg [106] has treated the problem convincingly and his conclusions can be summarized as follows. At large distances the relief of kinetic pressure indeed lowers the energy. This is

accompanied by a more diffuse wave function obtained via exponent optimization. At smaller distances the potential energy of attraction takes over, the wave function contracts, and gets an optimal exponent that is now higher than the value for separated atoms.

6. Interlude 2

Without the assistance of computers the application of quantum mechanics in chemistry could only progress slowly and, as foreseen by Dirac [107] in the second - never quoted - part of his famous pronouncement, subject to approximations. Two contributions stand out clearly, one is the famous Goeppert-Mayer/Sklar calculation on benzene [108] and the other the Coulson/Fischer treatment of the hydrogen molecule [109]. Both were ahead of their time as subsequent studies by Parr, Craig and Ross [110] and Garrett [111] showed. The Spin-Coupled Valence Bond method based on simultaneous optimization of the Coulson/Fischer type orbitals and the covalent VB-structure contributions has become a powerful calculational and especially interpretative tool [112].

The question of the benzene structure was taken up again by Lennard-Jones and Turkevich [113]. They showed, using molecular orbital arguments, that the π -system of $C_{2n}H_{2n}$ is unstable with respect to bond localisation. Wheland [114] questioned their result on the basis of an (unpublished) valence bond calculation. It would take almost sixty years before this seeming discrepancy was finally settled [115].

In this interlude the most influential book ever on quantum chemistry [116] appeared. One of the authors, Kimball, became the inventor of d^5 hybridisation [117]. This fact was completely forgotten, perhaps because it was entirely group theoretical, and, the

reduction table for 5-coordination contained a misprint in the d-function row! It turned up again in a series of articles on the five equivalent d-orbitals in the Journal of Chemical Education [118].

After the second world war the renewed interaction between theoreticians in Europe and the United States stimulated by the Shelter Island conference and the one in Paris in 1948 [119] became an important factor in the rapid development of quantum chemistry. Roothaan [120] saw that the numerical self-consistent field calculations on atoms and simple diatomic molecules would never do for polyatomic molecules, and accordingly developed the general procedure with the basis set as starting point. Pople [121] and Pariser/Parr [122] created the approximate method for π -electron systems that allowed for electron repulsion in the Hückel framework. The contribution from Cambridge (U.K.) was notable, with Lennard-Jones' extremely lucid but little known explanation of the importance of the Pauli Principle [123], and the introduction of the Gaussian basis functions and the "poly-deter" general configuration interaction method by Boys [124], as outstanding examples. Both are of fundamental importance, but are seldom cited. Lennard-Jones and his students also wrote an influential series of articles "The molecular orbital theory of chemical valency" that appeared in the Proceedings of the Royal Society [125]. In this series the idea of equivalent orbitals was exploited, which led to localized electron pairs in molecules from a molecular orbital viewpoint. Number VIII of the series by Hall [126], together with a paper by Moffitt [127], describes the SCF method for molecules in essentially the same way as Roothaan, so that both should be credited for the discovery. The important role of the U.K. becomes even more visible taking into account a competing series with Coulson as leading author [128]. Much more important than the papers was Coulson's role as a

teacher [129] and organizer of the Oxford Summer Schools in quantum chemistry. There and in the very different but equally important courses organized by Löwdin, in Uppsala (and Abisko) as well as later at Sanibel Island, Florida, younger people who would become active in the field, learned the trade and got to know each other.

7. The third period (1955-1980)

The field of computation of the energies and properties of organic molecules has gone through a number of stages. The Hückel method was suitable for π -electron systems, and although the self consistent field was introduced to incorporate electron repulsion in all-electron calculations, the approximate version was first applied to unsaturated systems. The Pariser/Parr/Pople (PPP) method gained immense popularity for the study of (hetero)aromatic molecules and derivatives. Because of the inclusion of configuration interaction it became possible to calculate spectroscopic properties, dipole moments and charge distributions in ground and excited states. This meant that reactivity could be investigated, and in many cases Hückel predictions could be tested and refined. In this area the Japanese school around Fukui [130] and others [131] became very active. All this was molecular orbital calculation, but it was shown by McWeeny [132] that the approximations could be incorporated in valence bond theory and that in this way a consistent scheme for calculations could be set up. The use of the so-called Löwdin orbitals [133] on the one hand secures the theoretical foundation of the approximations, but on the other hand necessitates the inclusion of many polar structures. The comparison was made by Campion and Karplus [134]. Somewhat earlier this had

also been seen by Craig [135]. He was the first showing the correct use of symmetry in the VB method. It is interesting to note, and easy to understand, that the necessity of the polar structures following from the orthogonality of the orbitals, is the opposite situation of what happens if one uses Coulson/Fischer orbitals, that are indeed heavily non-orthogonal.

Approximate methods applicable to all molecules were needed as long as *ab initio* calculations were still very expensive, and this meant that the idea of the PPP-method that integrals could be neglected as well as approximated or parametrized, was applied to σ -systems. Naturally, the zero differential overlap approximation that reduced the n^4 -dependence of the electronic repulsion integrals to a n^2 -dependence had to be reanalyzed. Pople again took the lead with the CNDO and INDO methods [136], and Dewar stayed close to organic chemistry with MINDO/1,2,3 [137], MNDO [138], and the later refinements AM1 and PM3. Hoffmann introduced Extended Hückel Theory [139] a little earlier. The contrast between the two ways of getting results for molecules of general geometries is marked. Whereas EHT has no explicit electronic repulsion, is non-SCF and keeps all non-neighbour resonance terms and all overlap integrals, the other methods are in every way antipodes. Later still, the improving performance of computers led Pople and coauthors to the development of the very successful Gaussian [140] series of *ab initio* programs.

Still another method designed for the same purpose, namely equilibrium ground state geometries, must be mentioned. Starting with Westheimer's early calculations [141], the molecular mechanics or force field methods became a very important tool, especially for very large molecular systems, as encountered in biochemical applications. Here the contributions by Lifson/Warshel [142] and Allinger [143] deserve attention.

The idea of the correlation diagram has been a cornerstone in electronic structure calculations from the beginning. Introduced by Mulliken, its importance was emphasized succinctly by van Vleck [144]. In the hands of Walsh [145] a very powerful method was developed to analyze the geometry of ground and excited states of simple molecules, using the angle dependence of molecular orbital energies. Later it was shown that VSEPR theory [146] could do a comparable job for even more complex systems. For orbitals the correlation diagram is a construct, but as soon as electronic configurations become the labels in the diagram and crossing and non-crossing arguments have the secure base discovered by von Neumann and Wigner [147], one is approximating real energy curves. In diatomic molecules this is (almost) the complete picture, but in polyatomic molecules with their many internal degrees of freedom, in general undefined cuts through the potential energy surface are obtained. The situation in triatomic molecules, with emphasis on degeneracies, has been described by Davidson [148].

The study of chemical reactions using correlation diagrams, although created by Longuet-Higgins and Abrahamson [149] in molecular orbital language, has its roots in valence bond theory [150]. In the discussion of "forbidden" and "allowed" reactions, a concept introduced by Woodward and Hoffmann [151] in the middle of this period, the ("avoided") crossing of potential energy curves (surfaces) plays a prominent role. This is due to van der Lugt and Oosterhoff [152] and also to Salem [153]. In this way reactions with opposite stereochemical outcome are differentiated from each other and photochemical reactions are explained. The discussion of photochemical and thermal reactions where diradicals are intermediate has profited immensely from Salem's [154] crystal-clear presentation.

Because of the unclear situation in polyatomic molecules Longuet-Higgins [155]

reexamined surface crossings. He obtained the proof of the existence of a degeneracy by means of the sign change in the wave function while describing a loop through configuration space around it, and he reestablished the fact that it is not so much symmetry which determines what happens, but the number of independent geometrical parameters of the system. This had been found before by Herzberg and Longuet-Higgins [156] and Teller [157]. The real state of affairs has been demonstrated analytically in a three-state model [158], but was described most clearly by Stone [159]. Landau and Lifchitz [160] is one of the few texts with the full story of the conical intersection. As it happens, careful scrutiny of Kauzmann's [161] book would have shown that he had also seen the consequences of Teller's contribution. If one wants to analyze the combined influence of symmetry and the number of independent coordinates, a result by Pople and coauthors [162] is indispensable. The conical intersection between two states of the same symmetry was discovered by London [163], but the general use of this concept in photochemistry is due to Robb, Olivucci and Bernardi [164]. It now seems quite certain that this is the "reaction funnel" as introduced by Michl [165]. This means that the radiationless transition through the avoided crossing area has become secondary. Barriers on the excited potential energy surface that make it more difficult for the system to reach a conical intersection, have taken the place of the less narrowly avoided crossing [166] in earlier explanations.

At this point a small digression is warranted. Both the general significance of the non-crossing rule and the opposite behaviour in photochemical reactions with respect to the corresponding thermal processes have the character of an aesthetic canon. There has been considerable resistance in the literature to the new insight that the true nature of von Neumann and Wigner's proof implies that conical intersections are everywhere, and that

the role of symmetry is considerably less important than previously thought. This means that an example in recent history of chemistry has been found, which fits McAllister's criteria [167] for a small revolution in spectroscopy and photochemistry of polyatomic molecules.

The general attack on the configuration interaction problem got a new impetus through the efforts of Paldus [168], who introduced the unitary group in quantum chemistry. That this group can be more than a formal device was emphasized by Matsen [169]. As in the earlier work by Koutecký [170] and van der Lugt [171] the π -electron system of benzene provided the benchmark for the calculations. Paldus also established the connection between the dimension of the full CI matrix and the dimension of the two-column irreducible representations of the unitary group. Subsequently he obtained the special closed-form general dimension formula, in non-closed form to be found in Weyl [172], that had been derived earlier [173] from combinatorial considerations. Shavitt [174], who had already participated in the early work with Boys [175], again played a major role in showing the great potential of graphical instead of algebraic representations.

During this period the important progress in reaction dynamics has been notable. This is amply demonstrated by the Nobel prizes for the experimentalists Polanyi [176] and Herschbach and Lee [177], but regarding the theoretical aspects Wyatt's [178] contributions deserve recognition. Furthermore Heller [179], developed the wave packet method, to treat reactions that start on the excited state potential energy surface. This whole area of research has profited from the workshops that were held in Orsay (1973, 1975, 1977, 1985) [180], under the auspices of the "Centre Européen de Calcul Atomique et Moléculaire" (CECAM), created by Carl Moser, its first director.

CECAM also provided opportunities for young as well as established scientists in physics and chemistry to interact and perform calculations for longer periods [181].

Naturally, when discussing reaction dynamics, the potential energy surface is taken for granted. Whereas in the early dynamics calculations this surface was no more than a model surface or the result of a semi-empirical calculation, nowadays *ab initio* and fully geometry optimized surfaces are available. The precise significance of the potential energy surface has been analyzed by Sutcliffe [182]. He reiterated the problem with the concept of molecular structure if the Born-Oppenheimer approximation, a prerequisite for the idea of a potential energy surface, is not assumed. This problem had been brought to the attention by Woolley [183]. The situation has been reviewed by Weininger [184].

In the calculation and application of potential energy surfaces the possibility of using analytical gradients and second derivatives has been of paramount importance. From the long list of people that have been involved, it may be sufficient to mention only Pulay [185], who is generally regarded as the pioneer in the field, Komornicki and McIver [186], Handy and Schaefer [187], Schlegel [188], Helgaker and Jørgensen [189] and Gauss and Cremer [190].

Before this development took place the definition of the reaction pathway on the surface, with the concomitant characterization of the transition state had solicited considerable attention. Fukui [191] contributed the idea of the intrinsic reaction coordinate (IRC), but earlier Murrell [192] had discussed the possible symmetries of transition states. This would turn up again in Salem's [193] treatment of the so-called narcissistic reactions. Stanton and McIver [194] and Pechukas [195] formulated general symmetry conditions for the transition vector. It was thought that symmetry could bring

order in the many possible pathways on the complicated potential energy hypersurface. It is conceivable that hardware and software improvements of the last few years carry with them the conviction that "brute force" will be the the only solution in the end, and that energy should be spent as such.

8. Epilogue

The extraordinary rich structure of organic chemistry guarantees that theoretical organic chemists will be occupied for quite some time. Will they be a separate breed in the future? It has been surmised on more than one occasion that there will come a time that the computer with its quantum chemistry software is fully comparable with for instance the NMR spectrometer with its software. Will specialists be needed? Will they still be writing programs? Are they mainly needed for educational purposes? Obviously theoretical chemists penetrate biology and physics in the same manner that their colleagues in these disciplines are entering chemistry. Is this phenomenon only due to "in the land of the blind one-eye is king"? Pople has once stated, and made clear in his famous diagram [196], that it is in the area of the medium-sized molecules, which is the heart of chemistry, where progress is the slowest. Undoubtedly he was, and is, right. There is something else, however.

The amount of detailed information that is available from experimental research in chemistry is staggering. Theory has the obligation of providing the terms of reference for an experimental science, such that valid predictions become possible. In chemistry this is only realized in a very general way [197]. It has been the experience, not altogether a happy one, of many theoreticians, that the questions they can answer are not always the

ones that are being asked. This may be more so in chemistry than elsewhere. Nevertheless, it has become routine to question one's friendly program on the behaviour of one's molecule of choice before measuring or synthesizing. Chemistry does not have theories of its own. It does have a lot of concepts, ideas and "rules", but they allow for exceptions, notwithstanding Woodward's [198]: "Violations. There are none!". Therefore it may be expected that the future of theoretical organic chemistry lies more in calculations than in new concepts of general validity. This may not be easy to accept for everyone, and it is at the heart of the perennial discussion between the Group I and Group II quantum chemists as Coulson [199] has named them. Still there is no real reason for pessimism as Karplus [200] has argued. When you get stuck in the plane of quantum chemistry, make a leap into the third dimension!

References

1. Theoretical Organic Chemistry, Proc.Kekulé Symposium, Butterworth, London, 1959.
2. A.Kekulé, Ann., **106** (1858) 129.
3. J.Walker, J.Chem.Soc., **121** (1922) 735.
4. D.W.Theobald, Chem.Soc.Rev., **5** (1976) 203.
5. M.J.Nye, From Chemical Philosophy to Theoretical Chemistry, Univ. Cal.Press, Berkeley and Los Angeles, CA, 1993.
6. P.E.van der Vet, Thesis, Amsterdam, 1987.
7. L.J.Oosterhoff, Position 12, Thesis, Leiden, 1949.
8. L.J.Oosterhoff, Inaugural Lecture, p.8, Univ.Press, Leiden, 1950.
The source is possibly the Leiden theoretical physicist Prof.Dr H.A.Kramers.
9. T.S.Kuhn, The Structure of Scientific Revolutions, Univ.Chicago Press, Chicago, Il, 1970.
10. J.W.McAllister, Beauty and Revolution in Science, Cornell Univ.Press, Ithaca, NY, 1996.
11. O.N.Witt, Ber., **9** (1876) 522; **21** (1888) 321.
12. W.Dilthey, J.prakt.Chem., [2] **109** (1925) 273.
13. R.Wizinger, Angew.Chem., **39** (1926) 564.
14. Th.de Donder, Compt.rend.Acad.Sci., **185** (1927) 698.
15. A.Haas, Quantenchemie, Akad.Verlag, Leipzig, 1929.
16. L.P.Hammett, Physical Organic Chemistry, McGraw-Hill, New York, NY, 1940.
17. A.Streitwieser,Jr, Molecular Orbital Theory for Organic Chemists, J. Wiley, New York, NY, 1961.
18. M.J.S.Dewar, The Molecular Orbital Theory of Organic Chemistry, McGraw-Hill, New York, NY, 1969.
19. F.Henrich, Theorien der Organischen Chemie, 5^e Aufl. Vieweg, Braunschweig, 1924.

20. G.E.K.Branch, M.Calvin, *The Theory of Organic Chemistry; an Advanced Course*, Prentice Hall, New York, NY, 1945.
21. G.W.Wheland, *Resonance in Organic Chemistry*, J. Wiley, New York, NY, 1955.
22. W.Hückel, *Theoretische Grundlagen der Organischen Chemie I, II*, Leipzig, Akad.Verlag, 1931-1956.
23. B.Pullman, A.Pullman, *Les Théories Électroniques de la Chimie Organique*, Masson Éd., Paris, 1952.
24. P.H.Hermans, *Inleiding tot de Theoretische Organische Chemie*, Elsevier, Amsterdam, 1952.
25. H.A.Staab, *Einführung in die Theoretische Organische Chemie*, Verlag Chemie, Weinheim, 1960.
26. A.Liberles, *Introduction to Theoretical Organic Chemistry*, McMillan, New York, NY, 1968.
27. T.H.Lowry, K.Schueller Richardson, *Mechanism and Theory in Organic Chemistry*, Harper Int.Ed., New York, NY, 1987.
28. C.Sandorfy, *Les Spectres Electroniques en Chimie Théorique*, Ed.Revue d'Optique, Paris, 1959.
29. S.S.Shaik, H.B.Schlegel, S.Wolfe, *Theoretical Aspects of Physical Organic Chemistry*, J. Wiley, New York, NY, 1992.
30. L.Salem, *Electrons in Chemical Reactions: First Principles*, J. Wiley, New York, NY, 1982.
31. A.Sc.Couper, *Compt.Rend.hebd.Séances Acad.Sci.*, **46** (1858) 1157.
32. E.Frankland, *Phil.Trans.*, **142** (1852) 417.
33. A.Crum Brown, *On the Theory of Chemical Combination*, Thesis, Edinburgh, 1861.
34. J.H.van't Hoff, *Sur les formules de structure dans l'espace*, *Arch.néerl.des sciences exactes et naturelles*, **9** (1874) 445.
35. J.A.Le Bel, *Bull.Soc.chim.France*, [2] **22** (1874) 337.
36. H.Mackle, *J.Chem.Ed.*, **31** (1954) 618.
37. D.H.Rouvray, *J.Mol.Struct.(THEOCHEM)*, **259** (1992) 1.
38. H.Kolbe, *J.prakt.Chem.*, [2] **10** (1874) 450.
39. S.Cannizzaro, *Nuovo Cimento*, **7** (1858) 321.

40. A.Kekulé, *Ann.*, **137** (1865-66) 158; *Ann.*, **162** (1872) 88.
41. H.A.Staab, *Angew.Chem.*, **70** (1958) 39.
"Dasselbe Kohlenstoffatom ist also in der ersten Zeiteinheit mit einem der beiden benachbarten, in der zweiten dagegen mit dem anderen der benachbarten Kohlenstoffatome in doppelter Bindung; und man sieht daher, dass jedes Kohlenstoffatom zu seinen beiden Nachbarn genau in derselben Beziehung steht."
42. J.Dewar, *Proc.Roy.Soc.Edinburgh*, (1866-67) 82.
43. E.Hückel, *Z.für Elektrochemie*, **43** (1937) 760.
44. A.Ladenburg, *Ber.*, **2** (1869) 140.
45. E.E.van Tamelen, S.P.Pappas, *J.Am.Chem.Soc.*, **85** (1963) 3297.
46. K.E.Wilzbach, J.S.Ritscher, L.Kaplan, *J.Am.Chem.Soc.*, **89** (1967) 1031.
47. T.J.Katz, N.Acton, *J.Am.Chem.Soc.*, **95** (1973) 2738.
48. Ad.Claus, *Theoretische Betrachtungen und deren Anwendung zur Systematik der organischen Chemie*, S.207, Freiburg in Br., 1867.
49. D.H.Rouvray, *J.Mol.Struct.(THEOCHEM)*, **185** (1989) 1.
50. A.Cayley, *Ber.*, **8** (1875) 1056.
51. W.K.Clifford, *Am.J.Math.*, **1** (1878) 126, 350.
52. J.J.Sylvester, *Am.J.Math.*, **1** (1878) 64.
53. P.Gordan, W.Alexejeff, *Z.physikal.Chem.*, **35** (1900) 610.
54. G.Rumer, E.Teller, H.Weyl, *Gött.Nachr.*, (1932) 499
55. J.Paldus, C.R.Sarma, *J.Chem.Phys.*, **83** (1985) 5135.
56. A.Bayer, *Ber.*, **18** (1885) 2269, 2277.
57. H.Sachse, *Ber.*, **23** (1890) 1363.
58. E.Mohr, *J.prakt.Chem.*, [2] **98** (1918) 315; *Ber.*, **55** (1922) 230.
59. C.W.Kern, R.M.Pitzer, O.J.Sovers, *J.Chem.Phys.*, **60** (1974) 3583.
60. J.Thiele, *Ann.*, **306** (1899) 87, 369.
61. H.E.Armstrong, *J.Chem.Soc.*, **51** (1887) 263.
62. E.Bamberger, *Ann.*, **257** (1890) 47.

63. D.Vorländer, Ann., **320** (1902) 111.
"Die Radicale haben ein Doppelnatur, der einerseits der negativen und positiven Natur der Elemente, andererseits dem gesättigten und ungesättigten Zustande entspricht."
64. J.W.van Spronsen, The Periodic System of Chemical Elements, Elsevier, Amsterdam, 1969.
65. R.Abbegg, Z.anorg.Chem., **39** (1904) 330.
66. P.Drude, Ann.Physik, **14** (1904) 722.
67. G.N.Lewis, J.Am.Chem.Soc., **38** (1916) 762; J.Chem.Phys., **1** (1933) 17.
68. N.Bohr, Phil.Mag., **26** (1913) 1, 476, 857.
69. N.Bohr, Z.für Physik, **2** (1920) 423.
70. I.Langmuir, Proc.Nat.Ac.Sci., **5** (1919) 252.
71. Ref.[5], p.204-205.
72. E.Weitz, T.König, Ber., **55** (1922) 2868.
"...und glauben dass der Absättigungszustand einzelnes Moleküls beliebig zwischen den beiden (real kaum existierenden) Extremformen a und b liegen kann. ändert sich dann nicht das Mengenverhältnis der beiden tautomeren (*sic!*) Molekülarnten, sondern die sämtliche Moleküle ändern Ihren Zustand"
73. F.Arndt, E.Scholz, P.Nachtwey, Ber., **57** (1924) 1906.
74. W.Heisenberg, Z.für Physik, **38** (1926) 411.
75. C.K.Ingold, E.H.Ingold, J.Chem.Soc., **129** (1926) 1310.
76. B.Eistert, Angew.Chem., **52** (1939) 358.
77. M.Born, J.R.Oppenheimer, Ann.der Phys., **84** (1927) 457.
78. J.W.Armit, R.Robinson, J.Chem.Soc., **127** (1925) 1605.
79. E.Hückel, Z.für Physik, **70** (1931) 204
80. G.W.Wheland, L.Pauling, J.Am.Chem.Soc., **57** (1935) 2086.
81. D.Vorländer, Ber., **52** (1919) 274.
82. T.M.Lowry, Chem. & Ind., (1925) 970.
"...that the quinonoid theory of colour makes the o and p positions similar and the meta position unique, whilst steric hindrance and the coordination observed by Sidgwick makes the o position unique and the m and p positions similar, whereas

the experimental facts appear to insist most strongly on the similarity of the o and m positions and the uniqueness of the p position in the isomeric di-derivatives of benzene."

83. R.Grinter, E.Heilbronner, *Helv.Chim.Acta*, **45** (1962) 2496.
84. L.Pauling, *The Nature of the Chemical Bond*, Cornell Univ.Press, Ithaca, NY, 1939-1960.
85. L.Pauling, *J.Chem.Phys.*, **1** (1933) 280.
86. J.C.Slater, *Phys.Rev.*, **35** (1930) 210.
87. H.C.Longuet-Higgins, *Proc.Phys.Soc.*, **60** (1948) 270.
88. J.A.Lennard-Jones, *Trans.Far.Soc.*, **25** (1929) 668.
89. G.W.Wheland, *Trans.Far.Soc.*, **33** (1937) 1499.
90. D.P.Craig, *Proc.Roy.Soc.*, **A202** (1950) 498.
91. J.J.C.Mulder, *Nouv.J.Chimie*, **4** (1980) 283.
92. U.Opik, M.H.L.Pryce, *Proc.Roy.Soc.*, **A238** (1957) 425.
93. H.A.Jahn, E.Teller, *Proc.Roy.Soc.*, **A161** (1937) 220.
94. G.B.Shaw, *J.Phys. A: Math.Nucl.Gen.*, **7** (1974) 1537.
95. R.S.Berry, S.A.Rice, J.Ross, *Physical Chemistry*, p.110, J. Wiley, New York, NY, 1980.
96. E.Heilbronner, *Helv.Chim.Acta*, **45** (1962) 1722; *The Resonance Formulation of Electronically Excited π -Electron States*, p.329 in *Molecular Orbitals in Chemistry, Physics, and Biology*, a Tribute to R.S.Mulliken, Ac.Press, New York, NY, 1964.
97. P.Durand, J.P.Malrieu, *Adv.Chem.Phys.*, **67** (1987) 321.
98. F.Bernardi, M.Olivucci, M.A.Robb, *J.Am.Chem.Soc.*, **114** (1992) 1606.
99. L.Pauling, *J.Am.Chem.Soc.*, **53** (1931) 1367.
100. J.C.Slater, *Phys.Rev.*, **37** (1931) 481.
101. D.B.Cook, *J.Mol.Struct.(THEOCHEM)*, **169** (1988) 79.
102. J.F.Ogilvie, *J.Chem.Ed.*, **67** (1990) 280.
103. W.Kutzelnigg, *J.Mol.Struct.(THEOCHEM)*, **169** (1988) 403.
"That localized orbitals often look as if they repel each other....is to a large extent an artifact of the localization procedure."

104. W.Kutzelnigg, Einführung in die Theoretische Chemie, 1.Quantenmechanische Grundlagen, 2.Die chemische Bindung, p.36, Verlag Chemie, Weinheim, 1978.
105. H.Hellmann, Einführung in die Quantenchemie, p.121, F.Deuticke, Leipzig, 1937.
106. K.Ruedenberg, Rev.Mod.Phys., **34** (1962) 326.
107. P.A.M.Dirac, Proc.Roy.Soc., **A123** (1929) 714.
108. M.Goeppert-Mayer, A.L.Sklar, J.Chem.Phys., **6** (1938) 645.
109. C.A.Coulson, I.Fischer, Phil.Mag., **40** (1949) 386.
110. R.G.Parr, D.P.Craig, I.G.Ross, J.Chem.Phys., **18** (1950) 1561.
111. J.Gerratt, W.N.Lipscomb, Proc.Natl.Acad.Sci., **59** (1968) 332; J.Gerratt, Adv.At.Mol.Phys., **7** (1971) 141.
112. D.L.Cooper, J.Gerratt, M.Raimondi, Nature, **323** (1986) 699.
113. J.E.Lennard-Jones, J.Turkevich, Proc.Roy.Soc., **A158** (1937) 297.
114. G.W.Wheland, Proc.Roy.Soc., **A164** (1938) 397.
115. J.J.C.Mulder, J.Chem.Ed., to be published.
116. H.Eyring, J.Walter, G.E.Kimball, Quantum Chemistry, J. Wiley, New York, NY, 1944-1960.
117. G.E.Kimball, J.Chem.Phys., **8** (1940) 188.
118. J.J.C.Mulder, J.Chem.Ed. **62** (1985) 377 and references therein.
119. Colloque International de la Liason Chimique, J.chim.phys., **46** (1949) 185, 497, 675.
120. C.C.J.Roothaan, Rev.Mod.Phys., **23** (1951) 69; J.Mol.Struct.(THEOCHEM), **234** (1992) 1.
121. J.A.Pople, Trans.Far.Soc., **49** (1953) 1375.
122. R.Pariser, R.G.Parr, J.Chem.Phys., **21** (1953) 466, 767.
123. J.Lennard-Jones, Chem. & Ind., (1954) 1156.
124. S.F.Boys, Proc.Roy.Soc., **A200** (1950) 542.
125. J.Lennard-Jones, Proc.Roy.Soc., **A198** (1949) 1, 14; G.G.Hall, J.Lennard-Jones, *ibid.* **A202** (1950) 155; J.Lennard-Jones, J.A.Pople, *ibid.* **A202** (1950) 166; J.A.Pople, *ibid.* **A202** (1950) 323; G.G.Hall, *ibid.* **A202** (1950) 336; G.G.Hall, J.Lennard-Jones, *ibid.* **A205** (1951) 357; G.G.Hall, *ibid.* **A205** (1951) 541;

- J.Lennard-Jones, J.A.Pople, *ibid.* **A210** (1951) 190; G.G.Hall, *ibid.* **A213** (1952) 102, 113; A.C.Hurley, J.Lennard-Jones, *ibid.* **A216** (1953) 1; A.C.Hurley, *ibid.* **A216** (1953) 424; A.C.Hurley, J.Lennard-Jones, *ibid.* **A218** (1953) 327; A.C.Hurley, *ibid.* **A218** (1953) 333; A.C.Hurley, J.Lennard-Jones, J.A.Pople, *ibid.* **A220** (1953) 446.
126. G.G.Hall, J.Mol.Struct.(THEOCHEM), **234** (1992) 13.
127. W.Moffitt, Proc.Roy.Soc. **A196** (1949) 510.
128. C.A.Coulson, H.C.Longuet-Higgins, Proc.Roy.Soc., **A191** (1947) 39; *ibid.* **A192** (1948) 16; *ibid.* **A193** (1948) 447, 456; *ibid.* **A195** (1948) 188; B.H.Chirgwin, C.A.Coulson, *ibid.* **A201** (1950) 196; C.A.Coulson, J.Jacobs, *ibid.* **A206** (1951) 287; C.A.Coulson, D.P.Craig, J.Jacobs, *ibid.* **A206** (1951) 297.
129. C.A.Coulson, Valence, Oxford Univ.Press, 1951, 1961;
R.McWeeny, Coulson's Valence, Oxford Univ.Press, 1979.
130. K.Fukui, T.Yonezawa, C.Nagata, J.Chem.Phys., **26** (1957) 831.
131. N.Mataga, K.Nishimoto, Z.Phys.Chem., **13** (1957) 140.
132. R.McWeeny, Proc.Roy.Soc., **A223** (1954) 63, 306; **A227** (1955) 288.
133. P.O.Löwdin, J.Chem.Phys., **18** (1950) 356.
134. W.Campion, M.Karplus, Mol.phys., **25** (1972) 921.
135. D.P.Craig, Proc.Roy.Soc., **A200** (1950) 272, 390, 401, 474.
136. J.A.Pople, D.L.Beveridge, Approximate Molecular Orbital Theory, McGraw-Hill, New York, NY, 1970.
137. M.J.S.Dewar, E.Haselbach, J.Am.Chem.Soc., **92** (1970) 590.
138. M.J.S.Dewar, W.Thiel, J.Am.Chem.Soc., **99** (1977) 4899.
139. R.Hoffmann, J.Chem.Phys., **39** (1963) 1397.
140. Gaussian '72 - '94, J.A.Pople *et al.*, Gaussian Inc., Pittsburgh, PA.
141. F.H.Westheimer, Chapter 12 in M.S.Newman, Steric Effects in Organic Chemistry, J. Wiley, New York, NY, 1956.
142. S.Lifson, A.Warshel, J.Chem.Phys., **49** (1968) 5116.
143. J.Ph.Bowen, N.L.Allinger, Chapter 3 in Reviews in Computational Chemistry, **2** K.B.Lipkowitz, D.B.Boyd (eds.), VCH Publ., 1991.

144. J.H.van Vleck, A.Sherman, *Rev.Mod.Phys.*, **7** (1935) 167.
p.175: "This diagram might well be on the walls of chemistry buildings, being almost worthy to occupy a position beside the Mendeléef periodic table so frequently found thereon."
145. A.D.Walsh, *J.Chem.Soc.*, (1953) 2260, 2266, 2288, 2296, 2301, 2306, 2318.
146. R.S.P.Gillespie, *Molecular Geometry*, Van Nostrand, New York, NY, 1972.
147. J.v.Neumann, E.Wigner, *Physik.Zeitschr.*, **30** (1929) 467.
148. E.R.Davidson, *J.Am.Chem.Soc.*, **99** (1977) 397.
149. H.C.Longuet-Higgins, E.W.Abrahamson, *J.Am.Chem.Soc.*, **87** (1965) 2045.
150. J.J.C.Mulder, *Valence Bond Theory for Chemical Reactions* p.355 in *Valence Bond Theory and Chemical Structure*, D.J.Klein and N.Trinajstić (eds.), Elsevier, Amsterdam, 1990.
151. R.B.Woodward, R.Hoffmann, *The Conservation of Orbital Symmetry*, Verlag Chemie, Weinheim and Acad.Press, New York, NY, 1970.
152. W.Th.A.M.van der Lugt, L.J.Oosterhoff, *Chem.Comm.*, (1968) 1235;
J.Am.Chem.Soc., **91** (1969) 6042.
153. Ref.[30], p.135-136.
154. L.Salem, C.Rowland, *Angew.Chem.*, **84** (1972) 86.
155. H.C.Longuet-Higgins, *Proc.Roy.Soc.*, **A344** (1975) 147.
156. G.Herzberg, H.C.Longuet-Higgins, *Disc.Far.Soc.*, **35** (1963) 77.
157. E.Teller, *J.Phys.Chem.*, **41** (1937) 109; *Isr.J.Chem.*, **7** (1969) 227.
158. C.M.Meerman-van Benthem, A.H.Huizer, J.J.C.Mulder, *Chem.Phys.Lett.*, **51** (1977) 93.
159. A.J.Stone, *Proc.Roy.Soc.*, **A351** (1976) 148.
160. L.Landau, E.Lifchitz, *Mécanique Quantique; théorie non relativiste*, p.334, Éditions Mir, Moscou, 1967.
161. W.Kauzmann, *Quantum Chemistry*, footnote on p.696, Ac.Press, New York, NY, 1957-1958.
162. J.A.Pople, Y.Aviva Sataty, E.Amitai Halevi, *Isr.J.of Chemistry*, **19** (1980) 290.
163. F.London, *Methoden der modernen Physik*, (Sommerfeld Festschrift), p.111, S.Hirzel, Leipzig, 1928.

164. F. Bernardi, M. Olivucci, M. A. Robb, *J. Am. Chem. Soc.*, **112** (1990) 1737.
165. J. Michl, V. Bonačić-Koutecký, *Electronic Aspects of Organic Photochemistry*, p.22, J. Wiley, New York, NY, 1990.
166. J. J. C. Mulder, L. J. Oosterhoff, *Chem. Comm.*, (1970) 305, 307.
167. Ref.[10], p.39: (i) Form of symmetry; (ii) Invocation of a model; (iii) Visualization and abstractness; (iv) Metaphysical allegiance.
168. J. Paldus, *J. Chem. Phys.*, **61** (1974) 5321.
169. F. A. Matsen, *J. Mol. Struct.(THEOCHEM)*, **259** (1992) 65.
170. J. Koutecký, K. Hlavatý, P. Hochmann, *Theor. Chim. Acta*, **3** (1965) 341.
171. W. Th. A. M. van der Lugt, L. J. Oosterhoff, *Mol. Phys.*, **18** (1970) 177.
172. H. Weyl, *The Theory of Groups and Quantum Mechanics*, p.385, Dover Publ. Inc., 1931.
173. J. J. C. Mulder, *Mol. Phys.*, **10** (1966) 479.
174. I. Shavitt, *Int. J. Quantum Chem.*, Symp. **12** (1978) 5.
175. S. F. Boys, G. B. Cook, C. M. Reeves, I. Shavitt, *Nature*, **178** (1956) 1207.
176. J. C. Polanyi, *Angew. Chem. Int. Ed. Engl.*, **26** (1987) 952.
177. Y. T. Lee, *Angew. Chem. Int. Ed. Engl.*, **26** (1987) 939.
178. R. E. Wyatt, *ACS Symp. Ser.*, **56**, State to State Chem. Symp., (1977) 185.
179. E. J. Heller, *Acc. Chem. Res.*, **14** (1981) 368.
180. *The Theory of Reaction Dynamics*, D. C. Clary ed., NATO ASI Series, Series C, Mathematical and Physical Sciences, **170**, D. Reidel, Dordrecht, 1986.
181. J. J. C. Mulder, J. S. Wright, *Chem. Phys. Lett.*, **5** (1970) 445.
182. B. T. Sutcliffe, *J. Mol. Struct.(THEOCHEM)*, **259** (1992) 29; **341** (1995) 217.
183. R. G. Woolley, *J. Am. Chem. Soc.*, **100** (1978) 1073; *J. Mol. Struct.(THEOCHEM)*, **230** (1991) 17.
184. S. J. Weininger, *J. Chem. Ed.*, **61** (1984) 939.
185. P. Pulay, *Adv. Chem. Phys.*, **69** (1987) 241.
186. A. Komornicki, J. W. McIver, *J. Chem. Phys.*, **70** (1979) 2014.

187. N.C.Handy, H.F.Schaefer III, *J.Chem.Phys.*, **81** (1984) 5031.
188. H.B.Schlegel, *Adv.Chem.Phys.*, **67** (1987) 249.
189. T.Helgaker, P.Jørgensen, *Adv.Quantum Chem.*, **19** (1988) 183.
190. J.Gauss, D.Cremer, *Adv.Quantum Chem.*, **23** (1992) 205.
191. K.Fukui, *Acc.Chem.Res.*, **14** (1981) 363.
192. J.N.Murrell, K.J.Laidler, *Trans.Far.Soc.*, **64** (1968) 371; J.N.Murrell, G.L.Pratt, *Trans.Far.Soc.*, **66** (1970) 1680.
193. L.Salem, *J.Am.Chem.Soc.*, **92** (1970) 4472.
194. R.E.Stanton, J.W.McIver, *J.Am.Chem.Soc.*, **97** (1975) 3632.
195. P.Pechukas, *J.Chem.Phys.*, **64** (1976) 1516.
196. J.A.Pople, *J.Chem.Phys.*, **43** (1965) S229.
197. J.J.C.Mulder, Thesis, Leiden, 1970.
p.12: "Such is the essential nature of chemistry that the subject is an impossible one if analogy arguments cannot be made operative."
198. Ref.[157], p.173.
199. C.A.Coulson, *Rev.Mod.Phys.*, **32** (1960) 170.
200. M.Karplus, *J.Phys.Chem.*, **94** (1990) 5435.

This Page Intentionally Left Blank

Inter-Relations between VB & MO Theories for Organic Π -Networks

D. J. Klein

Texas A & M University - Galveston
Galveston, Texas 77553-1675 USA

The rather disparately developed Valence-Bond (VB) and Molecular-Orbital (MO) theories are reviewed and compared, with special attention to the possibility of concordances of general theoretical predictions.

1. Broad Motivation and Aim - Graph Theory

Valence-Bond (VB) and Molecular-Orbital (MO) theories both were clearly formulated by the end of the first decade of quantum mechanics. Of course VB theory is connected to early conceptual roots in chemistry, as emphasized by Rumer [1] and more particularly by Pauling, in a review [2] and then in his masterwork [3] *The Nature of the Chemical Bond*. Thence for some period of time VB theory seems in the chemical community to have been viewed quite favorably.

The chemical relevance of MO theory was perhaps clearly demonstrated first by Hückel [4] in the early 1930s, but the dominance of MO theory did not set in until after further conceptual development from the "English school" and the systematic computational self-consistent-field format of Roothaan [5] (and Hall [6]). Still within the MO-theoretic framework various theorems were found [7,8], the "Woodward-Hoffmann" rules for concerted reactions were developed [9,10], and powerful "black-box" MO-based computer programs became both widely available and widely utilized.

But presumably because of VB theory's connections to classical chemical bonding concepts, work continued in this area and has even intensified in recent decades, there now being many new-theorematic, conceptual, and computational developments, as evidenced in the reviews in the-book [11]

Valence Bond Theory and Chemical Structure, where also some further detail in the here highly summarized history is provided.

A natural question concerns the possibility of mathematical interrelations between these VB and MO theories, especially since they each may be viewed to be based on rather disparate pictures: the VB picture being built from localized orbitals arranged into (one or more) correlated many-electron structures, whereas the MO picture is built from delocalized orbitals in a single uncorrelated (determinantal) structure. Thence naively one might expect these two theories to be contradictory in their simplest forms, but perhaps more fundamentally, one may view each picture as *complementary* in Bohr's sense [12]. That is, each theoretical picture is viewed as naught but a different representation of the same underlying quantum reality - a thesis which is in fact strangely reminiscent of Plato's metaphor [13] with our perceived world being only as differing views of shadows cast by a flickering fire on a cave wall, the reality casting the shadows being never directly viewed. With such ideas in mind one then might almost expect correspondences of prediction, despite the disparate underlying pictures. Of course, at least at the simplest levels of the two theories the possibility of predictive contradictions due to fundamental inadequacies might also be imagined.

Here then it is intended to note some inter-relations between VB and MO theories, particularly as regards general predictive correspondences. With the focus on *generality* of correspondences the emphasis naturally shifts to the simpler (typically semiempirical) models for which it is easier to obtain general (e.g. theorematic) results. Thence here attention is focused on the rather well studied such models for π -electron networks in neutral (i.e. non-ionic) organic molecules.

Some aspects of the questions about mathematical inter-relations concern the presentation in terms of molecular *graphs*, these being natural mathematical representations of the classical valence structures of molecules. Such a graph G consists firstly of sites corresponding to atomic π -centers and secondly of edges corresponding to bonds between pairs of atoms. Now G is conveniently represented by its *adjacency* matrix \mathbf{A} which has rows (and columns) labelled by sites of G and all of whose elements A_{ij} are 0 excepting unit elements corresponding to nearest-neighbor pairs of sites. Evidently the mathematical field [14] of graph theory provides [15] a natural framework for classical chemical bonding ideas and then presumably too for much of modern quantum chemistry, the simple VB and MO models for a particular molecule being entirely determined by their graphs. The general inter-relations considered between VB and MO models then can be expected to be expressed in terms of graph-theoretic language, for general classes of molecules.

That both models are expressed in terms of the molecular graph G (or the corresponding \mathbf{A}) implies that the various consequent properties are

also graph-theoretic. But at first glance each model seems to ascribe rather different graph-theoretic properties to the same physico-chemical properties, so that there is a natural question as to whether such different graphical properties might be inter-related to give like predictions for physico-chemical properties. Further manifestations that both models are graph-theoretic (as is already known *a priori*) is not the basic point that is here sought, though such manifestations are sometimes quoted as evidence of an inter-relation. For instance, the bond orders of Pauling (within the VB picture) and of Coulson (within the MO picture) are [15] both expressible neatly in terms of \mathbf{A} . But these graph-theoretic expressions are different (though both are functions of \mathbf{A} , this is only expected since both simple models are expressed in terms just of \mathbf{G} , or equivalently of \mathbf{A}), so this does not of itself imply an inter-relation of current interest, unless e.g. the two graph-theoretic expressions were shown to be numerically proportional - and this does not seem to have been done in an analytic manner, though starting with Ham & Ruedenberg [17] empiric numerical correspondences were noted. Here the present interest is with predictive physico-chemical correspondences of a general analytic nature.

2. VB and MO Models

Both the simple VB and MO models for organic π -networks are quantum-mechanical models explicitly expressible in terms of their molecular graphs. For the neutral molecule the (Pauling-Wheland) VB model assigns one π -electron (spin up or down) to each (carbon) center, so that for N sites the (2^N -dimensional) model space is spanned by products of N different electron spins. Then the simple VB Hamiltonian may be written as

$$H_{\text{VB}} = J \sum_{i-j} 2 \mathbf{s}_i \cdot \mathbf{s}_j$$

where J is the "exchange" parameter, the sum is over bonded pairs of sites in the graph G , and \mathbf{s}_i is the spin operator for site i . In physics the model is known as the (isotropic spin-1/2) *Heisenberg* model, and $2 \mathbf{s}_i \cdot \mathbf{s}_j + 1/2$ is the operator which effects electron exchange between the two centers.

The simple Hückel MO model assigns one spatial orbital to each site which thence may be occupied by 0, 1, or 2 π -electrons, so that the model space is spanned by 4^N orbital products with different numbers of electrons, and for the neutral molecule there is reduction to a subspace with a dimension which is the coefficient of x^N in $\{1+2x+x^2\}^N$. Then the simple MO Hamiltonian may be written as

$$H_{\text{MO}} = \beta \sum_{i-j} (E_{i+j} - E_{j+i})$$

where the sum is again over bonded pairs of sites in G , β is the "electron-transfer" parameter, and E_{i-j} is the operator transferring an electron from center i to j (while leaving its spin unchanged). The exchange operator (of the VB theory) also can be expressed in terms of the electron-transfer operators (of the MO model), namely as $E_{i-j} E_{j-i} = E_{i-i}$. But as long as one does not add corrections to H_{MO} for explicit electron-electron interaction the model is really mathematically simpler than the above equation might suggest - basically each electron may be treated independently and the energy given as a sum of orbital eigenvalues which (up to a factor β and sometimes a shift α) are just those of the adjacency matrix \mathbf{A} of the graph G .

Graph-theoretic methods and theorems might naturally be anticipated to apply for either of these simple models, or indeed for several modifications or even extensions of either type of model. That is, once the (site-number dependent) space on which the model acts is chosen, the remaining part of either type of model is specified entirely by the graph G , which in turn is specified by only part of the classical *valence structure*, namely that part with fixed (or localized) bonds. The dimensions of the full spaces on which either model is defined increase exponentially, with the number N of sites, and particularly for the VB model the direct numerical matrix-diagonalization then also increases exponentially with N , even after taking into account the cosmopolitan spin (or unitary group) symmetries - see e.g. [18]. For the MO model the reduction to the diagonalization of \mathbf{A} leads to just a standard numerical problem, scaling as N^3 . Though here the focus is more on general rules or theorematic results, the greater numerical computational efficacy of the simple MO model might be expected to translate to a greater wealth of theorematic results. But if the MO model is elaborated or the VB model is simplified one might anticipate this "wealth" to be more balanced.

There are various approximations or simplifications for the VB model. Most notable here are the resonance-theoretic reductions which involve different local pairing patterns for the π -electrons each associated to a site. As emphasized by Rumer [1], Pauling [2,3], and Wheland [19] such pairing patterns represent the coupling of disjoint pairs of neighbor electron spins to a (local π -bond) spin singlet pair in the overall electronic wave-function which in turn corresponds to a classical chemical-bonding pattern, called a *Kekulé structure*. But in fact this simplification step is but one in a chemically meaningful sequence [20] indicated in Figure 1, appearing on the next page. Here each restriction step entails a reduction to a subspace spanned by a chemically meaningful subset of states, and each orthogonalization step entails the systematic (so-called [21] "symmetric") orthogonalization of the associated natural spanning subsets. The 1st restriction step (separating out the covalent π -electron subspace) is within Pauling's [2] work, the 1st orthogonalization

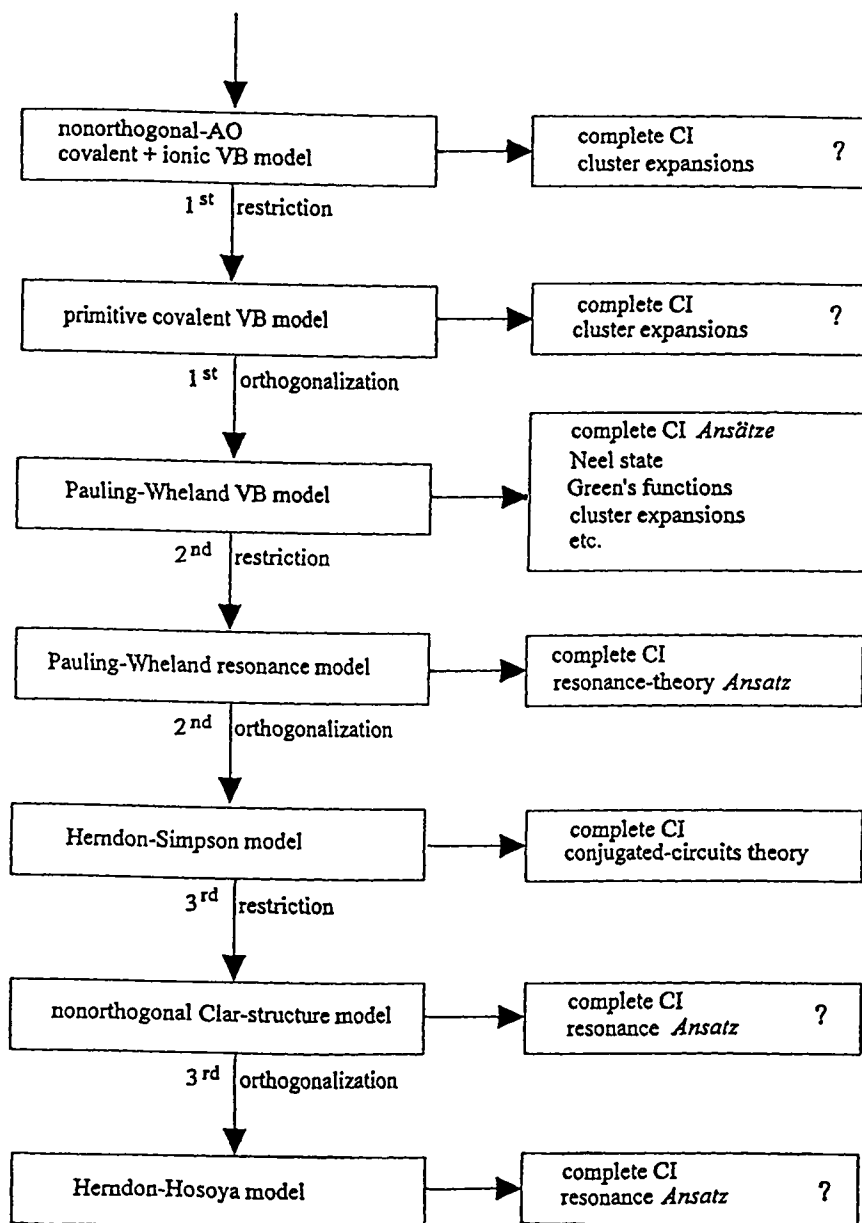


Figure 1. Hierarchical scheme for VB models and their solutions.

step (leading to H_{VB}) is well described in [20], the 2nd restriction step (to just the subspace spanned by Kekule structures) is that of Pauling and Wheland [19,22], and the 2nd orthogonalization step is developed in [23], though there is an (older) alternative existential route [24] to much the same end. From this consequent (Herndon-Simpson) model one may readily derive (via the invocation of a reasonable wave-function *Ansatz* for "benzenoid" structures, without unpaired electrons) the well-known "conjugated circuits" model [23,25,26], such being indicated in Figure 1 with the box to the right of the Herndon-Simpson model. Indeed this "conjugated-circuits" model can also be motivated from purely classical chemical-bonding ideas of Clar [27], as detailed elsewhere [28] - this empiric motivation is apparently close to that early independently made by Randić [29]. Further a 3rd restriction and subsequent 3rd orthogonalization lead to less-studied simpler models - again see [20] for some further comments and original references.

Finally as usual there is interest in computational methods for quantum-theoretically based models. In Figure 1 the boxes out to the right-hand side indicate something about such solution methods for the various models - the abbreviation CI is for "configuration interaction".

3. MO-Based Elaborations and Cross-Derivations

The derivational hierarchy of Figure 1 does not explicitly indicate any standard MO-theoretic model, though this can be rectified through the use of more elaborate many-body models, of the general PPP-Hubbard type [30,31], all defined on the same space as for H_{MO} , and still most simply dependent solely on the system graph G . Of these the first is the Hubbard model which is the sum of H_{MO} and a second electron-electron interaction term

$$V_1 = (U_1/2) \sum_i (E_{i-i}^2 - E_{i-i})$$

where U_1 is the Hubbard electron-repulsion parameter. When no more electron-electron interaction terms are to be added U_1 may be approximated [32] as the difference ($\gamma_0 - \gamma_1$) between the on-site and neighbor-site PPP parameters. Second is the Hubbard-PPP model which is the sum of H_{MO} , V_1 , and a third term

$$V_2 = U_2 \sum_{i-j} (1 - E_{i-i})(1 - E_{j-j})$$

where the sum is now over (bonded) neighbor pairs of sites of G and U_2 is a neighbor Coulomb interaction parameter. Here $1 - E_{i-i}$ is the operator counting the (instantaneous) net charge identified to a carbon π -center. When

$H_{MO}+V_1+V_2$ is used the coefficient U_1 should be just γ_0 while $U_2=\gamma_1-\gamma_2$, where γ_2 is a typical (or reference) next-neighbor PPP Coulomb repulsion parameter. Evidently the model could be further extended to include third or farther neighbor Coulomb interactions, possibly while still using a single new Coulomb interaction parameter U_3 referenced against a typical PPP Coulomb repulsion parameter γ_3 .

It is to be emphasized that all these mentioned models are reasonably taken as dependent on the particular molecular structure solely through just the molecular graph G , though the ordinary full PPP (or Pariser-Parr-Pople) model [30] depends on details of the molecular geometry, using Coulomb-interaction parameters (as well β_{ij} parameters) dependent on geometric distances rather than shortest-path graph distances. That is, the PPP parameters proposed for use in the graph-theoretic models utilize just γ_1 , γ_2 and γ_3 parameters taking just *typical (or reference)* values for nearest, next-nearest, and next-next-nearest neighbor Coulomb repulsions (and a similar modification could even be applied for the next-neighbor electron-hopping parameter β'). The purely G -dependent models are here referred to generically as *PPP-Hubbard* models. Purely G -dependent elaborations of the VB model have also been made [33]. Another type of modification would make the neighbor β -parameters of the Hückel model dependent on the Hückel-theoretic Coulson bond orders, the result remaining purely graph-theoretic. Rather similar comments apply to the J -parameters of the VB-model and the U -parameters of the PPP-Hubbard models. One sees in a fairly explicit manner that the purely graph-theoretic aspects of molecular π -electronic structure persist rather much beyond the simple Hückel model, which so often seems to be misconstrued as the limit of applicability of graph-theoretic ideas.

Now the diagram indicating the cross-derivational inter-relations between VB models and (MO-associated) PPP-Hubbard models is given in Figure 2, where for simplicity of presentation only a portion of Figure 1 more directly related to the VB results is reproduced. In Figure 2 the lead refinement step from the covalent+ionic VB model with nonorthogonal orbitals to the PPP-Hubbard model entails "symmetric" orthogonalization of the π -center atomic orbitals (AOs), as is quite distinct from the 1st orthogonalization step (of Figure 1) leading instead to the covalent VB model of Pauling and Wheland. In fact the present orbital orthogonalization step has been much studied [34,35] as to its effects in enhancing the electron-hopping (or resonance-integral) parameter β while simultaneously diminishing the electron exchange-parameter J . The restriction step leading from the PPP-Hubbard model to the covalent VB model has been repeatedly studied [33,36,37] and is different from the restriction steps of Figure 1 in that those of Figure 1 give a reasonable form of model at 1st order of degenerate perturbation theory while the step from the PPP-Hubbard model must proceed

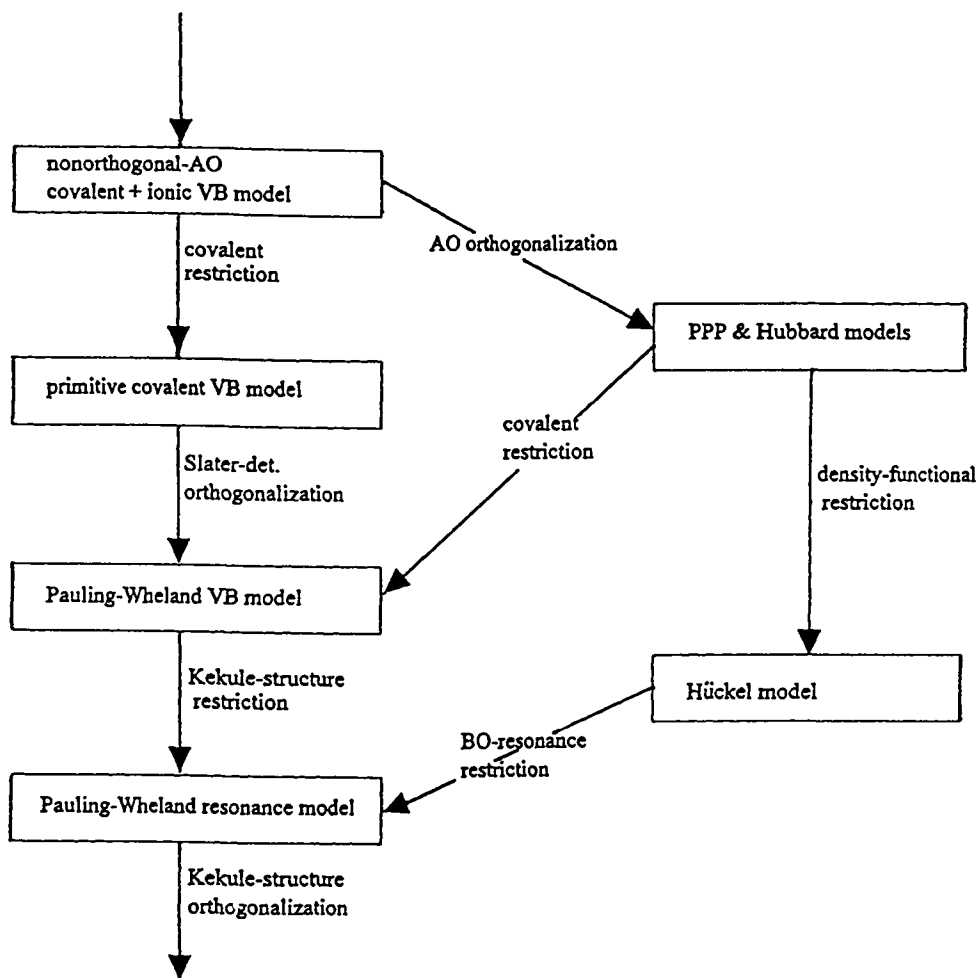


Figure 2. Cross-derivational inter-relations between the VB-hierarchy of models and the orthogonal-orbital (MO-associated) PPP-Hubbard and Hückel models.

to higher orders, with the qualitative nearest-neighbor form of H_{VB} being obtained at 2nd order, though for quantitative development of J in terms of the PPP-Hubbard model parameters requires much further higher orders, say as occur [33] in a straight-forward "cluster expansion".

The novel refinement step from the Hückel MO model to the Pauling-Wheland resonance-theoretic model entails a "bond-orbital" approximation due to Živković [38], and is found to apply best for the case of "benzenoid" graphs, being those without conjugated $4n$ -cycles in their Kekulé structures. First, bond-orbitals for G are formed from pairs of adjacent pairs of (orthogonalized) AOs; second, spin-up and -down pairs of electrons placed into such bond-orbitals are put together into determinants each of which then corresponds to a Kekulé structure; third, using these new "Kekulé structures" as a basis for a space on which to diagonalize the Hückel model, it is found that the matrix representation is essentially the same as that for the Pauling-Wheland resonance-theoretic model; and fourth, somewhat more ambiguously, this subspace on which the Hückel model is to be so diagonalized is argued to contain a good representation of the ground state. For non-benzenoids (with conjugated $4n$ -circuits) this bond-orbital-derived model yields a representation different than that of the simple nearest-neighbor Heisenberg (VB) model.

4. Hückel Rule

Hückel's rule indicating either stability or reactivity as a cycle has $4n$ or $4n+2$ electrons has been argued (e.g., in texts [39,40]) to be of central importance for all of theoretical organic chemistry. Moreover, contrary to occasional comments concerning this area there is a general degree of concordance between the predictions of VB and MO theory (though care needs to be invoked for the MO model avoiding the unjustified use of Hund's rule, and in the case of the VB model care needs to be taken with neutral $4n$ -cycles).

Hückel's MO work [4] in this area of course stands preeminent. But it has been much elaborated: first, as regards the possibility of a Möbius cycle (with one $\beta > 0$) whence [39,41] the electron-count conditions for stability and reactivity are interchanged, as indicated in Table I following; second, as regards the interpretation of Hückel rule as a basis for the Woodward-Hoffmann rules (at least for electrocyclic reactions); and third, as regards the rules applicability [39] to cycles even when embedded in more extensive π -networks.

Table 1. MO-based Hückel Rule for Annulenes

	$4n+2$	$4n$
regular cycle	closed shell	open shell
Möbius cycle	open shell	closed shell

The relevant VB-theoretic work seems less well known, with a first important result being with the work of Fischer & Murrell [42], which however is focused on the ionic case with the electron- and site-count being different. Basically they note a correspondence between the VB many-body basis states with net charges moved around and the AO basis of the 1-electron MO-model with electrons moved around.

The VB work of Oosterhoff *et al* [43] is relevant also to the neutral case but makes use of the non-orthogonalized VB model. Epiotis [44] also deals with the general case, possibly utilizing "anti-orthogonalized" AOs. Basically these workers note (beyond the exchange permutations) the importance of the cyclic permutations around the cycle, such as typically are discarded in the lowest order derivations to the Pauling-Wheland VB model. The inclusion of such terms is crucial most especially for 4-cycles - and such corrected models for quantitative work are available [33].

Even for PPP-Hubbard models some general ground-state theorems [45] are available, as regards spin and point-group symmetries of single cycles, as is summarized in Table 2. In this table the label of "biradical" is used as an indicator of a ground state of non-totally (i.e., not A_{1g}) symmetric point-group symmetry (as can only occur with radicaloid singly-occupied natural orbitals). The results agree with the already noted results for the VB models, and also bear a noted resemblance to the MO-based Table 1. But the VB model and the PPP model results of Table 2 give information about the exchange-splittings in the biradical species, such as is a somewhat delicate matter [46] when starting from the MO results, where a decision is needed as to the relevance of Hund's rule as is taken up in the next section in a more general context of possible structures.

Table 2. Correlated-electron Hückel Rule for Annulenes

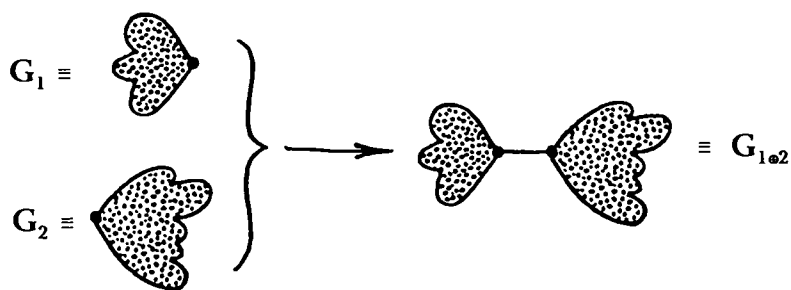
	$4n+2$	$4n$
regular cycle	$S = 0$	$S=1$ (ionic) or 0 (neutral) biradical
Möbius cycle	$S=1$ (ionic) or 0 (neutral) biradical	$S = 0$

Beyond the annulenes the ground-state spin symmetry and also the exchange splitting patterns amongst low-lying states of different spin multiplicities are of much interest, as is of relevance for a more general understanding of the magnetic properties of a molecule, novel possibilities arising especially with polyradical species. Though here one might anticipate that VB theory has a natural advantage over MO theory, here too there is a fair degree of concordance of predictions, once one goes beyond [46,47] the uncritical use of Hund's rule, the relevant MO-theoretic modification being originally developed in the work of Borden and Davidson [48]. For a bipartite (or alternant) species the ground-state spin evidently rather generally is just half the magnitude of the difference between the numbers ($\#_*$ and $\#_o$) of starred and unstarred sites

$$S = |\#_* - \#_o|/2$$

Indeed this is a theorem for the simple VB model [49] as well as the half-filled Hubbard model [50]. A related type of result applies [51] for π -network species decorated with carbene groups. Such results enable one to easily imagine various possibilities for high-spin species, as early noted by Ovchinnikov [52].

A final point might be mentioned as regards a note of Dewar and Longuet-Higgins [7] concerning a definition of the relative "parity" of Kekulé structures of a bipartite (or alternant) species. Two such Kekulé structures are given a *parity* difference which is the same as the parity of a permutation amongst the "starred" sites such as to carry one Kekulé structure into the other. After noting that $4n$ cycles have two Kekulé structures of opposite parity while $4n+2$ cycles have two of the same such parity, investigators have sometimes proposed that the magnitude of the difference between the numbers of Kekulé structures of different such parities might be a good indicator of stability, and suggested that such ideas constitute a "proper" derivation of VB-based ideas. But this mathematical observation has not been related to quantum underpinnings, VB- or MO-theoretic. This (non-negative) difference count often is called an *algebraic structure* count and denoted by $a(G)$. [Note that the Dewar-Longuet-Higgins "parity" is quite distinct from that which [53] earlier arose in the context of the VB model.] But now, it can argued that these algebraic-structure-count ideas are not well-related to quantum underpinnings, through the consideration of some judicious examples (as we now proceed to do). If one considers two π -network species G_1 and G_2 both of which involve Kekulé structures possibly with opposite parities, and one imagines these two species appended together, as



one obtains a new species $G_{1\bullet 2}$ which has $a(G_{1\bullet 2})=a(G_1) a(G_2)$. But because of this multiplicative feature one might imagine that it is $\log\{a(G)\}$ that is of physical relevance to measure a size-extensive thermodynamic stability, which is all reasonable at least if all the Kekulé structures are of the same parity. But now were G_2 to be just a 4-membered ring one obtains $a(G_{1\bullet 2})=0$ regardless of how large and how stable G_1 might be, so that it does not lead to what (say either $\log\{a(G_{1\bullet 2})\}$ or $\log\{1+a(G_{1\bullet 2})\}$) one would characterize as a size-extensive stability. That is, $a(G)$ does not seem to have a chemically plausible size-extensive behavior. But regardless it has been motivated from neither the VB- or MO-model - at most it has been identified a graph-theoretic invariant which exhibits some limited correlation with resonance energies.

5. Polymers and Excitations

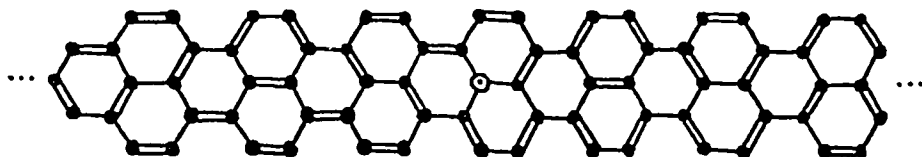
Beyond the question of similarity in general predictions for a system's ground state, there are questions concerning excited-state spectra. But results in this area seem to be more meager, perhaps because results for either type of model are more meager. Moreover, since the MO-model space includes the ionic structures while the simple covalent Pauling-Wheland model does not, there arises a natural limitation, in the absence of extension to include these excited structures.

Rather amusingly it turns out even at a very low level of description that there is a degree of concordance in general predictions concerning a class of conductive states at least for the class of "benzenoid" polymers. In particular within the framework of either the simplest Hückel model or of the simplest resonance-theoretic rationale it seems that the same structural conditions arise for the occurrence of Peierls-distortion and the sometimes associated solitonic excitations. For the simple Hückel model, starting with uniform β -parameters, such a structural condition is well-known [54-57] to be connected with a 0-band gap for which the fermi energy ϵ_F occurs at a rational multiple of the Brillouin-zone size, say at wave-vector $k=\pi p/q$ - then a distortion cutting the

Brillouin zone by q (and consequently increasing the unit cell by a factor q) opens a band gap at ϵ_F , whence a stabilization results shifting the occupied levels near ϵ_F downward and the unoccupied levels near ϵ_F upward. Moreover when there are degenerate such distortions solitonic excitations are expected [57] to arise at boundaries where a change from one such distortion to a degenerate one occurs. The now standard example is [55,56] that of *trans*-polyacetylene, where the Peierls distortion increases the unit-cell size from one to two π -centers via an alternation of bond lengths, with the two possible patterns of bond-length alternation being degenerate so that [56] solitonic excitations are expected in a region where a change from one pattern of bond-length alternation to the other is made. With the short and long bonds respectively identified as double and single a maximally rapid such change in such bond-length alternation can be pictured as



But in fact one can also view this as representing a VB structure for which there is a change in spin-pairing pattern from that associated to one of the two Kekulé structures to that associated to the spin-pairing pattern of the other Kekulé structure. Indeed for (cyclic-boundary-conditioned) *trans*-polyacetylene these two Kekulé structures differ in every position along the chain so that asymptotically they should not mix, and as a consequence the simple resonance-theoretic model gives [58] a doubly degenerate ground state corresponding to two patterns of alternation in Pauling bond orders and then too to two corresponding patterns of alternation in bond lengths. The VB structures with a lone unpaired electron separating these two spin-pairing patterns here also correspond to solitonic excitations. Even for finite (oligomeric) polyenes (without cyclic boundary conditions) such ideas [59] appear to be useful in understanding the ground state, the low-lying optically near-forbidden (homopolar) 1A and 3B excited states, and a related novel "sudden polarization" phenomenon. For rather general polymers of N monomer units there in fact remains a partitioning of Kekulé-structure patterns into different long-range-ordered classes. E.g., for



one sees that the Kekulé structure pattern as initiated on the left always has one horizontal double bond per naphthalene unit while that as initiated on the right always has two horizontal bonds per naphthalene unit - and with both types of patterns initiated on different parts of the polymer chain there must remain locally unpaired electrons in the region switching from one to the other. Such partitioning of Kekulé structures into different classes is such that every structure $|K\rangle$ of one class differs in $\sim N$ different positions from a structure $|K'\rangle$ in another class, and so can only weakly mutually interact, say (for normalized basis vectors) as $\langle K|H|K'\rangle \sim \kappa^N$ with $|\kappa| < 1$, so that in the high-polymer limit they are non-interacting. The different classes may be identified [60,61] with different *long-range spin-pairing orders* (where "order" here is used in the same sense as in statistical mechanics where it distinguishes different thermodynamic phases), and there are [61] a number of theorematic results. But also of much interest is [60] that the VB-based Peierls-distortion and associated possibility for solitons extends to arbitrary benzenoids in such a way that the final predictions correspond to those from the MO-based (or band-based) arguments. That is, within the resonance-theoretic model degenerate sets of non-interacting Kekule structures giving rise to solitonic excitations arise for the same benzenoid structures as one finds within the Hückel model lead to a zero band gap with degenerate Peierls distortions and associated solitonic excitations.

Some of these ideas seem to extend to make corresponding predictions concerning localized states at edges or ends of polymers. In particular, the different classes of Kekulé structures described seem strictly to occur with some sort of cyclic boundary conditions - that is, if one places particular ends on a polymer the spin-pairing patterns of no more than one of the classes seems consistently to pair all sites in the region of the ends. Thence [60] for a long polymer even if the class leading to the greatest bulk stability is not consistent with the considered polymer ends, it is still favored (overall) and leads to locally unpaired spins near the ends. There are entirely similar ideas applicable to the consideration of the (more extensive) edges, and rather interestingly quite similar final predictions seem [62] to come from a suitable consideration of the MO-based models, though here most of the work is as yet unpublished. But one important case has been considered by Dewar & Longuet-Higgins [7]: they show that MO- and resonance-theory unpaired spin densities for benzenoid monoradicals are in quite close correspondence. Herndon [63] discusses this relation.

6. Prospects

Evidently both VB and MO theories have a perhaps surprisingly wide degree of concordance in general predictions, as so far illuminated in our present review. It seems likely that there is more yet to be done, say as regards: extensions of theorems to more elaborated models, or new theorems for the already so studied species, or for ionic species, or for other possible molecular structures, say of non-alternants or of hetero-atomically substituted species. Aside from an analytic/theorematic approach to the correspondences between MO and VB approaches there too is the possibility of empiric computational investigations, many of which have been made, though they have not been closely followed here. Very often indeed one finds separate papers in either the MO or VB framework which separately conform to the same experimental reality, and thence also conform to one another. The view taken here makes an identification of predictions which are common to both disparate models and thence presumably more robust and reliable. Further the present approach provides a path for conceptual understanding. The rather wide-ranging correspondence of predictions between the VB and MO theories may be seen to be somewhat surprising in view of the historically frequent animosity between the advocates of the two theories.

Acknowledgement is made to the Welch Foundation of Houston, Texas.

References

- [1] G. Rumer, *Nachr. Ges. Wis. Gött., Math-Physik Klasse* (1932) 337.
- [2] L. Pauling, pages 1943-1983 in *Organic Chemistry*, vol. II, ed. H. Gilman, John Wiley & Sons, NY, 1938.
- [3] L. Pauling, *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, NY, 1939.
- [4] E. Hückel, *Zeits Phys.* **60** (1930) 423, **72** (1932) 310, & **76** (1932) 628.
- [5] C. C. J. Roothaan, *Rev. Mod. Phys.* **23** (1951) 69.
- [6] G. G. Hall, *Proc. Roy. Soc. A* **205** (1951) 541.
- [7] M. J. S. Dewar & H. C. Longuet-Higgins, *Proc. Roy. Soc. A* **214** (1952) 482.
- [8] C. A. Coulson & G. S. Rushbrooke, *Proc. Camb. Phil. Soc.* **36** (1940) 193.
H. C. Longuet-Higgins, *J. Chem. Phys.* **18** (1959) 265.
M. J. S. Dewar, *J. Am. Chem. Soc.* **74** (1952) 3341, 3345, 3350, 3353, 3355, & 3357.
- [9] R. B. Woodward & R. Hoffmann, *J. Am. Chem. Soc.* **87** (1965) 395 & 2511.

- H. C. Longuet-Higgins & E. W. Abrahamson, *J. Am. Chem. Soc.* **87** (1965)2045.
- [10] R. B. Woodward & R. Hoffmann, *The Conservation of Orbital Symmetry*, Springer-Chemie, Berlin, 1970.
- [11] D. J. Klein & N. Trinajstić (eds.), *Valence-Bond Theory and Chemical Structure*, Elsevier, Amsterdam, 1990.
- [12] N. Bohr, e.g., in chap. 1, vol. III of *Atomic Physics and Human Knowledge*, Oxbow Press, Woodbridge, CT, 1963.
- [13] Plato, in chap. XXV of *The Republic* [translation F. C. Cornford], Oxford University Press, Oxford, 1941.
- [14] F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA, 1969.
- [15] N. Trinajstić, *Chemical Graph Theory*, CRC Pub., Boca Raton, FL, 1992.
- [16] N. S. Ham, *J. Chem. Phys.* **29** (1958) 1229.
- [17] N. S. Ham & K. Ruedenberg, *J. Chem. Phys.* **29** (1958) 1215.
- [18] S. Ramasesha & Z. G. Soos, *Intl. J. Quantum Chem.* **25** (1984) 1003.
Z. G. Soos & S. Ramasesha, pages 81-110 in ref. [11].
S. A. Alexander & T. G. Schmalz, *J. Am. Chem. Soc.* **109** (1987) 109.
- [19] G. W. Wheland, *Resonance in Organic Chemistry*, J. Wiley & Sons, NY, 1955.
- [20] D. J. Klein, *Topics Curr. Chem.* **153** (1990) 57.
- [21] P.-O. Löwdin, *J. Chem. Phys.* **18** (1950) 365.
P.-O. Löwdin, *Adv. Quantum Chem.* **5** (1970) 185.
- [22] L. Pauling & Wheland, *J. Chem. Phys.* **1** (1933) 362.
- [23] D. J. Klein & N. Trinajstić, *Pure & Appl. Chem.* **61** (1989) 2107.
- [24] W. T. Simpson, *J. Am. Chem. Soc.* **75** (1953) 597.
- [25] W. C. Herndon, *J. Am. Chem. Soc.* **95** (1973) 2404.
W. C. Herndon, *Thermochim. Acta* **8** (1974) 225.
- [26] L. J. Schaad & B. A. Hess, Jr., *Pure & Appl. Chem.* **54** (1982) 1097.
- [27] E. Clar, *The Aromatic Sextet* (John Wiley & Sons, New York, 1972).
- [28] D. J. Klein, *J. Chem. Ed.* **69** (1992) 691.
- [29] M. Randić, *Tetrahedron* **33** (1977) 1905.
M. Randić, *J. Am. Chem. Soc.* **99** (1977) 444.
- [30] J. A. Pople, *Trans. Faraday Soc.* **49** (1953) 1375.
R. Pariser & R. G. Parr, *J. Chem. Phys.* **21** (1953) 466 & 767.
J. Koutecky, *Chem. Phys. Lett.* **1** (1967) 249.
- [31] J. Hubbard, *Proc. Roy. Soc. A* **272** (1963) 237 & **276** (1964) 238.
- [32] H. C. Longuet-Higgins & L. Salem, *Proc. Roy. Soc. A* **257** (1960) 445.
J. N. Murrell & L. Salem, *J. Chem. Phys.* **34** (1961) 1914.
- [33] R. D. Poshusta, T. G. Schmalz & D. J. Klein, *Mol. Phys.* **66** (1989) 317.
- [34] I. Fischer-Hjalmers, *J. Chem. Phys.* **42** (1965) 1962.
- [35] K. F. Freed, *J. Chem. Phys.* **60** (1974) 1765.
K. F. Freed, *Acc. Chem. Res.* **16** (1983) 137.

- [36] P. W. Anderson, *Solid St. Phys.* **14** (1963) 99.
- [37] L. N. Buleavski, *Zh. Eksp. & Teor. Fiz.* **51** (1966) 230.
W. A. Seitz & D. J. Klein, *Phys. Rev. B* **8** (1973) 2236.
E. N. Economou & C. T. White, *Phys. Rev. Lett.* **38** (1977) 289.
J.-P. Malrieu & D. Maynau, *J. Am. Chem. Soc.* **104** (1982) 3021.
D. Maynau, M. A. Garcia-Bach, & J. P. Malrieu, *J. Physique* **47** (1986) 207.
J.-P. Malrieu, pages 135-176 in ref. [11].
- [38] T. P. Živković, *Theor. Chim. Acta* **61** (1982) 363.
T. P. Živković, *Croatica Chemica Acta* **56** (1983) 29 & 525.
T. P. Živković, pages 437-467 in ref [11].
- [39] H. E. Zimmerman, *Quantum Mechanics for Organic Chemists*, Academic Press, NY, 1975.
- [40] C. A. Coulson, B. O'Leary, & R. B. Mallion, *Hückel Theory for Organic Chemists*, Academic Press, London, 1978.
- [41] E. Heilbronner, *Tetrahedron Lett.* (1964) 1923.
S. F. Mason, *Nature* **205** (1965) 495.
H. E. Zimmerman, *J. Am. Chem. Soc.* **88** (1966) 1565 & 1566.
H. E. Zimmerman, *Science* **153** (1966) 837.
- [42] H. Fischer & J. N. Murrell, *Theor. Chim. Acta* **1** (1963) 463.
- [43] W. T. A. M. v.d. Lugt & L. J. Oosterhoff, *Chem. Comm.* **1968**, 1235.
J. J. C. Mulder & L. J. Oosterhoff, *Chem. Comm.* **1970**, 305 & 307.
W. J. van der Hart, J. J. C. Mulder, & L. J. Oosterhoff, *J. Am. Chem. Soc.* **92** (1972) 5724.
- [44] N. D. Epiotis, *Unified Valence-Bond Theory*, Springer-Verlag, Berlin, 1983.
N. D. Epiotis, pages 377-412 in ref. [11].
- [45] D. J. Klein & N. Trinajstić, *J. Am. Chem. Soc.* **106** (1984) 8050.
- [46] W. T. Borden & E. R. Davidson, *Acc. Chem Res.* **14** (1981) 16.
- [47] D. J. Klein & S. A. Alexander, pages 404-419 in *Graph and Topology in Chemistry*, R. B. King & D. H. Rouvray (eds), Elsevier, Amsterdam, 1987.
- [48] W. T. Borden, *J. Am. Chem. Soc.* **97** (1975) 5968.
W. T. Borden & E. R. Davidson, *J. Am. Chem. Soc.* **99** (1977) 4587.
W. T. Borden, E. R. Davidson, & P. Hart, *J. Am. Chem. Soc.* **100** (1978) 388.
W. T. Borden, E. R. Davidson, & D. Feller, *J. Am. Chem. Soc.* **103** (1981) 5725.
- [49] E. H. Lieb & D. C. Mattis, *J. Math. Phys.* **3** (1962) 749.
D. J. Klein, *J. Chem. Phys.* **77** (1982) 3098.
- [50] E. H. Lieb, *Phys. Rev. Lett.* **62** (1989) 1201.
- [51] S. A. Alexander & D. J. Klein, *J. Am. Chem. Soc.* **110** (1987) 3401.
H. Iwamura & A. Izuoka, *J. Chem. Soc. Japan* **1987**, 595.

- [52] A. A. Ovchinnikov, *Theor. Chim. Acta* **47** (1978) 297.
- [53] L. Pauling, *J. Chem. Phys.* **1** (1933) 280.
- [54] R. E. Peierls, page 108 of *Quantum Theory of Solids*, Clarendon, Oxford, 1955.
- [55] H. C. Longuet-Higgins & L. Salem, *Proc. Roy. Soc. (London) A* **251** (1959) 171.
- [56] W. P. Su, J. R. Schrieffer, & A. J. Heeger, *Phys. Rev. B* **22** (1980) 2099.
M. J. Rice, *Phys. Lett. A* **71** (1979) 152.
- [57] I. Bozović, *Mol. Cryst. & Liq. Cryst.* **119** (1986) 475.
- [58] D. J. Klein, *Intl. J. Quantum Chem. S* **13** (1979) 293.
W. A. Seitz & T. G. Schmalz, pages 525-552 in ref. [11].
- [59] M. A. García-Bach, R. Valenti, S. A. Alexander, & D. J. Klein, *Croatica Chemica Acta* **64** (1991) 415.
- [60] D. J. Klein, T. G. Schmalz, W. A. Seitz, & G. E. Hite, *Intl. J. Quantum Chem. S* **19** (1986) 707.
- [61] N. E. Bonesteel, *Phys. Rev. B* **40** (1989) 8954.
D. J. Klein, T. P. Živković, & R. Valenti, *Phys. Rev. B* **43** (1991) 723.
- [62] W. A. Seitz, D. J. Klein, T. G. Schmalz, & M. A. García-Bach, *Chem. Phys. Lett.* **115** (1985) 139.
D. J. Klein, *Chem. Phys. Lett.* **217** (1994) 261.
- [63] W. C. Herndon, *Tetrahedron* **29** (1973) 3.

The Use of the Electrostatic Potential for Analysis and Prediction of Intermolecular Interactions

Tore Brinck

Department of Chemistry, Physical Chemistry, Royal Institute of Technology,
SE-100 44 Stockholm, Sweden

1. INTRODUCTION

The molecular electrostatic potential has been used extensively during the last three decades for the analysis of molecular interactions, including chemical reactions, hydrogen bonding, solvation processes and biomolecular recognition interactions [1-7]. In most of these studies, the objective has been to obtain a qualitative indication of molecular interaction tendencies, either by focusing on the extrema of the electrostatic potential, its most negative and positive values, or by analyzing the overall pattern in the potential. However, developments in recent years have provided new means for using the potential in quantitative analyses of intermolecular interactions. In this chapter we will review some of these new developments. We will also present examples of the use of these methods in studies of some specific chemical applications, related to hydrogen bonding, protonation, substituent effects and solvation.

2. METHODOLOGICAL BACKGROUND

2.1 Definition and Physical Significance

The electrostatic potential $V(\mathbf{r})$ at a point \mathbf{r} in the space surrounding a molecule is defined rigorously by

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \quad (1)$$

where Z_A is the charge on nucleus A, located at \mathbf{R}_A , and $\rho(\mathbf{r})$ is the electronic density function of the molecule. Thus, the first term is the contribution from the nuclei to $V(\mathbf{r})$ and the second term is the contribution from the electrons. Depending upon which term is dominant at \mathbf{r} , the potential has a positive or negative value at this point.

Since $V(\mathbf{r})$ is rigorously defined in terms of the nuclear and electronic charge distribution, it is a real physical property. Both experimentally and theoretically determined electrostatic potentials are frequently reported in the literature. Predominantly, X-ray crystallography is used for experimental determination of $V(\mathbf{r})$. However, at least for smaller molecules, the most common approach is to compute $V(\mathbf{r})$ from a theoretically determined charge distribution. Ab initio, semi-empirical and density functional theory (DFT) methods are routinely used for this purpose.

$QV(\mathbf{r})$ gives the interaction energy between a point charge of magnitude Q located at \mathbf{r} and the static charge distribution of the molecule. Furthermore, $QV(\mathbf{r})$ is the first order energy term in a perturbation theory treatment of the interaction between the charge and the molecule. It should be noted that $QV(\mathbf{r})$ does not account for contributions to the total interaction energy that arise from perturbations to the charge distribution induced by the interacting charge, e.g. polarization of the electron density and changes in the nuclear geometry. However, such contributions are most important within the van der Waals volume of the molecule and their importance decrease rapidly with increasing distance from the molecule. $V(\mathbf{r})$ can therefore be considered as equal to the initial potential an approaching charge or charge distribution will feel from the molecule.

2.2 Spatial Minima in the Electrostatic Potential

Figure 1 shows the HF/6-31G* computed $V(\mathbf{r})$ of acetone in the plane defined by the four non-hydrogen atoms of the molecule. We will refer to this plane as the molecular plane. As can be seen from the figure, there are both positive and negative regions in $V(\mathbf{r})$ for acetone. This is typical for a neutral molecule, while $V(\mathbf{r})$ for a neutral atom is everywhere positive. There exist two equivalent spatial minima (V_{\min}) in $V(\mathbf{r})$ for acetone. They are both located in the molecular plane and are positioned symmetrically around the oxygen atom. The distance from the oxygen is 1.21 Å and the angle with the O-C bond is 129°. The magnitude of the V_{\min} is -57.5 kcal/mol. The locations of the V_{\min} would correspond to the two possible equilibrium positions of a positive charge interacting with the molecule if the static charge distribution of the molecule was not allowed to relax. The interaction energy for such an interaction would be QV_{\min} where Q is magnitude of the charge. In reality, there will be a significant perturbation of the molecular charge distribution when the charge comes within the van der Waals volume of the molecule and QV_{\min} can therefore only be seen as an upper bound to the total interaction energy. To get an idea of the importance of the charge redistribution effects, we will first consider the interaction between acetone and a proton in the gas phase. As predicted from the

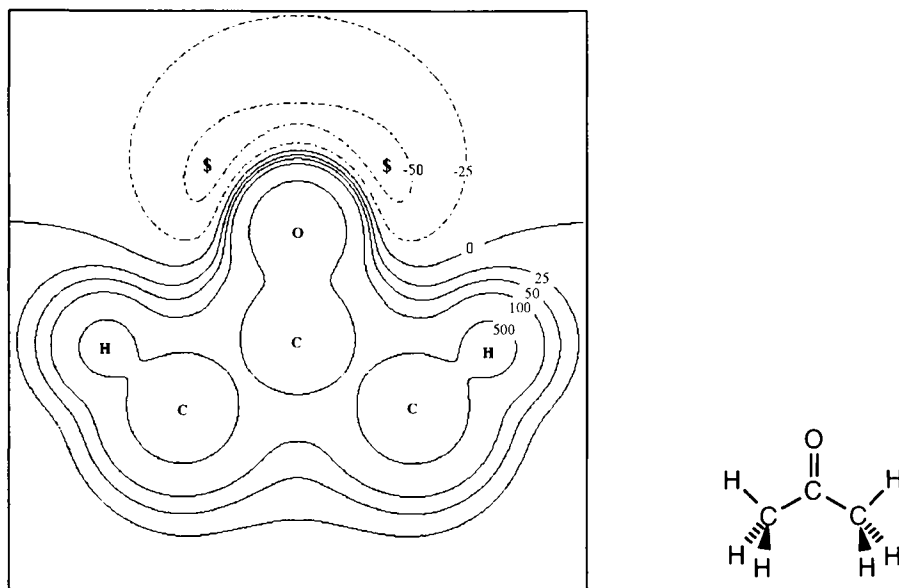
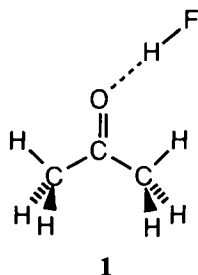


Figure 1. Computed HF/6-31G* $V(\mathbf{r})$ of acetone (in kcal/mol) in the plane containing the heavy atoms. Dashed contours correspond to negative $V(\mathbf{r})$. The positions of the two V_{\min} are depicted by \$. The V_{\min} value is -57.5 kcal/mol.

V_{\min} , ab initio calculations show that the equilibrium position of the proton is in the molecular plane. The O-H bond length is 0.96 Å and the H-O-C angle is 116° at the HF/6-31+G* level. Thus the bond length is 0.27 Å shorter and the bond angle is 13° smaller than predicted from the position of the V_{\min} . Still the V_{\min} prediction is remarkably good, particularly if we consider that the gas phase proton affinity of acetone is 196.7 kcal/mol, more than three times the magnitude of the V_{\min} .

The electrostatic potential and the V_{\min} are not only useful for the prediction of interactions with protons but also with other electrophiles, and in particular for interactions with hydrogen bond donors. For example, the charge distribution of hydrogen fluoride can crudely be approximated by a positive partial charge on hydrogen and an equal but negative charge on fluoride. It can therefore be expected that the hydrogen end of the molecule will be attracted to regions of negative potential. The HF/6-31+G* computed equilibrium geometry of the acetone-hydrogen fluoride complex with the lowest energy is depicted below, 1. The hydrogen fluoride lies in the molecular plane of acetone and the H-O-C angle is 129° in perfect agreement with the V_{\min} -O-C angle. The potential clearly does a better job at predicting the directionality of the much weaker interaction with HF than the strong interaction with the proton. An energy

decomposition analysis at the HF/6-31+G* level using the Morokuma scheme [8, 9] also shows that the electrostatic interaction energy (-14.8 kcal/mol) is the single largest component of the total Hartree-Fock interaction energy (-9.7 kcal/mol). The second largest contribution comes from the exchange repulsion energy (9.9 kcal/mol), which is of opposite sign to the electrostatic energy. The exchange repulsion is the non-classical repulsion between the electrons of the interacting molecules that arises as a consequence of the Pauli principle. The electrostatic energy and the exchange repulsion together constitute the first order energy term in the perturbation treatment of the total Hartree-Fock interaction energy [10]. However, the positive exchange-repulsion energy is largely canceled by the negative energy contributions from polarization and charge transfer, which is the reason that the electrostatic interaction energy is reasonably close to the total interaction energy. It should be noted that these calculations do not include electron correlation effects, which means that dispersion interactions are not accounted for. Calculations at the correlated MP2/6-31+G* level give a total interaction energy of -11.2 kcal/mol. This comparison of the interactions of acetone with a proton and with hydrogen fluoride is one example that the electrostatic potential is significantly better for analysis of non-covalent interaction between preferably polar molecules than for studying stronger interactions that lead to the formation of covalent bonds.



It is easily recognized that the existence of the two V_{\min} around the oxygen in acetone is in agreement with the well established concept that an sp^2 hybridized oxygen possesses two lone pairs. Also the locations of the two V_{\min} correspond to where we would expect to find these lone pairs. This is a general observation that the position and magnitude of V_{\min} can be used to characterize lone pairs in molecules [2, 5]. It can be noted that the existence of the lone pairs is not as easily deduced from an analysis of the electron density. There are, for example, no maxima in the electron density associated with lone pair regions. Other molecular regions that generally exhibit minima in the potential are π -regions, including multiple bonds and aromatic groups, and bent bonds. These

Table 1
Computed HF/6-31G* V_{\min} for some organic molecules

	V_{\min}^a	Type	Location
CH ₃ NH ₂	-84.3 (1)	Lone pair	Outside N, along C-N bond.
CH ₃ OH	-60.6 (2)	Lone pair	Outside O, above and below plane defined by C-O-H.
CH ₃ F	-30.6 (3)	Lone pair	Outside F, 124° angle with F-C bond.
C ₂ H ₄	-24.7 (3)	π -region	Outside C-C bond, above and below molecular plane.
C ₆ H ₆	-24.8 (12)	π -region	Inside C-C bond, above and below molecular plane.
Cyclopropane	-20.7 (3)	Bent bond	Outside C-C bond, in molecular plane.
C ₂ H ₆	-2.8 (2)	CH ₃ group	Outside C, along C-C bond.

^aSpatial minima in $V(\mathbf{r})$, i.e. (3,+3) critical points in $V(\mathbf{r})$. Numbers in parentheses refer to the number of equivalent minima.

are also types of regions that are known to be susceptible to electrophilic attack, which further confirms that minima in the potential are useful for characterizing sites for electrophilic attack. It should be noted that most neutral molecules have regions of negative potential and thereby also associated V_{\min} . However, for nonpolar molecules held together by σ -bonds, e.g. H₂, Cl₂ and alkanes, the magnitudes of the V_{\min} are often small, indicating low degrees of electrophilic susceptibility. It should therefore be stressed that the existence of minima should not be used by itself to identify sites for electrophilic attack but that also the magnitudes of the V_{\min} need to be considered. In Table 1 are listed calculated V_{\min} for some typical organic molecules.

2.3 Surface Electrostatic Potential

As can be seen from the potential map of acetone, sites for nucleophilic attack are not as easily identified from the electrostatic potential as are sites for electrophilic attack. Except for the negative region associated with the oxygen lone pairs, the electrostatic potential of acetone is everywhere positive. The regions of most positive potential are associated with the nuclei and the magnitude of the potential in these regions reflects the magnitude of the nuclear charges and, therefore, can not be assumed to indicate susceptibility toward nucleophiles. It has also been shown that maxima in the potential cannot exist at other positions than at the nuclei [11], where the potential becomes undefined if the nuclear charge distribution is represented by point charges as in eq. 1. Thus, there exists no criterion for analyzing nucleophilic processes that

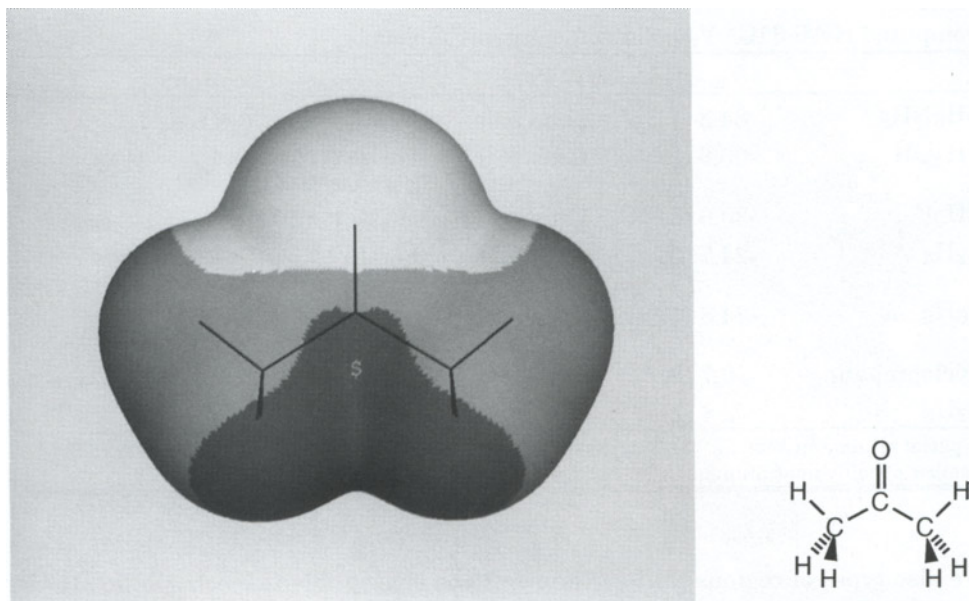


Figure 2. Computed HF/6-31G* $V(\mathbf{r})$ (in kcal/mol) on the 0.001 electron/bohr³ molecular surface of acetone; dark gray > 15.0, 15.0 > light gray > 0.0, white < 0.0. The $V_{S,\max}$ is designated by a \$; its value is 22.3 kcal/mol.

corresponds directly to V_{\min} for electrophilic processes. However, Politzer and co-workers have developed a very useful methodology for analysis of susceptibility towards nucleophilic attack [12-15]. In this method the potential is calculated and analyzed on a molecular surface that is defined by a constant contour of the electron density. To define the molecular surface in terms of the electron density has several advantages. First it can be noted that such a surface reflects features that are unique to a molecule, e.g. the formation of chemical bonds and lone pairs. Secondly, since the magnitude of the exchange-repulsion energy depends on the degree of electron overlap between the interacting molecules [16, 17], it seems logical to define the molecular boundary from the electron density. Bader and co-workers have also shown that the 0.001 and 0.002 a.u. contours of the electron density give molecular dimensions in agreement with intermolecular equilibrium distances observed in liquids and gases of nonpolar molecules [18, 19]. Polar molecules were found to approach each other more closely, which indicates that favorable interactions, such as electrostatic and polarization interactions, reduce their intermolecular distances [19].

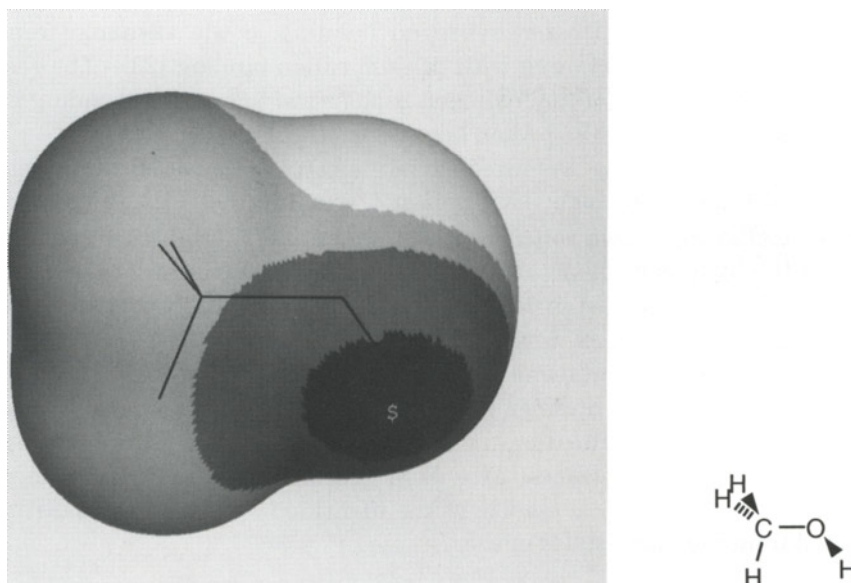


Figure 3. Computed HF/6-31G* $V(\mathbf{r})$ (in kcal/mol) on the 0.001 electron/bohr³ molecular surface of methanol; black > 40.0, 40.0 > dark gray > 15.0, 15.0 > light gray > 0.0, white < 0.0. The largest $V_{S,\max}$ is designated by a \$; its value is 47.9 kcal/mol.

Figure 2 shows the electrostatic potential computed on the 0.001 a.u. electron density contour of acetone. The depicted surface electrostatic potential has its largest maximum ($V_{S,\max}$) above the carbonyl carbon indicating this to be the site of the molecule that is most susceptible toward nucleophilic attack. This is in agreement with experimental observations. Ketones are, for example, known to undergo nucleophilic additions to the carbonyl carbon [20]. A number of studies have also shown that $V_{S,\max}$ allows the identification of sites for nucleophilic attack [13, 15, 21, 22]. Furthermore, the magnitudes of the maxima can serve as an indication of the relative reactivities at these sites. It can be noted that there are also two equivalent minima in the surface potential ($V_{S,\min}$), which can be interpreted as sites for electrophilic attack. In contrast to the V_{\min} for the same molecule, the $V_{S,\min}$ are located close to the top of the oxygen; the $V_{S,\min}$ -O-C angle is 170°. The $V_{S,\min}$ is therefore not as successful as the V_{\min} for predicting the directionality of protonation and hydrogen bonding of acetone. However, the lithium cation (Li^+), for example, binds in a linear configuration with the C-O bond. In relation to this, it can be noted that the exchange-repulsion plays an important role in the interaction with Li^+ [23], while it does

not contribute to the protonation processes, since the proton does not have any associated electrons. Also for hydrogen bonding is the exchange repulsion relatively small when compared with lithium cation binding [23]. The reason is that the electron density of the hydrogen is polarized towards its bonding partner (the heteroatom), leaving the proton largely unshielded.

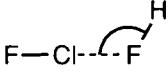
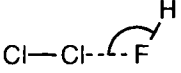
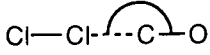
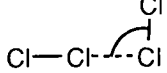
Figure 3 shows the computed surface electrostatic potential of methanol. A $V_{S,max}$ of large magnitude, 47.9 kcal/mol, is associated with the hydrogen of the hydroxyl group. Two much weaker maxima, 12.2 and 10.4 kcal/mol, are found on the hydrogens of the methyl group. In comparison with the $V_{S,max}$ on acetone (22.3 kcal/mol), the hydroxyl $V_{S,max}$ is significantly larger in magnitude. These results are consistent with methanol being a strong hydrogen bond donor, much stronger than acetone, and with the bonds being formed with the proton of the hydroxyl group. A number of studies have shown that surface maxima provide means for identification and ranking of hydrogen-bond-donating sites [15, 21, 24, 25]. Furthermore, as will be discussed in a later section of this chapter, the hydrogen bond acidity of the identified sites can be quantitatively predicted from the magnitudes of the $V_{S,max}$.

2.4 Geometries of Weak Complexes

In the previous sections, we have shown that the electrostatic potential can be used for characterization of sites in molecules that are susceptible towards interactions with electrophiles and nucleophiles. We have emphasized that the potential is especially suited for analysis of hydrogen bonding interactions. The importance of electrostatics for determining the directionality of hydrogen bond interactions is not surprising since the electrostatic part of the interaction energy has been shown to be dominating for many hydrogen-bonded complexes [2, 9, 26]. However, due to the large anisotropies in the electrostatic potentials of molecules, electrostatics often also determines the directionalities of interactions in which it makes only a minor contribution to the total interaction energy [27-31]. In Table 2 are listed a number of molecular complexes which have been denoted as van der Waals or charge-transfer complexes. The depicted structures have for all complexes except the chlorine dimer been determined by spectroscopical gas phase measurements [32-35]. The chlorine dimer structure corresponds to the geometry of the nearest neighbors in the chlorine crystal [36, 37]. Although the exact gas phase structure has not been determined, available spectroscopic data suggest that the depicted L-structure is the most stable form of the dimer [38]. This has also been confirmed by high-level ab initio calculations [39, 40]. The geometries of these types of complexes have often been rationalized in terms of HOMO-LUMO interactions. However, the canonical

Table 2

Intermolecular angles (in deg.) in weak complexes predicted from V_{\min} , ab initio computations and experiment

				
	2	3	4	5
V_{\min}	121 ^a	121 ^a	180 ^a	101 ^a
ab initio	121 ^b	131 ^b	180 ^c	90 ^c
exp.	125±3 ^d	125±3 ^e	180 ^f	104 ^g

^a V_{\min} angle of the donor molecule calculated at the HF/6-31G* level [40].
^bAngle determined by full geometry optimization at the MP2/6-311++G(2df,2p) level [40].
^cAngle determined by full geometry optimization at the MP2/6-311+G(2d) level [40].
^dExperimentally determined angle from ref. [32].
^eExperimentally determined angle from ref. [33].
^fExperimentally determined angle from ref. [34, 35].
^gAngle between the nearest neighbors in the Cl₂ crystal [36, 37].

Hartree-Fock orbitals fail to predict the correct geometries of complexes **1-3**. For both HF and CO the HOMO orbital is an almost pure p-orbital, which suggests that the structures **1-3** should be L-shaped with an bond angles close to 90°. Several studies have instead shown that the geometrical configurations of these types of weak complexes often are determined by electrostatics [27-31]. We have found that the orientation of the molecules in complexes **1-4** can be predicted from the computed electrostatic potentials of the isolated molecules by aligning the most positive $V_{S,\max}$ with the most negative V_{\min} [40]. The chlorine atoms in ClF and Cl₂ have strongly positive end regions, which explains why these molecules bind with a linear arrangement toward the electron donor. HF has a V_{\min} at a 121 degree angle with the HF bond, CO has it most negative V_{\min} in the end of the carbon atom, and Cl₂ has its rather weak minimum at a 90° angle with the Cl-Cl bond, which explains the orientations of the donor atoms in the above complexes. As can be seen from the table, the intermolecular angles predicted from the V_{\min} - $V_{S,\max}$ alignment agrees very well with the experimentally determined angles. High level ab initio calculations with large basis set also give good predictions, but at a significantly higher computational cost. Hartree-Fock calculations with moderate basis sets are less successful, mainly due to the basis set superposition error [40]. It should also be noted that the correct shapes of these complexes cannot be predicted from the lowest non-zero electrical moments of the interacting molecules.

2.5 Polarization Corrections to the Interaction Energy

Although electrostatics often play an important role for determining the directionalities of intermolecular interactions, quantitative analyses and predictions of interaction energies generally require consideration of other energy contributions as well. The interaction energy between a molecule and a classical point charge Q located at a position \mathbf{r} can be defined, using perturbation theory, as a power series in terms of Q :

$$E(Q, \mathbf{r}) = QV(\mathbf{r}) + Q^2P(\mathbf{r}) + Q^3P'(\mathbf{r}) + Q^4P''(\mathbf{r}) + \dots \quad (2)$$

The electrostatic potential $V(\mathbf{r})$ is the first order contribution to the interaction energy. The second order contribution $P(\mathbf{r})$, which is commonly referred to as a polarization correction to the electrostatic potential, is defined within the uncoupled Hartree-Fock perturbation theory by [41]

$$P(\mathbf{r}) = \sum_i^{\text{occ}} \sum_a^{\text{vir}} \frac{1}{\epsilon_i - \epsilon_a} \left[\sum_{\mu} \sum_{\nu} c_{\mu i} c_{\nu a} \int \frac{\chi_{\mu}(\mathbf{r}') \chi_{\nu}(\mathbf{r}') d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \right]^2 \quad (3)$$

The ϵ_i are the orbital energies and the $c_{\mu i}$ are the molecular orbital expansion coefficients in terms of the atomic orbital basis set χ_{μ} . Contributions from terms greater than second order are generally small and can for most chemical applications be neglected [41, 42]. Only a limited number of applications of $P(\mathbf{r})$ for analysis of intermolecular interactions have appeared in the literature. The most common approach of analysis has been to calculate a total interaction index, a "polarization-corrected electrostatic potential", defined by

$$E_p(\mathbf{r}) = QV(\mathbf{r}) + Q^2P(\mathbf{r}) \quad (4)$$

where Q is set to 1 a.u. or -1 a.u. depending upon whether nucleophilic or electrophilic processes are studied. The general consensus from the studies employing $E_p(\mathbf{r})$ is that it is better suited than $V(\mathbf{r})$ for the analysis of interactions with strong nucleophiles, or electrophiles, that lead to the formation of covalent bonds. Franci found $E_p(\mathbf{r})$ well suited for studying nucleophilic attack in vinylic systems [41]. Dive and Dehareng have demonstrated the use of $E_p(\mathbf{r})$ for characterization of sites susceptible to electrophilic and nucleophilic attack in aromatic systems [43]. It has also been shown that the correct ordering of the gas phase basicities of ammonia and its methyl derivatives cannot be predicted from the electrostatic potential unless a polarization correction is added [2]. A

major disadvantage with $E_p(\mathbf{r})$ is that its calculation requires considerably more computer time than a $V(\mathbf{r})$ calculation. This essentially prohibits the calculation of full potential maps for larger molecules. However, according to our experience, the anisotropy of $P(\mathbf{r})$ is generally much smaller than for $V(\mathbf{r})$, and the interaction sites predicted from $V(\mathbf{r})$ therefore often correspond very well to those predicted from $E_p(\mathbf{r})$. Despite this, the inclusion of polarization can be very important for the relative ranking of different sites, particularly when the interaction tendencies of different molecules are considered. A useful approach is therefore to compute $P(\mathbf{r})$ only at the possible interaction sites that have been identified from the potential. In particular, for investigating the interaction tendencies towards electrophiles, we compute $P(\mathbf{r})$ at the positions of the V_{\min} . This quantity will hereafter be referred to as $P_{V_{\min}}$. Another problem with the use of $E_p(\mathbf{r})$ is the assumption that the polarizing charge is of unitary magnitude. Whereas this may be appropriate for interactions with protons or small ions, it is harder to estimate what charge to use for other interactions. The approach we usually take for quantitative analysis of interaction tendencies towards electrophiles is not to assume an apparent charge of the electrophile, but rather to treat V_{\min} and $P_{V_{\min}}$ as separate quantities and let their relative contributions be determined from multilinear correlations with known interaction energies.

2.6 Charge Transfer and the Average Local Ionization Energy

It can be noted that eq. 3 only accounts for effects due to polarization of the charge within the molecule, or to be more exact, it only allows for redistribution of electrons within the subspace spanned by the basis functions. However, the close interaction between two chemical species is often accompanied by a flow of electrons between them, which is normally referred to as a charge transfer interaction. In protonation processes, for example, there is usually a transfer of electrons towards the proton that leads to a build up of electron density around the proton and the formation of a covalent bond. Within the molecular orbital picture, charge transfer corresponds to transfer of electrons from occupied orbitals in one species to unoccupied orbitals in the other. It should be noted that the differentiation between polarization and charge transfer is strongly dependent on the basis set used for describing the electronic charge distribution. In the limit of an infinite basis set, it is no longer possible to distinguish between the two types of interactions [9]. However, for practical purposes it may still be warranted to treat the two terms separately. We have found $P(\mathbf{r})$ useful for prediction of polarization, while the average local ionization energy is very well suited for prediction of the electron donating capacities of molecules. The average local ionization energy $\bar{I}(\mathbf{r})$ is rigorously defined within the framework of the Hartree-Fock theory by [44]

$$\bar{I}(\mathbf{r}) = \sum_i^{\text{occ}} \frac{\rho_i(\mathbf{r})|\epsilon_i|}{\rho(\mathbf{r})} \quad (5)$$

where $\rho_i(\mathbf{r})$ is the electronic density of the i^{th} molecular orbital at the point \mathbf{r} and ϵ_i is the orbital energy. According to Koopmans' theorem, the absolute values of the Hartree-Fock orbital energies are good approximations to the ionization energies, and $\bar{I}(\mathbf{r})$ can therefore be viewed as the average energy needed to ionize an electron from a point \mathbf{r} in the space of an atom or molecule.

A number of studies have shown that $\bar{I}(\mathbf{r})$ calculated on molecular surfaces defined by contours of constant electron density provide an effective tool for analysis of reactivity towards electrophiles [44-49]. The positions on a molecular surface where $\bar{I}(\mathbf{r})$ has its lowest values, the local surface minima ($\bar{I}_{S,\text{min}}$), are viewed as the locations of the least tightly bound electrons, and thus as the sites most likely to interact with an electrophile. However, in contrast to the electrostatic potential, $\bar{I}(\mathbf{r})$ reflects a molecule's ability to undergo charge transfer rather than its electrostatic interaction tendencies. The $\bar{I}_{S,\text{min}}$ are therefore better suited than the V_{min} for analyses of strong interactions that lead to the formation of covalent bonds. For example, it has been shown that the $\bar{I}_{S,\text{min}}$ of aromatic systems can be used to identify and rank the sites most likely to undergo electrophilic attack that leads to electrophilic aromatic substitution [44]. The V_{min} are not as successful in the same type of analysis. On the other hand, V_{min} are much better suited than $\bar{I}_{S,\text{min}}$ for characterization of hydrogen-bond-accepting sites [21].

The average local ionization energy is the topic of another chapter in this book, and the interested reader is referred to that chapter for further information about this property. In this chapter we will focus on how $\bar{I}(\mathbf{r})$ can complement $V(\mathbf{r})$ in quantitative analyses of intermolecular interactions.

2.7 Characters of the Different Interaction Quantities

In Table 3 are listed the V_{min} , $P_{V_{\text{min}}}$ and $\bar{I}_{S,\text{min}}$ together with some experimentally determined interaction indices for a diverse group of organic bases, containing oxygen, sulfur and nitrogen. The table shows the different natures of the three computed quantities. Considering first the V_{min} values, we can see that the V_{min} of the sulfur compounds are of smaller magnitudes than of the nitrogen and oxygen compounds. This is consistent with sulfur being less electronegative than the two first row compounds and that the negative charge on sulfur is less concentrated due to the larger size of the sulfur atom. The more negative potentials associated with the sp^3 nitrogen compounds compared to the

Table 3

Computed HF/6-31G* quantities^a and experimentally determined interaction indices for some oxygen, nitrogen and sulfur bases

	V_{\min}	PV_{\min}	$\bar{I}_{S,\min}$	Π	$-\Delta H_{Ph}^b$	$\Delta\nu_{OH}^c$	$-\Delta H_H^{+d}$
	(kcal/mol)	(kcal/mol)	(eV)	(kcal/mol)	(kcal/mol)	(cm^{-1})	(kcal/mol)
(CH ₃ CH ₂ S) ₂	-31.5	-39.3	10.88	9.0	[3.8]	75	[195.6]
<i>p</i> -Dithiane	-33.0	-40.6	10.43	10.6	[4.0]	121	[198.9]
(CH ₃) ₂ S	-37.9	-37.4	10.18	9.8	4.6	137	200.6
Cl ₃ CCN	-39.1	-47.9	16.56	7.3	[3.5]	23	175.8
(CH ₃ CH ₂) ₂ S	-39.2	-42.1	10.05	7.4	4.6	146	205.0
Tetrahydro- thiophene	-39.4	-41.1	10.04	8.4	4.9	154	204.6
(CH ₃ CH ₂) ₃ P=S	-47.7	-42.7	9.64	13.9	[5.3]	195	[208.2]
ClH ₂ CCN	-48.5	-49.6	15.95	16.9	4.2	48	179.5
<i>p</i> -Dioxane	-53.7	-52.6	15.15	11.0	5.6	126	193.8
(3,5-Cl ₂)- pyridine	-55.1	-72.2	13.45	7.8	[6.5]	200	[214.3]
(CH ₃ CH ₂) ₂ O	-55.8	-57.0	14.86	6.4	6.0	150	200.2
C ₆ H ₅ CHO	-55.8	-53.6	15.16	12.7	[5.2]	65	200.2
CH ₃ CO ₂ CH ₃	-56.6	-51.5	15.06	12.9	4.8	77	197.8
CH ₃ CO ₂ C ₂ H ₅	-57.4	-52.1	15.02	11.2	4.8	83	200.7
CH ₃ COCH ₃	-57.5	-53.0	14.92	14.2	5.1	115	196.7
CH ₃ CN	-57.8	-51.3	15.36	21.1	4.6	75	188.2
HCO ₂ CH ₃	-58.3	-47.0	15.00	22.2	[4.5]	54	188.4
CH ₃ OH	-60.6	-41.9	14.85	12.6	[5.2]	116	181.9
CH ₃ CH ₂ OH	-61.0	-44.2	14.77	11.6	[5.4]	120	188.2
Tetrahydro- furan	-62.0	-52.4	14.90	8.9	6.0	158	198.8
Pyrimidine	-62.8	-69.4	13.14	12.8	[6.8]	213	210.8
(CH ₃) ₂ NCN	-64.4	-54.7	15.01	18.1	5.4	117	205.0
F ₃ CCH ₂ NH ₂	-66.7	-50.0	12.36	13.0	[6.7]	250	202.5
(CH ₃) ₂ NCHO	-67.6	-54.3	14.36	15.6	6.1	150	(211.4)
(CH ₃) ₂ NCOCH ₃	-68.8	-58.3	14.30	13.7	6.8	179	(216.2)
Quinoline	-69.1	-78.0	12.60	11.1	[7.9]	[303]	226.5
(CH ₃ CH ₂) ₃ N	-69.7	-83.7	11.30	3.9	9.1	429	232.3
Pyridine	-70.8	-72.8	12.62	11.6	8.0	286	220.8

Continued

Table 3

Computed HF/6-31G* quantities^a and experimentally determined interaction indices for some oxygen, nitrogen and sulfur bases (continued)

	V_{\min}	$P_{V_{\min}}$	$\bar{I}_{S,\min}$	Π	$-\Delta H_{Ph}^b$	$\Delta\nu_{OH}^c$	$-\Delta H_H^d$
	(kcal/mol)	(kcal/mol)	(eV)	(kcal/mol)	(kcal/mol)	(cm^{-1})	(kcal/mol)
(CH ₃) ₃ N	-71.3	-69.9	11.40	5.1	8.8	[359]	225.1
(4-CH ₃)pyridine	-73.0	-73.8	12.51	11.1	[8.2]	304	225.2
Quinuclidine	-76.6	-75.7	11.30	5.2	9.0	[400]	233.1
(CH ₃ CH ₂) ₂ NH	-76.7	-67.4	11.33	5.7	8.6	398	225.9
(CH ₃) ₂ S=O	-76.8	-57.0	13.33	19.2	6.9	205	211.3
Pyridine	-77.0	-69.0	12.90	18.1	7.9	278	220.3
<i>N</i> -oxide							
Cyclopropyl- amine	-77.5	-55.6	11.73	8.9	[8.2]	310	215.2
(CH ₃) ₂ NH	-78.1	-59.0	11.41	7.4	8.6	[353]	220.6
(CH ₃) ₃ P=O	-78.6	-57.2	12.82	17.1	[7.5]	266	217.1
(1-CH ₃)- imidazole	-82.1	-68.4	12.41	16.9	[8.4]	313	228.9
CH ₃ CH ₂ NH ₂	-83.4	-51.5	11.50	9.1	8.6	[348]	217.0
CH ₃ (CH ₂) ₃ NH ₂	-83.8	-51.9	11.50	7.2	[8.8]	354	217.9
CH ₃ NH ₂	-84.4	-47.5	11.52	11.2	8.6	344	214.1
NH ₃	-87.8	-36.8	11.83	18.9	7.8	275	204.0

^aComputed quantities are from ref. [50].

^bExperimentally determined enthalpies for 1:1 phenol-base complexation in apolar solvents [51]. Values in brackets are predicted from eq. 10.

^cExperimentally determined OH frequency shifts for methanol-base complexes in carbon tetrachloride. These values were obtained by Berthelot and co-workers [52-54]. Values in brackets are predicted from eq. 11.

^dExperimentally determined gas phase proton affinities [55]. Values in parentheses were not included in the correlation with computed quantities. Values in brackets are predicted from eq. 12.

oxygen compounds, despite the larger electronegativity of oxygen, is partly a consequence of that the negative charge on the nitrogen in these compounds is concentrated to only one lone pair. The $P_{V_{\min}}$ values mainly reflect the characters of the groups that are bonded to the heteroatom. Large and easily polarized groups that are bonded directly to the heteroatom increase the magnitude of $P_{V_{\min}}$. This is most clearly seen for ammonia and its methyl derivatives where $P_{V_{\min}}$ changes dramatically with the substitution of hydrogens

for methyl groups. As has already been mentioned, the polarization correction is needed to correctly predict the relative basicities of the amines. The $\bar{I}_{S,\min}$ values, on the other hand, reflect more strongly the chemical nature of the heteroatom. The sulfur containing compounds have much lower $\bar{I}_{S,\min}$ values than those containing oxygen, consistent with sulfur being softer and more easily ionized than oxygen. For the amines, the $\bar{I}_{S,\min}$ values decrease with increased substitution in the same manner as the $P_{V,\min}$ values. However, the changes in the $\bar{I}_{S,\min}$ values are much smaller.

3. ANALYSIS OF SITE-SPECIFIC INTERACTIONS

3.1 Hydrogen Bonding

A number of theoretical studies have shown that hydrogen bonds between neutral molecules are largely electrostatic in nature [2, 9, 23, 26, 27, 56]. The importance of electrostatics in hydrogen bonding is understandable if one considers the electrostatic potentials of typical hydrogen bond donors and acceptors. The electrostatic potential of a good hydrogen bond donor is always strongly positive on the outside of the donating hydrogen. This is a consequence of that the electron density of the hydrogen is polarized towards the electronegative atom it is bonded to, typically a fluorine, oxygen or nitrogen, leaving the proton largely unshielded on the side opposite to the bond. In a similar way, good acceptor sites, which generally are associated with the lone pairs of nitrogens or oxygens, are characterized by regions of strongly negative potential. It should be noted that the acceptor strength is not only determined by the ability of the hydrogen-bond-accepting atom to withdraw charge from its environment but also on the number of lone pairs this charge is distributed on. This explains why the hydrogen-bond-accepting ability decreases in the order $N(sp^3) > O \gg F$ despite that the electronegativity decreases in the opposite order. As can be seen from Table 1, this charge concentration effect is fully reflected in the electrostatic potential minima of the lone pair regions.

The electrostatic potential can also be used for quantitative predictions of the strengths of hydrogen bond interactions. This was first demonstrated in a study by Kollman et al [57]. They found for a group of twelve small (1-2 non-hydrogen atoms) hydrogen bond donors a linear relationship between the electrostatic potential at a distance of 2 Å from the acidic proton and the complexation energy with ammonia. The linear correlation coefficient was 0.94. A good relationship was also observed between the negative electrostatic potential at a fixed distance outside the heteroatom and the complexation energy with hydrogen fluoride for a similar set of hydrogen bond acceptors. In this case the correlation coefficient was even better, 0.985. The correlations are

surprisingly good, especially since both the donor and acceptor set contain molecules with first row atoms (C, O, N) as well as second row atoms (S, P, Cl). However, it should be noted that both the complexation energies and the electrostatic potentials were computed at the Hartree-Fock level and with the 4-31G basis set, which is known to overestimate electrostatic effects. Based on that both these relationships have intercepts close to zero, Kollman later suggested that a product of the positive potential associated with the donor and the negative potential of the acceptor should be able to estimate the association energy of an arbitrary donor-acceptor complex [27]:

$$\Delta E = \text{constant } V_{\text{donor}} V_{\text{acceptor}} \quad (6)$$

This equation was found to give good estimates for a number of complexes, including π -complexes, but failed for complexes where the donor is a CH group.

A number of empirical parameters and models for the analysis and prediction of hydrogen bond interactions have been developed over the years. In particular, Kamlet, Taft, Abboud and Abraham and their coworkers have made significant contributions to this field by the development of linear solvation energy relationships (LSER) [58, 59]. They have by means of a limited number of empirical parameters been able to correlate more than 250 biological, chemical and physical properties involving solute-solvent interactions for a large number of compounds [60, 61]. The most important parameters within the LSER approach are the hydrogen-bond-donating and accepting parameters, α and β , respectively. These were originally defined as solvent parameters and determined from the solvent effects on the spectroscopic properties of some reference solutes [59]. However, the original solvent parameters have now largely been substituted for the more recently developed solute parameters, α_2^H and β_2^H , which have been determined from the free energies of 1:1 acid-base complexation in carbon tetrachloride [62-65]. For more than 1300 combinations of hydrogen bond donors and acceptors, the following relationship has been shown to hold to a very high accuracy [66]:

$$\log K_{\text{HB}} = 7.354 \alpha_2^H \beta_2^H - 1.094 \quad (7)$$

N = 1312, R = 0.996, SD = 0.093, F = 14788

K_{HB} is the equilibrium constant in carbon tetrachloride for the 1:1 complexation, and α_2^H and β_2^H are the parameters for the donor and acceptor, respectively. Available data indicates that eq. 7 with slightly different coefficients also can describe complexation in 1,1,1-trichloroethane and the gas phase [67]. The

resemblance with the eq. 6 is noticeable. It could also be argued that eq. 7 supports the notion that hydrogen bonding is largely electrostatic in nature. If hydrogen bonding were more covalent in its character, it is less likely that it would be possible to reproduce the total interaction energy by just a product of the donating and accepting capabilities of the monomers. Murray and Politzer have also shown that the LSER hydrogen bond parameters can be related to the molecular electrostatic potential. In their original work, they found for families of molecules (azines, ethers, primary amines and molecules containing double bonded oxygens), taken separately, good correlations between the heteroatom V_{\min} computed at the HF/STO5G level and the solvent hydrogen-bond-accepting parameter β [68]; the correlation coefficients range from 0.94 to 0.98. Relationships of similar quality were also demonstrated to exist between $V_{S,\max}$ and the solvent hydrogen-bond-donating parameter, α , for alcohols and molecules with alkyl groups as donors [15]. It was later shown that V_{\min} and $V_{s,\max}$ correlate equally well with the solute parameters, α_2^H and β_2^H , respectively [24]. A subsequent study showed that the quality of the correlations with β_2^H degrades significantly when V_{\min} is substituted for $V_{S,\min}$ [21].

Kenny followed the approach of Murray and Politzer and investigated correlations between the electrostatic potential and hydrogen bond basicity for a set of 23 heterocycles with nitrogen as the acceptor [69]. A very good linear relationship ($R = 0.981$) was found between the HF/6-31G* computed V_{\min} and $\log K_{HB}$ for formation of 1:1 complexes with 4-nitrophenol in 1,1,1-trichloroethane. It was pointed out that the V_{\min} versus $\log K_{HB}$ relationship is significantly better than the correlations between aqueous basicity and hydrogen bond basicity (pK_a versus $\log K$) for the same types of system. The predictive capability of the relationship was further demonstrated for a set of five heterocycles in which all molecules contains two or more non-equivalent nitrogen donors. Four of the predictions are within 0.30 units of the experimental values, while the fifth, tetrazole, is overestimated by 0.51 units.

Kenny also studied how the correlation between the potential and the basicity is affected when the potential is computed at a fixed distance from the nitrogen [69]. It was found that for a distance of 1.4 Å the correlation is almost as good as that for V_{\min} . However, for distances beyond 1.6 Å the quality of the correlation degrades quickly. Good correlations was also found between the absolute value of the electric field outside the nitrogen and the hydrogen bond basicity. In this case the best correlation was achieved for a distance of 2.5 Å from the nuclei while shorter and longer distances gave significantly worse correlations. Finally, it was demonstrated that if both the potential and the field are used in a dual parameter relationship, it is possible to get a good correlation for any distance between 1.0 and 2.7 Å from the basic nitrogen.

One disadvantage with the original correlations of Murray and Politzer and that of Kenny is that they refer to families of related molecules. Since Murray and Politzer only used a minimal basis set for their potential calculations, we decided to investigate if less family dependent correlations for hydrogen bond basicity and acidity can be obtained if the potential is calculated at a higher level of theory. Both V_{\min} and $V_{S,\min}$ were computed at the HF/6-31G* level for a diverse group of 33 hydrogen-bond-accepting molecules whose $\log K_{\text{HB}}$ values for 1:1 complexation with 4-nitrophenol had been determined experimentally [52]. In view of the earlier results of Murray and Politzer and those of Kenny, we were surprised to find a better correlation between $V_{S,\min}$ and $\log K_{\text{HB}}$ than between V_{\min} and $\log K_{\text{HB}}$. However, the $V_{S,\min}$ versus $\log K_{\text{HB}}$ relationship is still of much lower statistical significance ($R=0.902$) than the previously reported family-dependent relationships. Our study indicated that it might be necessary to include also charge transfer and polarization parameters in order to get family independent correlations for the hydrogen bond basicity.

In Table 4 are listed experimentally derived α_2^{H} values, together with HF/6-31G* computed $V_{S,\max}$ values, for a group of 18 hydrogen bond donors of different types, including CH, NH and OH donors. There is an excellent linear relationship between $V_{S,\max}$ and the statistically corrected α_2^{H} values [52].

$$\alpha_2^{\text{H}}(\text{corrected}) = 0.0196 V_{S,\max} - 0.556 \quad (8)$$

$N = 18, R = 0.991$

(Note that a statistical correction is necessary to apply to the α_2^{H} values of those molecules that have several equivalent hydrogens available for hydrogen bonding.) The excellent correlation indicates that the hydrogen bond acidity of a molecule can be predicted to a high accuracy directly from its computed $V_{S,\max}$ value without the inclusion of any specific charge transfer or polarization terms.

The limited success of our family independent correlations between the potential minima (V_{\min} and $V_{S,\min}$) and $\log K_{\text{HB}}$ suggested that the hydrogen bond basicity of a molecule is not determined entirely by its electrostatic properties, and that other energy terms, such as polarization and charge transfer, also should be considered. However, our studies also indicated that some of the problems was associated with that we were correlating a free energy rather than an enthalpy, since the entropical contributions to the free energy might be irregular and hard to estimate. For example, the application of statistical corrections to the hydrogen bond basicity is not as straightforward as for the hydrogen bond acidity; it is not clear if the number of donating atoms or the number of lone pairs should be counted. The solvent effects on the free

Table 4

HF/6-31G* computed $V_{S,max}$ and experimental hydrogen bond acidities α_2^H ^a

	$V_{S,max}$ (kcal/mol)	α_2^H	Predicted α_2^H (corrected) ^b
CH ₃ COCH ₃	22.3	0.04 (-0.13)	-0.12
CH ₃ CN	27.8	0.09 (-0.01)	-0.01
CH ₂ Cl ₂	30.2	0.13 (0.07)	0.04
CH ₃ NO ₂	33.5	0.12 (0.02)	0.10
C ₆ H ₅ NH ₂	36.3	0.26 (0.20)	0.15
CHCl ₃	36.7	0.20	0.16
CH ₃ CH ₂ OH	47.0	0.33	0.36
CH ₃ OH	47.8	0.37	0.38
Pyrrole	48.3	0.41	0.39
Indole	49.6	0.44	0.41
CH ₃ COOH	56.4	0.55	0.55
C ₆ H ₅ OH	58.4	0.60	0.59
2-Naphtol	59.7	0.61	0.61
CF ₂ CH ₂ OH	62.1	0.57	0.66
<i>p</i> -C ₆ H ₄ (Cl)OH	64.1	0.67	0.70
(CF ₃)COH	70.1	0.86	0.82
<i>p</i> -C ₆ H ₄ (OH)NO ₂	72.5	0.82	0.86
CF ₃ COOH	73.3	0.95	0.88

^aAll data are taken from ref. [52]. The α_2^H values were originally obtained from R. W. Taft. A statistical correction to α_2^H has been applied to those molecules having N equivalent hydrogens available for hydrogen bonding, as indicated by parentheses. α_2^H is defined by $\alpha_2^H = (\log K_a + 1.1) / 4.636$, where K_a is the equilibrium constant for a 1:1 complex of the donor and a reference acceptor. $K_a = NK'_a$, where K'_a is the corrected value. Accordingly, $\log K_a = \log N + \log K'_a$
 $\alpha_2^H(\text{corrected}) = \alpha_2^H - (\log N)/4.636$

^b Predicted using eq. 8.

energies may also be hard to estimate. In Table 3 are listed computed HF/6-31G* computed quantities and experimental enthalpies for the formation of 1:1 complexes with phenol in apolar solvents for a diverse group of oxygen, nitrogen and sulfur donors. The enthalpies are taken from compilations by Drago and co-workers [51, 60, 61, 70]. They have argued that the solvent effects on complexation enthalpies can be minimized by using alkane solvents for sulfur donors and strong nitrogen donors, such as pyridines and amines, and carbon

tetrachloride for oxygen donors and weak nitrogen donors, such as nitriles [71, 72]. However the physical basis for this selection criteria has been questioned [73, 74], and it cannot be excluded that some of the enthalpies have significant contributions from solvent effects. It can first be noted that there is a fair correlation between V_{\min} and the experimental enthalpies with a correlation coefficient of 0.84. A much better correlation is obtained if the polarization parameter, PV_{\min} , and the charge transfer parameter $\bar{I}_{S,\min}$ are included in the correlation [50]:

$$-\Delta H_{Ph} = -0.0778 V_{\min} - 0.0513 PV_{\min} - 0.345 \bar{I}_{S,\min} + 3.20 \quad (9)$$

t-stat	17.85	5.34	13.08	4.53
--------	-------	------	-------	------

N = 25, R = 0.978, SD = 0.38, F = 160

The correlation is improved further by the inclusion of a fourth parameter, Π [50].

$$-\Delta H_{Phen} = -0.0867 V_{\min} - 0.0369 PV_{\min} - 0.245 \bar{I}_{S,\min} - 0.0693 \Pi + 2.93 \quad (10)$$

t-stat	18.02	5.97	6.63	4.42	5.65
--------	-------	------	------	------	------

N = 25, R = 0.989, SD = 0.27, F = 231

Π is computed from the entire surface electrostatic potential and reflects the polarity of the molecule. We will discuss the Π parameter in more detail in a later section. For the above correlations, we have listed the statistical t-scores of the different parameters. According to the t-scores for the four-parameter correlation, V_{\min} is by far the most important parameter followed by $\bar{I}_{S,\min}$, PV_{\min} and Π . This is consistent with the earlier findings concerning the importance of electrostatics in hydrogen bonding. However, it is clear that for such a diverse group of bases as this, the hydrogen-bond-donating ability cannot be described properly without explicit consideration of polarization and charge transfer effects. For example, in order to account for the complexation enthalpies of the sulfur bases, it is necessary to consider for their large charge transfer capacities, as indicated by the low $\bar{I}_{S,\min}$ values, since their V_{\min} are very weak. In a similar manner, the enhanced enthalpies of the secondary and particularly the tertiary amines cannot be reproduced without a polarization parameter. The function of the Π parameter is more difficult to interpret. Especially, since the negative sign of the Π coefficient shows that the magnitude of the complexation enthalpy decrease with an increase in polarity. A plausible explanation is that Π to some degree accounts for the solvent effects on the enthalpies; large polarity strengthens the binding to the solvent and consequently reduces the magnitude of the complexation enthalpy.

3.2 Frequency Shifts

Frequency shifts have been used extensively as measures of the strengths of hydrogen-bonding interactions [51, 53, 54, 58, 59, 72]. In particular, it has been shown for several alcohols that the shift in the O-H stretching frequency upon 1:1 complexation with nitrogen and oxygen donors is linearly correlated to the complexation enthalpy [51, 72]. Similar relationship does also exist between the O-H shift and the enthalpy for complexation with sulfur donors [72]. However, the sulfur donors cannot be put together with the nitrogen and oxygen donors, since the sulfur donors induces relatively larger shifts compared to the their low enthalpies. It has been suggested that the reason is that the magnitude of the frequency shift depend upon the increase in the electron density of the O-H bond due to electron transfer from the base to the alcohol [72]. Complexation with the softer sulfur donors results in relatively larger degrees of charge transfer and consequently bigger shifts. In Table 3 are listed methanol O-H frequency shifts for 1:1 complexation with hydrogen bond acceptors in carbon tetrachloride. The frequency shifts are well correlated by the same type of relationship as the phenol-base complexation enthalpies [50].

$$\Delta\nu_{\text{OH}} = -4.71V_{\text{min}} - 2.38 P_{V_{\text{min}}} - 27.8 \bar{I}_{\text{S,min}} - 5.39 \Pi + 201.1 \quad (12)$$

t-stat	17.57	6.86	15.3	6.37	7.45
--------	-------	------	------	------	------

N = 37, R = 0.985, SD = 19, F = 258

According to the t-scores, V_{min} is the most important parameter also for this correlation. However, the relative importance of $\bar{I}_{\text{S,min}}$ is much larger than in the correlation with the phenol-base complexation enthalpies, consistent with the larger effect of charge transfer on frequency shifts compared to enthalpies.

3.3 Protonation

As has already been mentioned, protonation is a considerably stronger interaction than hydrogen bonding. A protonation reaction generally also involves a considerable degree of charge transfer and polarization. This can be realized from Table 3, which shows that the maximum magnitude of the electrostatic interaction energy between a neutral base and a proton, as given by V_{min} , for most of the listed bases corresponds to less than one third of the total gas phase proton affinity (ΔH_{H^+}). Thus, polarization, charge transfer and to some degree relaxation of nuclear geometry, together, contribute in most cases to more than two third of the total interaction energy. It can also be noted that there is no significant overall correlation between V_{min} and ΔH_{H^+} for the molecules in Table 2; the linear correlation coefficient is not better than 0.67. However, it

should be noted that correlations between V_{\min} and both gas phase and aqueous acidities have been reported for families of molecules [21, 25, 47, 49, 75-77]. The existence of such correlations implies that the charge transfer and polarization effects are either constant or varies linearly with the electrostatic effect within each family of molecules. In particular, we have found that for families of substituted aromatics the electrostatic interaction energy often varies linearly with the charge transfer energy [49, 76, 77], while the polarization energy seems to be nearly independent of substitution [77]. This is indicated by linear correlations between V_{\min} and $\bar{I}_{S,\min}$ and nearly constant PV_{\min} values. Substituent effects on gas phase and solution basicities will be discussed in more detail in the next section of this chapter.

We have found that family independent relationships for gas phase basicities can be obtained if not only electrostatic effects but also charge transfer and polarization effects are considered explicitly. The experimental gas phase proton affinities of the molecules in Table 2 have been found to follow the following relationship [50]:

$$-\Delta H_{H^+} = -0.294 V_{\min} - 0.698 PV_{\min} - 4.37 \bar{I}_{S,\min} + 206.5 \quad (13)$$

t-stat	5.87	12.49	12.59	31.4
	N = 36, R = 0.975, SD = 3.6, F = 206			

According to the statistical t-scores, PV_{\min} and $\bar{I}_{S,\min}$ are the two most important parameters in this correlation. This is consistent with our findings from above that polarization and charge transfer generally constitutes more than two third of the total protonation energy. It is also consistent with the results from earlier studies in which Morokuma decomposition analysis have been used to investigate the relative contributions of electrostatic, polarization and charge transfer energies to the proton affinities of neutral bases [23, 78]. It should be noted that inclusion of the Π parameter did not improve the correlation, which is consistent with our interpretation that Π reflects solute-solvent interactions.

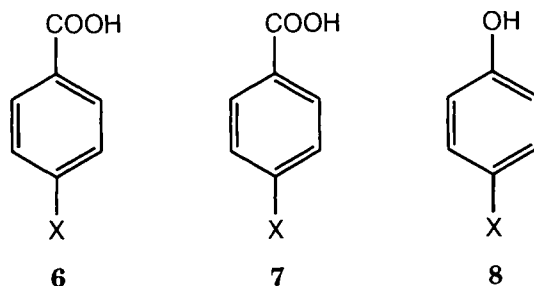
The correlation is surprisingly good considering that only properties of the isolated bases are included and that protonation is a strong interaction which leads to more than a small perturbation of the charge distribution of the base. However, it should be noted that dimethylformamide and dimethylacetamide are not included in the correlation. The relationship underestimates their proton affinities by 12.0 and 12.6 kcal/mol, respectively. We have suggested that these discrepancies can be attributed to large resonance stabilization effects in the protonated forms of these compounds [50].

4. ANALYSIS OF SUBSTITUENT EFFECTS ON CHEMICAL REACTIVITY

4.1 Background

Substituent constants have been widely used in organic chemistry for the interpretation of electronic effects on chemical reactivity [79, 80]. However, it is well known that the character of the electronic effects can differ quite substantially between different chemical systems and that it is not possible to find a universal set of substituent constants. In this section we will demonstrate how the electrostatic potential can be used as an alternative and a complement to the traditional constants in analyses of electronic substituent effects. The main advantage with this approach is that we can study systems whose substituent effects are unknown, or systems for which the substituent effects cannot be described by the traditional constants. An analysis of the electrostatic potential can also provide a direct insight to the effects the substituents have on the electronic structures of the molecules. We will both discuss the classical examples of substituent effects on the acidities of benzoic acid and phenol, and a more novel example, i.e. the substituent effects on the O-H bond dissociation energy in phenols.

4.2 Acidities of Aromatic Systems



Some of the more commonly used substituent constants are the σ_p , σ_p^0 and σ_p^- scales. These constants were originally defined from the aqueous acidities of systems **6**, **7** and **8**, respectively [79, 80]. In all three systems, it is observed that electron donating substituents decrease the acidity, whereas electron accepting substituents have the opposite effect. The observed substituent effects mainly reflect the stabilizing and destabilizing effects of the substituents on the negative charge associated with the oxygens in the ionized forms of the systems. While the substituents have stabilizing and destabilizing effects also on the neutral molecules, these are generally of smaller magnitude. Both inductive and resonance effects are considered to be important for the stabilization. In

particular, through-resonance interactions can in some cases lead to large substituent effects.

By computing the V_{\min} associated with the oxygens, it is possible to get a direct estimate of the delocalization of the oxygen charge by resonance and inductive interactions in the ionic forms of these systems. We will first consider the benzoic acids **6**, which is the system Hammett used in his original definition of substituent constants. In Table 5 are listed anion V_{\min} , computed gas phase deprotonation energies (ΔE) and experimental relative free energies of deprotonation ($\Delta\Delta G_g$) for a group of substituted benzoic acids. It can first be noted that there is a good agreement between the relative computed deprotonation energies ($\Delta\Delta E$) and the experimental $\Delta\Delta G_g$ values, indicating that the chosen computational level is capable of reproducing the substituent effects in this system. There is also a direct agreement between the calculated V_{\min} shifts (ΔV_{\min}) and $\Delta\Delta G_g$. The largest deviation between ΔV_{\min} and $\Delta\Delta G_g$ is found for the strongest acceptor NO_2 and is only 1.3 kcal/mol. It is interesting to note that for the donor substituents the agreement between ΔV_{\min} and $\Delta\Delta G_g$ is actually better than between $\Delta\Delta E$ and $\Delta\Delta G_g$. While the one-to-one correspondence between ΔV_{\min} and $\Delta\Delta G_g$ might be slightly fortuitous, it should be remembered that V_{\min} corresponds to the largest possible electrostatic interaction energy between a molecule and a proton. The substituent effects on the gas phase acidity in this system can therefore be interpreted as largely reflecting the change in the electrostatic interaction energy between the anion and the proton upon substitution. However, it should be noted that there also exists an excellent linear relationship between the oxygen $\bar{I}_{S,\min}$ and $\Delta\Delta G_g$ [77]. This suggests that the charge transfer and electrostatic substituent effects are linearly correlated. The calculated PV_{\min} are, on the other hand, all within 0.8 kcal/mol [77], indicating that the polarization contribution to the total interaction energy is nearly independent of substitution.

Table 5 also includes the values of the σ_p , σ_p^0 substituent constants. The σ_p values are directly proportional to the relative free energies of ionization in solution for the benzoic acid systems. There is a good linear relationship between our computed V_{\min} values and σ_p , with a correlation coefficient of 0.971. Since there is a near one-to-one correspondence between V_{\min} and $\Delta\Delta G_g$, an equally good relationship exists between $\Delta\Delta G_g$ and σ_p . This shows that the substituent effect in solution to a great extent is linearly related to the substituent effect in the gas phase. However, the substituent effect in the gas phase is nearly 11 times larger than in solution [81]. Thus, the relative solvation effect is proportional but opposite in sign to the relative gas phase acidity, and therefore the solvation effect must also be linearly related to the oxygen V_{\min} . This is not entirely surprising, since the solvation effect is likely to be dominated

Table 5

Computed oxygen V_{\min} for benzoate anions, computed deprotonation energies and experimental gas phase acidities for benzoic acids, and experimentally derived substituent constants

	V_{\min}^a	ΔV_{\min}^b	$\Delta E^{a,c}$	$\Delta\Delta E^d$	$\Delta\Delta G^e$	σ_p^f	σ_p^{0f}
	(kcal/mol)	(kcal/mol)	(kcal/mol)	(kcal/mol)	predicted	(kcal/mol)	(kcal/mol)
<i>p</i> -NH ₂	-189.4	-2.2	353.5	-3.6	-2.3	-0.30	-0.57
<i>p</i> -MeO	-188.0	-0.8	351.7	-1.9	-0.8	-0.12	-0.28
<i>p</i> -Me	-188.0	-0.8	351.2	-1.4	-1.1	-0.14	-0.14
H	-187.2	0.0	349.8	0.0	0.0	0.00	0.0
<i>p</i> -F	-184.3	2.9	346.6	3.2	2.9	0.15	0.06
<i>p</i> -Cl	-181.9	5.3	344.7	5.1	4.4	0.24	0.22
<i>p</i> -CF ₃	-179.3	7.9	341.9	8.0	n.a.	0.53	0.53
<i>p</i> -CN	-176.1	11.1	338.6	11.2	10.9	0.71	0.71
<i>p</i> -NO ₂	-174.2	13.0	336.7	13.1	11.7	0.81	0.81

^a V_{\min} and ΔE were computed at the HF/6-31+G*/6-31G* level and are taken from ref. [77].

^b $\Delta V_{\min} = V_{\min}(\text{X-C}_6\text{H}_4\text{COO}^-) - V_{\min}(\text{C}_6\text{H}_5\text{COO}^-)$.

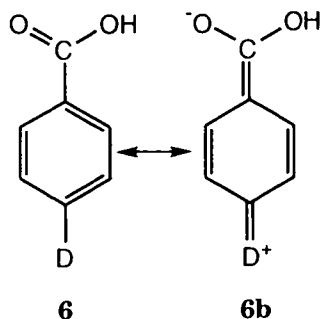
^c ΔE for the reaction $\text{X-C}_6\text{H}_4\text{COOH} \rightarrow \text{X-C}_6\text{H}_4\text{COO}^- + \text{H}^+$.

^d ΔE for the reaction $\text{X-C}_6\text{H}_4\text{COO}^- + \text{C}_6\text{H}_5\text{COOH} \rightarrow \text{X-C}_6\text{H}_4\text{COOH} + \text{C}_6\text{H}_5\text{COO}^-$.

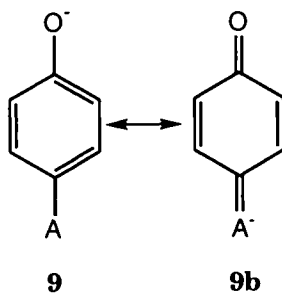
^e $\Delta G(600\text{K})$ for the reaction $\text{X-C}_6\text{H}_4\text{COO}^- + \text{C}_6\text{H}_5\text{COOH} \rightarrow \text{X-C}_6\text{H}_4\text{COOH} + \text{C}_6\text{H}_5\text{COO}^-$.

^fReference [79].

by the solvation energy of the anion, which largely is determined by the strengths of the hydrogen bonds with the oxygens [81]. Consequently, it is expected that the solvation effect will correlate with oxygen V_{\min} , since V_{\min} correlates with hydrogen bond strength. In relation to this, it is interesting to note that both V_{\min} and $\Delta\Delta G_g$ correlates better with σ_p^0 than with σ_p . Both σ_p^0 relationships have correlation coefficients better than 0.99, whereas the correlations coefficients for the σ_p relationships only are slightly better than 0.97. The superior correlations with σ_p^0 is a result of the more negative σ_p values for the resonance donors OCH₃ and NH₂. These substituents can stabilize the neutral benzoic acid by through-resonance interactions as illustrated by the resonance structure **6b** shown below. This stabilization is expected to be much more important in aqueous solution than in the gas phase, since structure **6b** is favorably solvated [81]. The σ_p^0 scale is defined by system **7** in which resonance donor cannot participate in through-resonance interactions, which explains why it better than the σ_p scale reflects the substituent effects of resonance donors on the gas phase acidities of benzoic acids.



As has already been mentioned, the aqueous acidities of substituted phenols were originally used to define the σ_p^- scale of substituent constants. The main difference between the σ_p and σ_p^- scales is that the σ_p^- constants for resonance acceptors are significantly larger. The reason for this discrepancy is that the phenoxide anion is stabilized by through-resonance interactions with acceptor substituents as shown below. Another difference between the phenol and benzoic acid systems is that the neutral phenol molecule cannot be stabilized by through-resonance with donors. The σ_p^- constants for resonance donors are therefore more similar to σ_p^0 than to σ_p . We have previously shown that the oxygen V_{\min} of the phenoxide anion correlates very well with σ_p^- for a large number of substituents, including strong resonance donors [76]. An equally good relationship was also found between V_{\min} and the gas phase acidity. Furthermore, the variations in V_{\min} are about two times larger in the phenol system than in the benzoic acid systems. This is consistent with the observations that the substituent effects on both gas phase and solution acidities are much larger for phenol than for benzoic acid [81]. Thus, this shows that V_{\min} calculations cannot only be used to understand the effect of different substituents on a specific reaction system but can also provide an indication of the relative magnitudes of substituent effects in different systems.



4.3 O-H Bond Dissociation Energies in Phenols

The analysis of acidities in aromatic systems in terms of linear free energy relationships is one of the many examples of the use of substituent constants for understanding reactions that involve heterolytic bond cleavage. Substituent effects on homolytic bond dissociation energies, on the other hand, are not as well investigated and understood as the heterolytic dissociation process. Several studies have shown that the homolytic O-H bond dissociation energy (BDE) in substituted phenols is linearly correlated to the σ_p^+ constant [82-84]. The BDE increases with the electron-donating capacity of the substituent. It has been suggested that substituent effects on BDEs can be divided into polar and radical stabilization effects [84-90]. The polar effect refers to the relative stabilization of the parent molecule by resonance and inductive interactions. The radical effect is the substituent effect on the radical stability and is expected to depend on the spin delocalization of the unpaired electron. In the case of O-H BDEs in phenols there has been a controversy regarding which of the two effects that is dominating [87, 88]. We decided to investigate if an analysis of the electronic structures of substituted phenols and phenoxy radicals could provide some guidance regarding the relative importance of the two effects. First of all it should be noted that the selected computational method, B3LYP/6-31G**, provide accurate phenol BDEs. This was demonstrated by a comparison with experimentally determined BDEs for a group of phenols [91].

A commonly used concept in chemistry is that the energy required to break a bond depend on the properties of the bond itself [92, 93]. For example, bond strength has been shown to correlate with properties such as the bond length, the bond force constant and the electron density minimum along the bond path [93-97]. However, in the case of phenols, we found that these properties show very small substituent effects despite the rather large substituent effect on the O-H BDE [91]. As an alternative explanation, Bordwell has suggested that the polar effect on the O-H BDE in phenols reflects the substituent's ability to stabilize the phenol by delocalization of the oxygen lone pair [84]. To test this hypothesis, we decided to use the magnitude of the oxygen V_{\min} as a measure of the lone pair delocalization. In Figure 4 we have plotted the computed V_{\min} for a group of substituted phenols versus their computed relative BDEs (Δ BDEs). There is a fair linear relationship between V_{\min} and Δ BDE for the phenols with electron-withdrawing substituents. However, for the electron donating substituents, V_{\min} changes more slowly, which does not reflect the large changes in the BDE. Thus, our results indicate that lone pair delocalization is important for determining the Δ BDEs of phenols with electron-donating substituents, while other effects dominate for electron-accepting substituents.

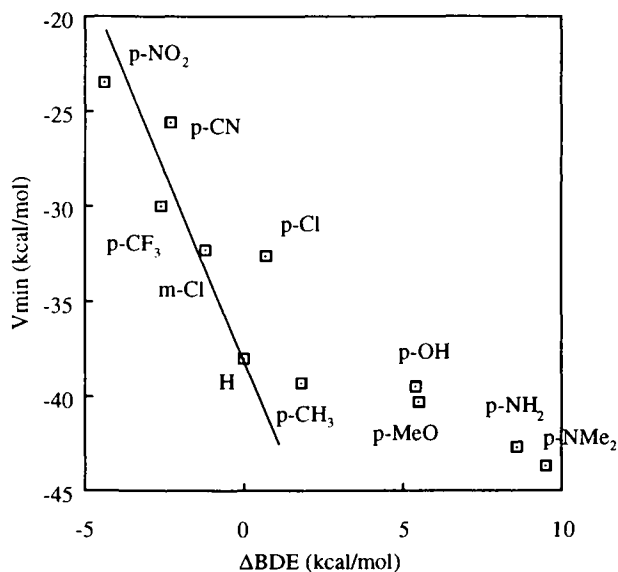


Figure 4. Plot of oxygen V_{\min} versus ΔBDE for phenols. Reprinted with permission from ref. [91]. Copyright 1997 American Chemical Society.

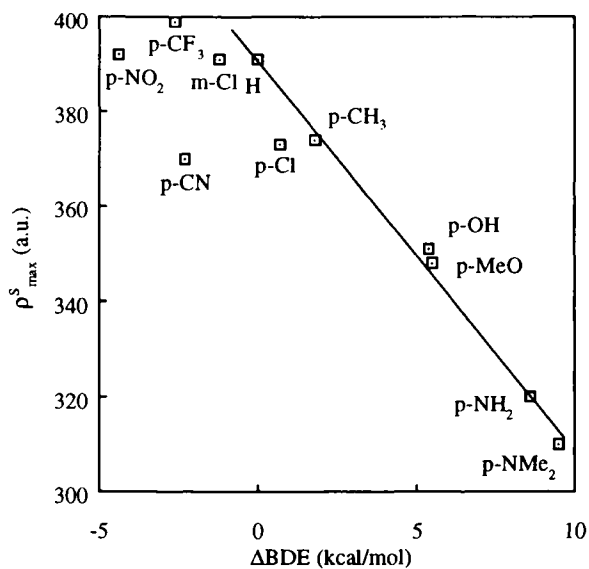


Figure 5. Plot of oxygen ρ_{\max}^S for phenoxy radicals versus ΔBDE for phenols. Reprinted with permission from ref. [91]. Copyright 1997 American Chemical Society.

In order to estimate the radical stabilization effect on the BDE, we decided to study how the spin delocalization of the phenoxyl radical varies with the substituent. For this purpose, we computed the surface maxima in the spin density associated with the oxygens of the substituted phenoxyl radicals. The molecular surface was defined by the 0.002 a.u. contour of the electron density. By calculating the spin density on the molecular surface rather than at the nuclei, the spin density will emphasize the spin delocalization of the valence electrons, which is expected to be the most important for reactivity. For example, the spin density at the nuclei does not reflect the spin delocalization of the π -electrons, since these generally have zero densities at the nuclear positions.

In Figure 5 we have plotted the oxygen surface maxima in the spin density, designated as $\rho^{\text{S}_{\text{max}}}$, versus the computed ΔBDE for the phenols. There is a linear correlation between $\rho^{\text{S}_{\text{max}}}$ and ΔBDE for the phenols with electron-donating substituents. This indicates that the ΔBDE s of these compounds mainly are determined by the ability of the substituents to stabilize the radical through spin delocalization of the unpaired electrons. For the phenols with electron donating substituents, on the other hand, the variations in the $\rho^{\text{S}_{\text{max}}}$ are much smaller, and they don't reflect the changes in the ΔBDE .

On the basis of the linear correlations between V_{min} and ΔBDE for electron-accepting substituents and $\rho^{\text{S}_{\text{max}}}$ and ΔBDE for electron-donating substituents, we decided to investigate if a dual parameter relationship of the following type could correlate the ΔBDE of all phenols:

$$\Delta\text{BDE} = a \Delta V_{\text{min}} + b \Delta \rho^{\text{S}_{\text{max}}} \quad (14)$$

where $\Delta V_{\text{min}} = V_{\text{min}}(\text{X-C}_6\text{H}_4\text{OH}) - V_{\text{min}}(\text{C}_6\text{H}_5\text{OH})$ and $\Delta \rho^{\text{S}_{\text{max}}} = \rho^{\text{S}_{\text{max}}}(\text{X-C}_6\text{H}_4\text{O}\cdot) - \rho^{\text{S}_{\text{max}}}(\text{C}_6\text{H}_5\text{O}\cdot)$. We found a very good linear relationship with a correlation coefficient of 0.993 (see Figure 6). We interpret $-a \Delta V_{\text{min}}$ as the relative stabilization energy of the phenol and $b \Delta \rho^{\text{S}_{\text{max}}}$ as the relative stabilization energy of the radical. We call these two quantities ΔPSE (the relative polar stabilization energy) and ΔRSE (the relative radical stabilization energy), respectively. Note that positive values in these quantities implies a net stabilization relative the unsubstituted phenol or phenoxyl radical.

In Table 6 are listed the calculated ΔPSE and ΔRSE , together with computed and predicted ΔBDE s for the substituted phenols. According to the derived stabilization energies, the ΔBDE s of the phenols with electron-withdrawing substituents are mainly determined by the polar stabilization of the parent molecules. The polar effect is less important for the phenols with electron-donating substituents, but is in all cases destabilizing. For these substituents, it is instead the spin delocalization that has the greatest effect on

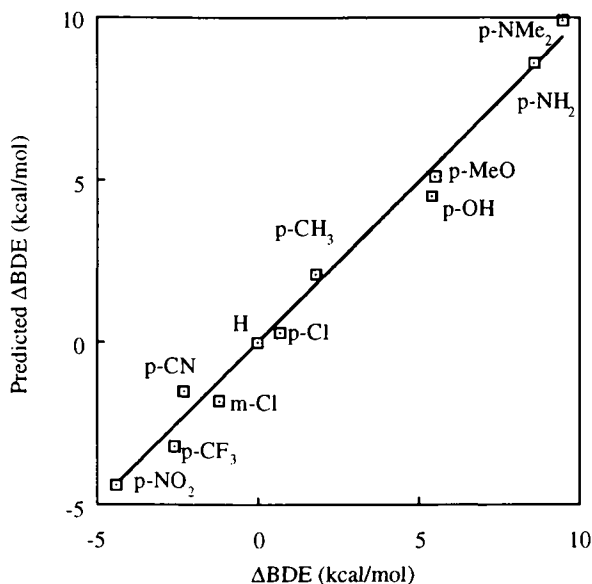


Figure 6. ΔBDE predicted from eq. 14 versus computed ΔBDE for phenols. Reprinted with permission from ref. [91]. Copyright 1997 American Chemical Society.

the BDE by stabilizing the radical. The electron-withdrawing substituents have much smaller effects on the radical stability. Considering the whole data set, the radical effect is of greater importance than the polar effect for determining the relative BDE. CF_3 is the only substituent that has a net destabilizing effect on the radical. This is consistent with the behavior of this substituent in other radical systems [98, 99].

Finally, we like to point out that our results can explain the observations that O-H BDEs in phenols correlate with σ_p^+ . Because of the direct conjugation between the oxygen lone pair and the substituent, the polar stabilization of the phenol can be expected to follow a linear relationship with σ_p^- rather than with σ_p^+ . This is also consistent with our computed ΔPSE which correlates linearly with σ_p^- with a correlation coefficient of 0.984. Since the σ_p^- and the σ_p^+ scales differ in that σ_p^+ predicts much larger substituent effects for resonance donors (e.g. OCH_3 , OH and NH_2) and relatively smaller substituent effects for resonance attractors (e.g. CN and NO_2), the overall relationship between ΔBDE and σ_p^+ can be explained by the observed extra stabilization of the radical by electron donating substituents. Thus to understand the substituent effects on the O-H

Table 6
B3LYP/6-31G** computed molecular properties, O-H bond dissociation energies and stabilization energies for some phenols^a

	ΔV_{\min}^b (kcal/mol)	$\Delta \rho_{\max}^S{}^c$ (a.u.)	ΔBDE^d (kcal/mol)	ΔBDE^e predicted	ΔPSE^f (kcal/mol)	ΔRSE^g (kcal/mol)
<i>p</i> -NMe ₂	-5.7	-81	9.5	9.9	-1.7	8.2
<i>p</i> -NH ₂	-4.7	-71	8.6	8.6	-1.4	7.2
<i>p</i> -MeO	-2.3	-43	5.5	5.1	-0.7	4.4
<i>p</i> -OH	-1.5	-40	5.4	4.5	-0.4	4.1
<i>p</i> -Me	-1.3	-17	1.8	2.1	-0.4	1.7
<i>p</i> -Cl	5.4	-18	0.7	0.2	1.6	1.8
H	0	0	0	0	0	0
<i>m</i> -Cl	5.7	0	-1.2	-1.8	1.7	0.0
<i>p</i> -CN	12.4	-21	-2.3	-1.5	3.6	2.1
<i>p</i> -CF ₃	8.0	8	-2.6	-3.2	2.4	-0.8
<i>p</i> -NO ₂	14.5	1	-4.4	-4.4	4.3	-0.1

^aAll data are taken from ref. [91].

^b $\Delta V_{\min} = V_{\min}(X-C_6H_4OH) - V_{\min}(C_6H_5OH)$.

^c $\Delta \rho_{\max}^S = \rho_{\max}^S(X-C_6H_4O\cdot) - \rho_{\max}^S(C_6H_5O\cdot)$.

^dCorresponds to ΔE for the reaction $X-C_6H_4O\cdot + C_6H_5OH \rightarrow X-C_6H_4OH + C_6H_5O\cdot$. Note that a positive value indicates that the bond in the substituted phenol is weaker than in phenol.

^e $\Delta BDE = -0.294 \Delta V_{\min} - 0.102 \Delta \rho_{\max}^S$.

^f $\Delta PSE = -0.294 \Delta V_{\min}$ (the relative polar state stabilization energy).

^g $\Delta RSE = -0.102 \Delta \rho_{\max}^S$ (the relative radical stabilization energy).

BDE in phenols, it is necessary to consider both polar and radical stabilization effects.

5. STATISTICALLY-BASED INTERACTION INDICES

5.1 Background

In the preceding sections of this chapter, we have focused on the use of extrema in the electrostatic potential, i.e. V_{\min} , $V_{S,\min}$ and $V_{S,\max}$, for the interpretation and prediction of site-specific molecular interactions. However, additional information about a molecule's ability to interact with other molecules can be obtained by an analysis of the overall pattern of the electrostatic potential on the molecular surface. Politzer and co-workers have in recent years developed a number of statistically-based interaction indices that are defined in terms of the entire surface electrostatic potential [100-102]. They have further shown

that there exist quantitative relationships between analytical functions of these quantities and a number of macroscopic properties that reflect molecular interaction tendencies. The properties that have been correlated include partition coefficients [103, 104], solubilities in supercritical fluids [101, 105], critical constants (temperatures, pressures and volumes) [102], boiling points [102], heats of fusion [106], and diffusion coefficients [107].

It is not within the scope of this article to present a full review of all the applications of statistically-based interaction indices. We will instead give a short presentation of the most important indices followed by an example of the use of these indices in the analysis of a solvation process; the partitioning of solutes between octanol and water.

5.2 Definitions

As a measure of local polarity, a quantity Π has been defined by [100]

$$\Pi = \frac{1}{n} \sum_{i=1}^n |V(\mathbf{r}_i) - \bar{V}_S| \quad (15)$$

where $V(\mathbf{r}_i)$ is the potential at the i^{th} point on the surface and \bar{V}_S is the average surface potential. Π can be viewed as the average deviation of the electrostatic potential on the molecular surface. For a spherically symmetric system, such as a ground state atom, $\Pi=0$, since $V(\mathbf{r}) = V(r) = \bar{V}_S$. It should be noted that Π permits a quantitative assessment of the total local polarity, even in a molecule that has a zero dipole moment. We have shown for a group of solvents that Π correlates well with the solvatochromic polarity/polarizability parameter ($\pi^* + d\delta$) [59], with $d=-0.4$ [100]. It has been suggested that this term mainly reflects polarity when $d=-0.4$ [108]. We also found a fair correlation between Π and the dielectric constant [100]. In this chapter we have already discussed the use of Π in correlations with phenol-base complexation enthalpies and O-H frequency shifts in methanol-base complexes. It was suggested that Π in these relationships mainly reflect non-specific solute-solvent interactions. In Table 3 are listed the Π values for the bases included in the correlations. It is easily recognized that there is no direct correct correlation between V_{\min} and Π . This is expected, since V_{\min} corresponds to the electrostatic potential at one particular point in the space of a molecule, while Π is a global quantity, which is computed from the entire surface electrostatic potential.

The quantity σ_{tot}^2 , which is defined by eq. 16, reflects the variability of the electrostatic potential on the molecular surface [101].

$$\sigma_{\text{tot}}^2 = \sigma_+^2 + \sigma_-^2 = \frac{1}{m} \sum_{i=1}^m [V^+(\mathbf{r}_i) - \bar{V}_S^+]^2 + \frac{1}{n} \sum_{i=1}^n [V^-(\mathbf{r}_i) - \bar{V}_S^-]^2 \quad (16)$$

The first summation is over the surface points with positive potential and the second over the points with negative potential. \bar{V}_S^+ and \bar{V}_S^- are the positive and negative surface averages in $V(\mathbf{r})$, respectively. Since the terms in eq. 10, are squared, σ_{tot}^2 is, in contrast to Π , particularly sensitive to the extremes in $V(\mathbf{r})$. The two quantities have also been found to be quite different and even been found to vary in opposite directions for some groups of molecules [106]. σ_{tot}^2 is considered to be indicative of a molecule's electrostatic interaction tendencies. For example, σ_{tot}^2 has been used in conjunction with measures of molecular size, i.e. surface area or volume, for correlating solubilities in supercritical fluids [101, 105]. It has been suggested that σ_{tot}^2 in these relationships reflect solute-solute interactions, since the supercritical solubility mainly is determined by the solute vapor pressure [105].

Finally, a balance parameter v has been defined by [102]

$$v = \frac{\sigma_+^2 \sigma_-^2}{[\sigma_{\text{tot}}^2]^2} \quad (17)$$

v reaches its limiting value 0.250 when σ_+^2 and σ_-^2 approaches the same magnitude. Thus v gives an indication of the balance in a molecule's negative and positive electrostatic potential. For example, polar aprotic solvents, such as diethyl ether, acetone or DMSO, generally have very low v values while the values for protic solvents like alcohols are considerably larger. The product $v\sigma_{\text{tot}}^2$ has been found to be useful for correlating properties that reflect interactions between molecules of the same kind, e.g. boiling points, critical temperatures and critical pressures [102, 109].

5.3 Predictions of Octanol/Water Partition Coefficients

The octanol/water partition coefficient is one of the most frequently used descriptors in biological quantitative structure activity relationships. It is considered to reflect the hydrophobicity of a molecule and therefore to be relevant both for correlating the transport properties and the receptor binding of biologically active molecules. Since pharmacological and toxicological research often concerns poorly characterized or not yet synthesized molecules, there is a

great need for methods that allow accurate predictions of partition coefficients without the use of experimental data. Several empirical methods for the prediction of prediction octanol/water partition coefficients have also been proposed. Among the more widely used are the fragment approaches of Recker [110], and Hansch and Leo [111]. While the predictive power of these methods is very good for simple organic molecules, they are less accurate for complicated drug molecules with several interaction sites [112].

The linear solvation energy approach(LSER) has been used successfully for correlating octanol/water partition coefficients [61, 108]. It has been shown for a large number of molecules that the logarithm of the partition coefficient ($\log P_{ow}$) can be expressed as a linear combination of a cavitation term (molecular volume), a polarity/polarizability term, and hydrogen bond acidity and basicity terms [108]. Unfortunately, the LSER approach is not directly applicable to the prediction of partition coefficients of novel drug molecules, since the LSER parameters are determined by elaborate experimental procedures. However, the functional form of the relationship provides insight into the molecular characteristics that are important for the partitioning between water and octanol. Thus, the LSER equation can be helpful in the development of relationships with theoretical descriptors. The LSER studies showed the cavitation term and the hydrogen bond basicity term to be most significant [108]. The polarity/polarizability term was also found to be highly significant, while the hydrogen bond acidity term made a very small but still significant contribution to the overall relationship. On the basis of this information, we decided to investigate the use of statistically-based interaction indices for correlating octanol/water partition coefficients. Surface electrostatic potentials were computed at the HF/STO-5G**//STO-3G* level for a set of 70 organic molecules of various types and sizes [103]. It can first be noted that a fairly good correlation can be obtained by the simple relationship

$$\log P_{ow} = 0.0243 (\text{area}) - 0.0109 \sigma_{tot}^2 - 0.261 \quad (18)$$

N = 67, R = 0.936, SD = 0.552

The correlation improves significantly when σ_{tot}^2 is substituted for σ_-^2 and the term (area) Π is introduced

$$\log P_{ow} = 0.0298 (\text{area}) - 0.000849 (\text{area})\Pi + 0.00912\sigma_-^2 - 0.529 \quad (19)$$

N = 67, R = 0.961, SD = 0.437

This relationship resembles the LSER equation in many ways. In both relationships a cavitation term (surface area and molecular volume, respectively) is

found to be highly important. This reflects that the free energy required to form a cavity for the solute in water is significantly higher than in octanol. The σ_-^2 term can be considered to correspond to the hydrogen bond basicity term in the LSER equation, since σ_-^2 emphasizes the negative extrema in the potential, which are known to be important for the hydrogen bond accepting ability. The large significance of this term indicates that the better ability of water, in comparison with octanol, to donate hydrogen bonds to the solute is of great importance for the partitioning. Finally, the (area) Π term in our relationship can be compared with the polarity/polarizability term of the LSER equation.

While eq. 19 was found to provide a rather good overall correlation, a number of problems with the equation was also observed. It was found to almost consistently overestimate the $\log P_{ow}$ of nitrogen donors and underestimate the $\log P_{ow}$ of oxygen donors. In addition, the relationship was in some cases found to give poor predictions for molecules with multiple interaction sites. These results indicate that the relationship cannot account fully for the important hydrogen bond interactions between the solute and the solvents. In a more recent study, we therefore tried to improve the hydrogen bond description to find a better method for prediction of partition coefficients [113]. A new data set consisting of 74 molecules was designed. Efforts were made to select molecules with large variations in size, shape and functional groups. This set include among others, heterocycles, amines, amides, alkanes and molecules with more than one functional group. Compared with the earlier study, we chose to do the calculations at a higher theoretical level (HF/6-31G*) to get a better description of the molecular charge distribution. A number of different descriptors were tested, including site specific hydrogen bond descriptors. The following equation was found to give the best three parameter relationship.

$$\log P_{ow} = 0.0278 (\text{area}) - 4.180 \times 10^{-6} (\text{area}) \overline{V_-^2} + 0.0164 \Sigma V_{\min} - 0.894 \quad (20)$$

N = 74, R = 0.979, SD = 0.316, F = 545

$\overline{V_-^2}$ is defined from the surface electrostatic potential by

$$\overline{V_-^2} = \frac{1}{n} \sum_{i=1}^n [V^-(\mathbf{r}_i)]^2 \quad (21)$$

where the summation is over all negative points on the surface. The ΣV_{\min} parameter is the sum of spatial minima in the electrostatic potential. Only V_{\min} more negative than -35 kcal/mole are included in the summation. In addition, if two V_{\min} are within 2.1 Å of each other, only the most negative is included. We

Table 7
Predictions of $\log P_{ow}$ for some biologically active molecules using eq. 21^a

	Experimental $\log P_{ow}$	Predicted $\log P_{ow}$	Residual
Caffeine	-0.07	0.18	-0.25
Clonidine	1.57	2.08	-0.51
Morpholine	-0.86	-0.42	-0.44
Nicotine	1.17	1.64	-0.47
Sulfanilamide	-0.62	-0.60	-0.02
Vanillin	1.37	1.41	-0.04

^aData taken from ref. [113]. Experimental values are from [114].

view this term as a hydrogen-bond-accepting parameter, which sums up the contributions from all the strong acceptor sites in the molecule. The inclusion of charge transfer and polarization corrections to this term did not improve the correlation. It is noteworthy that the polarity parameter $\overline{V_-^2}$, which emphasizes only the variations in the negative potential, was found to perform better than polarity parameters that reflect both the negative and positive potential regions. However, considering that the hydrogen-bond-accepting ability of the solute molecule is more important for the partitioning than its hydrogen-bond-donating ability, this is not so surprising. In particular, we found this term to be important for correlating the partition coefficients of aromatic molecules, which indicates that it accounts for hydrogen bond interactions with their π -regions. Finally, it should be noted that the relationship did not improve upon the addition of hydrogen-bond-donating terms.

To test the predictive power of this relationship, six biologically active molecules were used as a validation set. Vanillin was included in this set to investigate the predictive capability for molecules with internal hydrogen bonds. The predicted $\log P_{ow}$ for this data set are listed in Table 3 together with experimentally determined values. Several of the predicted values are very close to the experimental, e.g. caffeine and vanillin. The largest deviation is found for clonidine, for which our relationship overestimate the $\log P_{ow}$ value by only 0.51 units. These results clearly indicate that eq. 20 has a predictive capability also for more complex molecules.

We believe that relationships like eq. 20 that combines the use of global statistically based interaction indices with local interaction indices, such as the V_{min} , can be very useful for studying solvation processes. In particular, for

processes where hydrogen bonding plays an integral part, the inclusion of local indices is likely to be crucial.

6. SUMMARY

The use of the electrostatic potential in quantitative analyses of intermolecular interactions and stabilization/destabilization of molecules is the focus of this chapter. We have shown that spatial minima in the potential (V_{\min}) can be used for the identification and ranking of sites susceptible toward electrophilic attack. Surface maxima in the electrostatic potential ($V_{S,\max}$) serve the same purpose for nucleophilic attack. In particular, we have found that the hydrogen bond acidity of a molecule is directly correlated to the magnitude of its most positive $V_{S,\max}$. Also the hydrogen bond basicity has been found to be largely electrostatic in nature, as indicated by good correlations between V_{\min} and empirical hydrogen-bond-accepting scales within families of molecules. However, in order to get family independent correlations for the hydrogen bond basicity, V_{\min} has to be supplemented with parameters that reflect contributions from other energy terms than electrostatics. We have shown that a general equation, containing terms that represent electrostatics, polarization, charge transfer and polarity, can correlate phenol-base complexation enthalpies, O-H stretching frequency shifts for methanol-base complexation and proton affinities. Based on the different nature of these interactions, it is likely that this equation has the potential to correlate a wide range of acid-base interactions and provide means for predicting the interacting tendencies of poorly characterized bases.

We have also discussed the use of the electrostatic potential for the analysis of substituent effects in aromatic systems. Substituent effects on gas phase and solution acidities of benzoic acids and phenols are dominantly determined by the relative stabilization of the negative charge in the ionized forms of these systems. The oxygen V_{\min} is an excellent tool for the analysis of this stabilization effect. On the other hand, we have found that the homolytic O-H bond dissociation energy in phenols depends both on the substituent's ability to stabilize the parent molecule (the phenol) and the radical. The relative stabilization energies of the parent molecule and the radical can be estimated from their computed V_{\min} and surface maxima in the spin density, respectively.

Finally, we have discussed the use of statistically-based indices for prediction of octanol/water partition coefficients. These indices are calculated from the entire surface electrostatic potential and give an indication of a molecule's global interaction tendencies. We have found good correlations for octanol/water partition coefficients, using subsets of these quantities together with molecular surface area. However, to correctly account for the importance of hydrogen bonding on the partitioning, the global indices have to be supplemented

with a local hydrogen-bond-accepting term, i.e. the sum of V_{\min} . We believe that the combined use of global and local interaction indices can become a fruitful approach for the prediction of solvation energies and other macroscopic properties that reflect intermolecular interactions.

In conclusion, it is clear that considerable information about a molecule's inherent stability, and its ability to interact with other chemical species, can be deduced from the electrostatic potential and some other well defined properties that reflect the molecular charge distribution. It should be emphasized that this approach only requires the wavefunction of the isolated molecule to be calculated, and it is therefore considerably more economical than the conventional supermolecule approach for calculation of intermolecular interaction energies. In particular, we believe that this methodology can be very useful for studying interactions in biological systems, since these often involve large molecules with several interaction sites.

ACKNOWLEDGMENT

Financial support by the Swedish Natural Science Research Council is gratefully acknowledged.

REFERENCES

1. E. Scrocco and J. Tomasi, in *Topics in Current Chemistry*, No. 42, Springer-Verlag, Berlin, 1973.
2. E. Scrocco and J. Tomasi, *Adv. Quant. Chem.*, 11 (1978) 115.
3. P. Politzer and K. C. Daiker, in B. M. Deb (eds.), *The Force Concept in Chemistry*, Van Nostrand Reinhold Company, New York, 1981, Chap. 6.
4. P. Politzer and D. G. Truhlar, (eds.), *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, Plenum Press, New York, 1981.
5. P. Politzer and J. S. Murray, in K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry*, VCH Publishers, New York, 1991, Chap. 7.
6. J. S. Murray and P. Politzer, in P. Politzer and J. S. Murray (eds.), *Quantitative Treatments of Solute/Solvent Interactions*, Elsevier, Amsterdam, 1994, Chap. 8.
7. J. S. Murray and K. Sen, (eds.), *Molecular Electrostatic Potentials: Concepts and Applications*, Elsevier, Amsterdam, 1996.
8. K. Kitaura and K. Morokuma, *Int. J. Quant. Chem.*, 10 (1976) 325.

9. K. Morokuma, in P. Politzer and D. G. Truhlar (eds.), *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, Plenum Press, New York, 1981.
10. S. Scheiner, in K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry*, VCH Publishers, New York, 1991
11. R. K. Rathak and S. R. Gadre, *J. Chem. Phys.*, 56 (1990) 6715.
12. P. Sjoberg and P. Politzer, *J. Phys. Chem.*, 94 (1990) 3959.
13. P. Sjoberg, J. S. Murray, T. Brinck, P. Evans and P. Politzer, *J. Mol. Graphics*, 8 (1990) 81.
14. J. S. Murray, P. Lane, T. Brinck and P. Politzer, *J. Phys. Chem.*, 95 (1990) 844.
15. J. S. Murray and P. Politzer, *J. Org. Chem.*, 56 (1991) 6715.
16. O. Engkvist, P. Åstrand and G. Karlström, *J. Phys. Chem.*, 100 (1996) 6950.
17. J. H. Jensen and M. S. Gordon, *Mol. Phys.*, 89 (1996) 1313.
18. R. F. W. Bader and H. J. T. Preston, *Theoret. Chim. Acta*, 17 (1970) 384.
19. R. F. W. Bader, M. T. Carroll, J. R. Cheeseman and C. Chang, *J. Am. Chem. Soc.*, 109 (1987) 7968.
20. S. J. Weininger and F. R. Stermitz, (eds.), *Organic Chemistry*, Academic Press, Orlando, 1984, 648.
21. J. S. Murray, T. Brinck, M. E. Grice and P. Politzer, *J. Mol. Struct. (Theochem)*, 256 (1992) 29.
22. T. Brinck, J. S. Murray and P. Politzer, *Int. J. Quant. Biol., Biol. Symp.*, 19 (1992) 57.
23. P. Kollman and S. Rothenberg, *J. Am. Chem. Soc.*, 99 (1977) 1333.
24. J. S. Murray and P. Politzer, *J. Chem. Res.*, (s) (1992) 110.
25. T. Brinck, J. S. Murray and P. Politzer, *Int. J. Quant. Chem.*, 48 (1993) 73.
26. K. Morokuma, *Acc. Chem. Res.*, 10 (1977) 294.
27. P. A. Kollman, *J. Am. Chem. Soc.*, 99 (1977) 4875-4894.
28. P. A. Kollman, in P. Politzer and D. G. Truhlar (eds.), *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, Plenum Press, New York, 1981.
29. A. D. Buckingham and P. W. Fowler, *J. Chem. Phys.*, 79 (1983) 6426.
30. A. D. Buckingham and P. W. Fowler, *Can. J. Chem.*, 63 (1985) 2018.
31. A. P. L. Rendell, G. B. Bacskay and N. S. Hush, *Chem. Phys. Lett.*, 117 (1985) 400.
32. S. E. Novick, K. C. Janda and W. Klemperer, *J. Chem. Phys.*, 65 (1977) 5115.
33. F. A. Baiocchi, T. A. Dixon and W. Klemperer, *J. Chem. Phys.*, 77 (1982) 1632-1638.

34. S. W. Bunte, J. B. Miller, Z. S. Huang, J. E. Verdasco, C. Wittig and R. A. Beaudet, *J. Phys. Chem.*, 96 (1992) 4140.
35. W. Jäger, X. Yunjie and M. C. L. Gerry, *J. Phys. Chem.*, 97 (1993) 3685.
36. E. D. Stevens, *Mol. Phys.*, 37 (1978) 27.
37. E. D. Stevens, personal communication.
38. S. J. Harris, S. E. Novick, J. S. Winn and W. Klemperer, *J. Chem. Phys.*, 61 (1974) 3866-3867.
39. W. B. Almeida, *J. Phys. Chem.*, 97 (1993) 2560. In this article the global minimum of the Cl₂ dimer is reported to be a "near T-shaped structure". However, a closer examination of the presented geometrical parameters shows that it is an L-structure.
40. T. Brinck, (1997) submitted for publication.
41. M. M. Francl, *J. Phys. Chem.*, 89 (1985) 428.
42. I. Alkorta, H. O. Villar and J. J. Perez, *J. Phys. Chem.*, 97 (1993) 9112.
43. G. Dive and D. Dehareng, *Int. J. Quant. Chem.*, 46 (1993) 127.
44. P. Sjöberg, J. S. Murray, T. Brinck and P. Politzer, *Can. J. Chem.*, 68 (1990) 1440.
45. T. Brinck, J. S. Murray, P. Politzer and R. E. Carter, *J. Org. Chem.*, 56 (1991) 2934.
46. T. Brinck, J. S. Murray and P. Politzer, *J. Org. Chem.*, 56 (1991) 5012.
47. J. S. Murray, T. Brinck and P. Politzer, *Int. J. Quant. Chem., Biol. Symp.*, 18 (1991) 91.
48. J. S. Murray, T. Brinck and P. Politzer, *J. Mol. Struct. (Theochem)*, 255 (1992) 271.
49. M. Haeberlein, J. S. Murray, T. Brinck and P. Politzer, *Can. J. Chem.*, 70 (1992) 2209.
50. T. Brinck, *J. Phys. Chem. A*, 101 (1997) 3408.
51. R. S. Drago, G. C. Vogel and T. E. Needham, *J. Am. Chem. Soc.*, 93 (1971) 6014.
52. H. Hagelin, J. S. Murray, T. Brinck, M. Berthelot and P. Politzer, *Can. J. Chem.*, 73 (1995) 483.
53. M. Berthelot, J. F. Gal, C. Helbert, C. Laurence and P. C. Maria, *J. Chim. Phys.*, 82 (1985) 427.
54. M. Berthelot, G. Grabowski and C. Laurence, *Spectrochim. Acta*, 41A (1985) 657.
55. S. G. Lias, J. E. Bartmess, J. F. Liebman, J. L. Holmes, R. D. Levin and W. G. Mallard, *J. Phys. Chem. Ref. Data*, 17 (1988).
56. U. Dinur and A. T. Hagler, in K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry*, VCH Publishers, New York, 1991.

57. P. Kollman, J. McKelvey, A. Johansson and S. Rothenberg, *J. Am. Chem. Soc.*, 97 (1975) 955.
58. M. J. Kamlet, J.-L. M. Abboud and R. W. Taft, *J. Am. Chem. Soc.*, 99 (1977) .
59. M. J. Kamlet, J.-L. M. Abboud, M. H. Abraham and R. W. Taft, *J. Org. Chem.*, 48 (1983) 2877.
60. M. J. Kamlet, R. W. Taft, G. R. Famini and R. M. Doherty, 41 (1987) 589.
61. M. H. Abraham, *Chem. Soc. Revs.*, 22 (1993) 73.
62. M. H. Abraham, P. P. Duce, P. L. Grellier, D. V. Prior, J. J. Morris and P. J. Taylor, *Tetrahedron Lett.*, 29 (1988) 1587.
63. M. H. Abraham, P. L. Grellier, D. V. Prior, P. P. Duce, J. J. Morris and P. J. Taylor, *J. Chem. Soc. Perkin Trans. 2*, (1989) 699.
64. M. H. Abraham, P. L. Grellier, D. V. Prior, J. J. Morris, P. J. Taylor, C. Laurence and M. Berthelot, *Tetrahedron Lett.*, 29 (1989) 2571.
65. M. H. Abraham, P. L. Grellier, D. V. Prior, J. J. Morris and P. J. Taylor, *J. Chem. Soc. Perkin Trans. 2*, (1990) 521.
66. M. H. Abraham, P. L. Grellier, D. V. Prior, R. W. Taft, J. J. Morris, P. J. Taylor, C. Laurence, M. Berthelot, R. M. Doherty, M. J. Kamlet, M. Abboud, K. Sraidi and J. Guiheneuf, *J. Am. Chem. Soc.*, 110 (1988) 8534.
67. R. W. Taft and J. S. Murray, in P. Politzer and J. S. Murray (eds.), *Quantitative Treatments of Solute/Solvent Interactions*, Elsevier, Amsterdam, 1994, Chap. 8.
68. J. S. Murray, S. Ranganathan and P. Politzer, *J. Org. Chem.*, 56 (1991) 3734.
69. P. W. Kenny, *Journal of the Chemical Society, Perkins Transactions 2*, (1994) 199.
70. R. S. Drago and B. B. Wayland, *J. Am. Chem. Soc.*, 87 (1965) 3571.
71. W. Partenheimer, T. D. Epley and R. S. Drago, *J. Am. Chem. Soc.*, 90 (1968) 3886.
72. G. C. Vogel and R. S. Drago, *J. Am. Chem. Soc.*, 92 (1970) 5347.
73. E. M. Arnett, E. J. Mitchell and T. S. S. R. Murty, *J. Am. Chem. Soc.*, 96 (1974) 3875.
74. C. Laurence, G. Guiheneuf and B. Wojtkowiak, *J. Am. Chem. Soc.*, 101 (1979) 4793.
75. P. Nagy, K. Novak and G. Szasz, *J. Mol. Struct. (Theochem)*, 201 (1989) 257.
76. M. Haeberlein and T. Brinck, *J. Phys. Chem.*, 100 (1996) 10116.
77. A. K. Chandra, K. Bhattacharya and T. Brinck, (1997) submitted for publication.
78. H. Umeyama and K. Morokuma, *J. Am. Chem. Soc.*, 98 (1976) 4400.

79. L. P. Hammett, *Physical Organic Chemistry*, McGraw-Hill, New York, 1970.
80. O. Exner, *Correlation Analysis of Chemical Data*, Plenum Press, New York, 1988.
81. T. B. McMahon and P. Kebarle, *J. Am. Chem. Soc.*, 99 (1977) 2222.
82. P. Mulder, O. W. Saastadt and D. Griller, *J. Am. Chem. Soc.*, 110 (1988) 4090.
83. J. Lind, X. Shen, T. E. Eriksen and G. Merényi, *J. Am. Chem. Soc.*, 112 (1990) 479.
84. F. G. Bordwell and J.-P. Cheng, *J. Am. Chem. Soc.*, 113 (1991) 1736.
85. A. A. Zavitsas and J. A. Pinto, *J. Am. Chem. Soc.*, 94 (1972) 7390.
86. J. M. Dust and D. R. Arnold, *J. Am. Chem. Soc.*, 105 (1983) 1221; and references therein.
87. K. B. Clark and D. D. M. Wayner, *J. Am. Chem. Soc.*, 113 (1991) 9363.
88. F. G. Bordwell, X.-M. Zhang, A. V. Satish and J.-P. Cheng, *J. Am. Chem. Soc.*, 116 (1994) 6605.
89. W. M. Nau, H. M. Harrer and W. Adam, *J. Am. Chem. Soc.*, 116 (1994) .
90. Y.-D. Wu, C.-L. Wong and K. W. K. Chan, *J. Org. Chem.*, 61 (1996) 746.
91. T. Brinck, M. Haeberlein and M. Jonsson, *J. Am. Chem. Soc.*, 119 (1997) 4239.
92. L. Pauling, *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, NY, 1960.
93. T. L. Cottrell, *The Strengths of Chemical Bonds*, Butterworth Scientific Publications, London, Great Britain, 1958.
94. P. Politzer, *J. Chem. Phys.*, 50 (1969) 2780.
95. P. Politzer and D. Habibollahzadeh, *J. Chem. Phys.*, 98 (1993) 7659.
96. R. F. W. Bader, *Atoms in Molecules - A Quantum Theory*, Oxford University Press, Oxford, 1990.
97. J. J. M. Wiener, J. S. Murray, M. E. Grice and P. Politzer, *Mol. Phys.*, 90 (1997) 425.
98. D. D. M. Wayner and D. R. Arnold, *Can. J. Chem.*, 62 (1984) 1164.
99. G. Leroy, M. Sana and C. Wilante, *J. Mol. Struct. (Theochem)*, 234 (1991) 303.
100. T. Brinck, J. S. Murray and P. Politzer, *Mol. Phys.*, 76 (1992) 609.
101. P. Politzer, P. Lane, P. Murray and T. Brinck, *J. Phys. Chem.*, 96 (1992) 7938.
102. J. S. Murray, P. Lane, T. Brinck, K. Paulsen, M. E. Grice and P. Politzer, *J. Phys. Chem.*, 97 (1993) 9369.
103. T. Brinck, J. S. Murray and P. Politzer, *J. Org. Chem.*, 58 (1993) 7070.
104. J. S. Murray, T. Brinck and P. Politzer, *J. Phys. Chem.*, 97 (1993) 13807.

105. P. Politzer, J. S. Murray, P. Lane and T. Brinck, *J. Phys. Chem.*, 97 (1993) 729.
106. J. S. Murray, T. Brinck and P. Politzer, *Chem. Phys.*, 204 (1996) 289.
107. P. Politzer, J. S. Murray and P. Flodmark, *J. Phys. Chem.*, 100 (1996) 5538.
108. M. J. Kamlet, R. M. Doherty, M. H. Abraham, Y. Marcus and R. W. Taft, *J. Am. Chem. Soc.*, 92 (1988) 5244.
109. J. S. Murray, T. Brinck, P. Lane, K. Paulsen and P. Politzer, *J. Mol. Struct. (Theochem)*, 307 (1994) 55.
110. R. F. Recker, *The Hydrophobic Fragmental Constant*, Elsevier, New York, 1977.
111. C. Hansch and A. J. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley & Sons, Inc., New York, 1979.
112. R. Mannhold, R. Rekker, C. Sonntag, A. M. ter Laak, K. Dross and E. E. Polymeropoulos, *J. Pharm. Sci.*, 84 (1995) 1410.
113. M. Haeberlein and T. Brinck, *J. Chem. Soc. Perkin Trans. 2*, (1997) 289.
114. C. Hansch, A. Leo and D. Hoekman, *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*, American Chemical Society, Washington, DC, 1995.

This Page Intentionally Left Blank

Exploring Reaction Outcomes Through the Reactivity-Selectivity Principle Estimated by Density Functional Theory Studies

Branko S. Jursic

Department of Chemistry, University of New Orleans, New Orleans
Louisiana 70148, USA

1. INTRODUCTION

The principle of reactivity-selectivity has been used widely in organic chemistry to explain the outcome for many different chemical reactions. A general (empirical) principle that is often referred to is that the more reactive the reagent is, the less selective it is in the reactions, *e.g.*, in the chlorination of alkanes, the rate of hydrogen substitution is tertiary>secondary>primary. If the temperature is raised above 300°C, chlorine becomes more reactive and the substitution now becomes less selective, the rates of the substitution being the same for primary, secondary and tertiary hydrogens. The simplest explanation for the reactivity-selectivity principle is that the more reactive the reagent is, the more likely it will react at every collision. In other words, there would be less discrimination between the various positions. Additionally, the less reactive the reagent is, the more the electronic distribution in the substrate will play a part, thereby resulting in more discrimination at the various positions. This simple picture of reactivity-selectivity is a useful principle in the hands of organic chemists if the information on the electronic distribution of both reactant and substrates is available. With accumulating information on many organic chemical reactions, continuously experimental organic chemists can very successfully put this principle to use. The most reliable qualitative results can be obtained for a particular reaction series when an increase of the reactivity of the reactant will result in a corresponding decrease in selectivity of that species. If applied to another reaction series, the reactivity-selectivity principle might fail. In the majority of cases, the reason for failure is that the electronic structures of both a reaction and a substrate are not known. Little has been done to introduce a theoretical basis of the reactivity-selectivity principle. There are some reports that raised serious questions concerning its validity [1-3]. Comparison of selectivity of a species produced by a number of different methods often allows for similarities and differences in these species. The observation of a reactivity-selectivity relationship for a given reaction series suggests a certain uniformity in the mechanism while a sudden break or the total failure suggests the opposite.

In this part of the book, the usefulness of computational approaches to evaluate reactivity-selectivity with a particular emphasis on using Density Functional Theory (DFT) computational methods [4-7] is focused upon. They have been shown by many to be highly reliable for the computation of the structures and energies of chemical systems. The goal is to demonstrate that computational tools are very useful for exploring organic reaction mechanisms, and if properly used, will produce reliable results.

2. COMPUTATIONAL METHODOLOGY

All the calculations were performed with the GAUSSIAN 94 computational package [8]. Traditional Hartree-Fock (HF) [9-14], Møller-Plesset energy correction (MP2, MP3, MP4) [14-19] and Quadratic Configuration Interactions, Singlet and Doublet substitution with Triples contribution to the energies. [QCISD(T)] [20] *ab initio* methods were used. The Gaussian theoretical models (G1, G2, and G2MP2) [21-24] which are known to produce highly accurate energies, were also used to compute the energies.

Three hybrid DFT methods were applied. Becke's three-parameter functional (B3) [25] was used in combination with LYP [26,27] (B3LYP), P86 [28] (B3P86), and PW91 [29] (B3PW91) correlation functionals. Three exchange DFT functionals: Xa [30], Slater's exchange functional (HFS or S in combination with the correlation functionals) [30] and Becke's 88 (HFB or B in combination with the correlation functionals) [31] were used in combination with the local (non-gradient corrected) functional of Perdew (PL) [32], with the gradient-corrected Perdew 86 (P86) correlation functional [28], with Perdew and Wang's 1991 (PW91) gradient-corrected correlation functional [29], with local spin density correlation (LSD) of Vosko, Wilk and Nusair (VWN) [33], and with the fifth functional of VWN [33].

For all the calculations, the Gaussian type basis sets [3-21G, 6-31G(d), 6-311G(d,p); 6-311G(2d,2p), etc.] were employed. The explanation and abbreviations of the basis sets are as found in the GAUSSIAN manuals [34, 35]. The finding, optimization, and verification of the transition states were performed as explained above.

3. BASICS FOR THE REACTIVITY-SELECTIVITY APPROACH

A central point to every mechanistic study of any chemical reaction is the nature of the transition state. There are not yet any experimental techniques that can directly observe transition state structures, although there are some recent experimental studies that, in the near future, could pave the way for the development of these techniques [36-37]. The chemical reaction properties that are associated with the transition state must be relied upon. Through them, the structural and energy profiles of the transition state structures can be generated.

There are many postulates that are associated with the nature of the transition state structure. Probably the most popular is the Hammond postulate [38]. The postulate is founded on the principle that the interconversion of two states of similar energies on a reaction pathway should cause only small structural changes. In principle, this leads us to conclude that for highly exothermic reactions, there should be a small structural difference between the reactant and corresponding transition state; whereas, for highly endothermic reactions, the rearrangement in the reactant structure toward the transition state structure should be substantial. This is because the transition state structure is closer to the product structure. There is an even more fundamental relationship which treats the whole spectrum of reaction types. In this relationship, the transition state is viewed as gradually changing from reactant-like in highly exothermic reactions to product-like in highly endothermic reactions. Hence, the derivation of the corresponding free energy relationship:

$$\delta G^\ddagger = \alpha \delta G_P + (1-\alpha) \delta G_R \quad (1)$$

The G^\ddagger , G_P , and G_R , are the free energies of the three stationary points on the reaction potential energy surface; δ is an operator which indicates the difference introduced as caused by some perturbation (e.g. a substituent or medium change); α is a factor that determines the position of the transition state in regard to reactants and products. It varies from 0 to 1 and is closer to 1 for an endothermic reaction or zero for exothermic reactions. Equation (1) can be rearranged into the form

$$\delta \Delta G^\ddagger = \alpha \delta \Delta G^\circ \quad (2)$$

which indicates that for a given reaction, a perturbation of the free energy change (ΔG°) will be only partially reflected in the transition state energy (ΔG^\ddagger). This is perfectly demonstrated in Figure 1. In the case presented, the perturbation effects

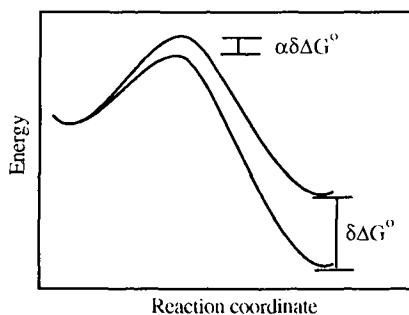


Figure 1. The reaction energy profile for a reaction when perturbation influences products but not reactants.

are only dominant in the product but not in the reactants. By moving along the reaction coordinates from the reactant to the product, the free energy between the two pathways gradually increases toward its maximum value. The transition states energy differences depend on their distance along the reaction coordinate. The relationship presented here is based on models proposed by Polanyi, Bell, Dewar, and Pross [39-43].

Dewar applied a similar approach to explain a series of the same type of reactions that produce products of different stabilities [41]. Before this case is presented, let the reaction of a simple bimolecular substitution, $A + B-C$ and $A-B + C$ be examined. This reaction can be simplified by two separate reactions, those being the B-C bond dissociation reaction and the other the A-B association reaction. These two potential energies are presented in Figure 2. The crossing of these two potential

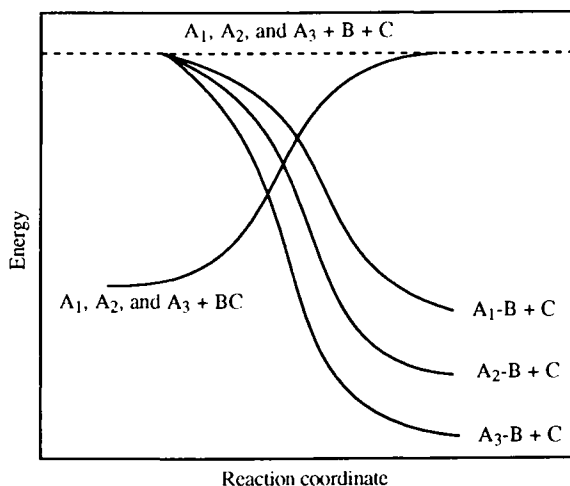


Figure 2. The Bell-Evans-Polanyi treatment for the reaction $A_1, A_2, \text{ and } A_3 + BC \rightarrow A_1-B, A_2-B, \text{ and } A_3-B + C$ with the effect of product stability on the transition state position.

energy surfaces (for B-C dissociation and for A-B association) represents position (a) of the transition state on the potential energy surface, although the transition state energy is substantially lower than the energy of the crossing point. For a series of competitive reactions between substrate B-C with different species (reactants) $A_1, A_2, \text{ and } A_3$ as presented in Figure 2, one can conclude that (a) by stabilizing the product, the reaction barrier will be decreased, (b) by stabilizing the product, the reaction is becoming more exothermic and the transition state structure resembles reactants more and (c) the continuous increase in stabilization of the product brings about a progressively smaller stabilization of the transition state.

In light of this observation, selectivity and reactivity on the basis of the energy changes along reaction coordinate for two or more competitive reactions can now be defined. If the reaction of reactant A with two competitive reagents B and C is being followed experimentally, the rate of the reaction for A + B and the rate for reaction A + C with rate constants k_A and k_B respectively, can be measured. The Selectivity is defined as:

$$S = \log k_A/k_B \quad (3)$$

Since the difference is in the free energy of activation, $\Delta\Delta G^\ddagger$, for two concurrent reactions is $\Delta G^\ddagger_{AB} - \Delta G^\ddagger_{AC}$. And, since there is a linear relationship with k_A/k_B , the selectivity is proportional to $\Delta\Delta G^\ddagger$. This is a very simplified approach to selectivity explanation and it must be noted that many assumptions must be fulfilled for its validity. The fundamental assumption for this conclusion is that the reaction under consideration obeys a rate-equilibrium relationship. For example, the principle cannot be applied for reactions that are diffusion controlled. It is also doubtful that this principle can be applied for reactions that involve very reactive species such as carbenes, radicals, and carbonium ions [1].

In order to acquire a meaningful value that can be used by experimentalists to explain selectivity of the reaction, the definition of reactivity in regard to its absolute value is necessary. This is a very speculative approach because there is no uniform way to define reactivity. In general, the reactivity of a species is determined by its comparison to some standard reagent. Even reactivity obtained in a such way must be taken with precaution, because the other reaction conditions, a such as polarity of a solvent, might also reverse the order of reactivity. A classical example that demonstrates this behavior is the nucleophilic reactivity of halide ions towards some standard substrates. One order of reactivity is obtained in aprotic solvents and other in protic solvents. Furthermore, the nature of the substrate is also a crucial factor. For example, the order of nucleophilic reactivity in methanol with triarylmethyl cation as a substrate (S_N1 mechanism) is $CN^- < CH_3O^- < N_3^-$ while it is opposite for a neutral substrate such as methyl iodide (S_N2 mechanism) [44-45]. This effect was even used to determine the nature of the nucleophilic substitution [46]. There are many examples in which the nature of the substrate, solvent, pH, and salt can strongly effect the order of reactivity. One of the ways an experimentalist tries to solve this problem is through the use of reactive parameters in a quantitative fashion, combined with a stipulation of a reference compound as well as reaction conditions (solvent, temperature, pH,...). Thus, it is not surprising that the most readily available parameters that determine reactivity are rate constants with the standard substrates. This is also not an adequate choice because the rate order might change at different temperatures due to varying activation parameters.

Experimentally, it is not always possible to obtain absolute rate constants, an example being the case in which a reaction is occurring very rapidly. In this case, experimentalists use relative rate constants for determining reactivity. In multistep reactions, the rate constant usually represents the rate for the slowest

reaction step (a rate determining chemical transformation). This is usually the case with solvolysis reactions and, in many cases, the rate determining step of the reaction is the generation of the carbocation. The more stable carbocation is less reactive. While these definitions are generally acceptable, there are some cases where they cannot be applied. For example, both *E*- and *Z*-1-bromo-3,3-dimethyl-1-(4'-methoxyphenyl)-1-butene produce the same vinyl carbocations as intermediates in their solvolysis. The k_E/k_Z ratio is 1640, which corresponds to a difference of 5.2 kcal/mol in favor of the *E* isomer which is more reactive[47]. On the other hand, there is a considerable difference in the reactants' stability.

It should be noted that whatever experimental approach is used for to determine the reactivity, there are inherent limitations due to choosing the standard and approximation for the reaction mechanism. Therefore, one alternative procedure for determining the reactivity should be sought in the computational evaluation of the reactivity through determining the reaction barrier, or for some very reactive reaction pathways, the energy difference between reactants and the reactive intermediate. With today's advanced capabilities of computational methods, the reaction barrier for almost any kind of the reaction can be evaluated with both highly accurate *ab initio* and Density Functional Theory (DFT) methods. As a result, reactivity can be defined as the inverse value of the reaction barrier or inverse value of energy difference between the reactive intermediate and the reactant.

Now imagine a reaction of species A with two reagents B and C. As mentioned previously, experimentally determined selectivity is the logarithm of the ratio of the rate constants for two competitive reactions ($\log k_{A-C}/k_{A-B}$). There are numerous experimental ways to determine selectivity and some of them are presented in Table 1. More information about this experimental evaluation of reactivity and selectivity can be found elsewhere [48 - 51].

Table 1. Some of the experimental ways to determine selectivity of the reaction

Name	Equation	Reactivity	Selectivity
Hammett [48]	$\log k/k_0 = \sigma\rho$	σ	ρ
Winstein-Grunwald [49]	$\log k/k_0 = mY$	Y	m
Swain-Scott [50]	$\log k/k_0 = ns$	n	s
Brønsted [51]	$\log k = -\alpha pK_a + \log G$	pK_a	α

Because computation of the activation barrier for competitive reactions is a straightforward process [52], it is now possible to define selectivity of the reactions through computed activation barriers of the reactions. A classical example of a chemical transformation for which the computed reactivity-selectivity principle can be applied is presented in Figure 3. As mentioned above, the computed reactivity of the reactants is defined as $1/\Delta G^\ddagger$. Of course, the same reactant can have different reactivities depending upon what product would be formed or which reaction path would be chosen. For example, the reactivities of A ($r_{A \rightarrow B}$ and $r_{A \rightarrow C}$) are $1/\Delta G_B^\ddagger$ and $1/\Delta G_C^\ddagger$ for reactions $A \rightarrow B$ and $A \rightarrow C$, respectively.

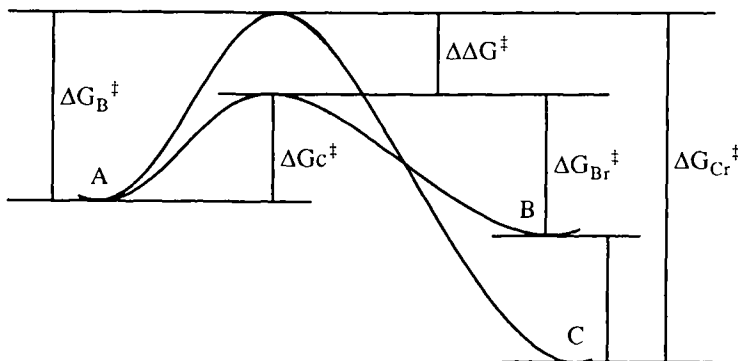


Figure 3. The energy profile for two competitive reactions that can be both kinetically and thermodynamically controlled.

The selectivity of the chemical transformation is necessarily defined as the differences of reactivities for two competitive reactions. For example, selectivity for formation of product B under a kinetically controlled reaction is simply $s = 1/\Delta G_B^\ddagger - 1/\Delta G_C^\ddagger$. On the other hand, selectivity for preparation of product C under thermodynamically controlled reaction conditions (Figure 3) is $1/\Delta G_B^\ddagger - 1/\Delta G_C^\ddagger$. If a reaction is kinetically controlled, that is, experimentally achieved by carrying out the reaction at low temperature, the path of higher reactivity of substrate A will be dominant. The selectivity of the reaction is responsible for the ratio of the products. The higher the difference in the selectivity of the two reaction paths, the more one product will dominate over the other. The inverse reaction (Figure 3) is thermodynamically controlled (high temperature controlled reaction). because the reaction is reversed and there is an interest in keeping the product. The less selective pathway, A-C (actually B \rightarrow A \rightarrow C), will be dominant, keeping C as the major product of the reaction.

Now, the example of the reactivity-selectivity principle computed on the examples for which experimental data exist will be reviewed.

4. THE DIELS-ALDER REACTION

The most important cycloaddition reaction from the point of view of synthesis is the Diels-Alder reaction. The Diels-Alder reaction is the addition of an alkene to a diene to form a cyclohexene. It is called a [4 + 2]-cycloaddition reaction because four π electrons from the diene and two π electrons from the alkene are directly involved in the bonding change. The reaction has been the object of extensive theoretical and mechanistic studies, as well as synthetic application [52, 53]

4.1. Diels-Alder Reaction of Cyclopropene with Butadiene

Wiberg reported the Diels-Alder reaction of butadiene and cyclopropene [53] and Baldwin estimated from the reaction between cyclopropene and 1-deuteriobutadiene at 0°C that 99.4% of the formed cycloadduct was the *endo* isomer [54]. There are many suggestions which attempt to explain *endo* selectivity in Diels-Alder reactions (Alder's rule [55]), but none are firmly established. According to Woodward and Hoffmann [56], the preference is the result of favorable Secondary Orbital Interactions (SOI) or secondary orbital overlap [57-59] between the diene and dienophile in the corresponding transition state structure. One can also find an explanation for the reaction preference in the difference between primary overlap [60], volumes of activation [61], and the polarity of the transition states [62]. Secondary orbital overlap between the diene and the dienophile does not lead to bonds in the adduct, but primary orbital overlaps do.

There is no doubt that the driving force for cyclopropene as the dienophile for a Diels-Alder reaction is the release of angle strain energy in the course of the reaction. This is demonstrated by its relatively low activation barrier. For example, cyclopropene reacts with cyclopentadiene and butadiene at 0°C or at room temperature, producing almost exclusively the *endo* cycloadduct [53]. This addition can be explored by computing activation barriers for two isomeric transition state structures. In this way, nonbonding interactions between diene and dienophile in two isomeric transition state structures can be closely evaluated. The reactivity and selectivity for two concurrent reaction pathways can also be computed.

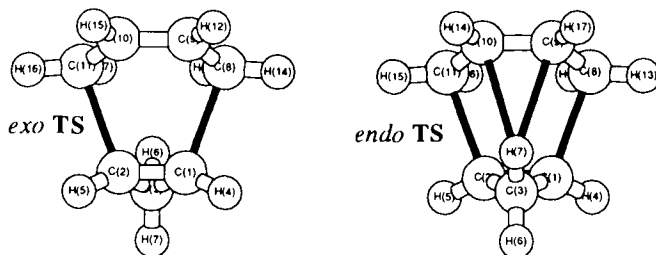


Table 2. Some geometric parameters for the *exo* and *endo* transition state structures of cyclopropene addition to butadiene

Theory Model	$r_{81}/\text{Å}$	$r_{98}/\text{Å}$	$r_{109}/\text{Å}$	$r_{21}/\text{Å}$	$r_{36}/\text{Å}$	$a_{218}/^\circ$	$a_{321}/^\circ$
<i>Exo</i> Transition State Structure							
HF	2.260	1.366	1.407	1.336	1.489	110.0	63.4
B3LYP	2.378	1.370	1.423	1.338	1.500	109.6	63.5
BLYP	2.418	1.381	1.431	1.349	1.513	109.5	63.5
<i>Endo</i> Transition State Structure							
HF	2.259	1.366	1.408	1.335	1.479	109.9	63.2
B3LYP	2.384	1.370	1.424	1.340	1.492	109.5	63.3
BLYP	2.424	1.383	1.433	1.351	1.505	109.4	63.3

r = bond distances; a = bond angles.

Typically, the most significant structural parameters which are subject to change with different theory levels are the newly forming bond distances. It was demonstrated, by using different theory levels, that the transition state structures, for the most part, did not vary substantially [63-65]. This was true when the usual dienes and dienophiles were utilized. Cyclopropene, due to its angle strain, is an unusual dienophile which should require an electron correlational computational method to correctly compute its transition state geometry. This was aptly demonstrated by the bond distance of the bonds in formation. All applied computational methods predicted transition state structures for the concerted synchronous formation of both C-C bonds. As was expected, HF *ab initio* methods produced considerably shorter bond distances. The C-C bond in formation computed by HF was more than 0.1 Å shorter than the one computed by both the B3LYP and BLYP DFT methods. On the other hand, BLYP has been known to produce slightly longer bond distances and, in this respect, it resembles MPn *ab initio* methods [66]. The DFT computed newly forming C-C bond distances (r_{81} , Table 2) for the *endo* transition state structure that were slightly longer than were the ones for the *exo* transition state structure. This indicated that the former structure was also closer to the reactants. By referring to the Hammond postulate [38], one can see that the transition state structure, which was closer in geometry to the reactants, will have a lower activation energy. The product formed through an *endo* transition state structure should be dominant in a kinetically-controlled cycloaddition reaction. As pointed out earlier, Wiberg and Barley observed only an *endo* cycloadduct for the butadiene reaction with cyclopropene [53]. Two concurrent reactions can also be explored through non-bonding interactions in corresponding transition state structures.

Table 3. The Mullikan bond orders (BO) and frontier orbital energies computed for the two isomeric transition state structures of cyclopropene addition to butadiene

	BO ₁₋₈	BO ₁₋₉	BO ₇₋₉	BO ₆₋₉	HOMO	LUMO
<i>Exo</i> Transition State Structures						
A	0.30920	-0.00217	0.00149	0.00287	-0.29742	0.13487
B	0.26910	0.00515	0.00195	0.00360	-0.21929	-0.01634
C	0.27617	0.00223	0.00223	0.00380	-0.18311	-0.03924
<i>Endo</i> Transition State Structures						
A	0.31535	-0.00092	0.01726	0.00057	-0.30223	0.14025
B	0.26906	0.00403	0.02352	0.00094	-0.22331	-0.01221
C	0.27574	0.00479	0.02446	0.00102	-0.18684	-0.03556

A=HF/6-31G(d); B=B3LYP/6-31G(d); C=BLYP/6-31G(d).

One way of determining nonbonding interactions between two chemical systems is by computing bond orders [67-68], as well as by Frontier Molecular Orbital (FMO) [69-71] interactions in the transition state. It is well known that FMO can be used to explain the reactivity of a diene and dienophile for cycloaddition reactions [72-74]. There was no noteworthy difference between the

computed bond orders for the bonds involved in the formation (C1-C8 or C2-C11) of the two isomeric transition state structures with hybrid or gradient-corrected DFT methods. However, there was a noticeable difference in the secondary orbital interactions (SOI) between C1-C9 (C2-C10) of the diene-dienophile π -bonds. If these interactions were of a dominant nature, the *exo* transition state structure would have had the lower energy, which was not the case. Other nonbonding interactions were between the methylene hydrogen of the cyclopropene moiety with the π -orbitals of the butadiene moiety of the transition state structures. This interaction was only present in the *endo* transition state structure. In fact, this secondary molecular orbital overlap was higher than the C1-C9 secondary molecular orbital overlap in the *exo* transition state structure. This properly suggested that the *endo* transition state structure should have had a substantially lower energy than the *exo* transition state structure. The frontier molecular orbitals of the transition state structures also indicated that there were additional stabilization interactions which decreased the frontier orbital energies in the *endo* transition state structure and therefore, made the *endo* transition state lower in energy. The computed activation barriers and reactivity-selectivity parameters were more reliable values for predicting the outcome of the reaction (Table 4). The reaction barriers were computed with several *ab initio* and DFT methods. From previous studies, the QCISD(T) *ab initio* and B3LYP DFT computed values were the ones which proved to be most trustworthy [52]. In fact, they are in many cases identical. The B3LYP/6-

Table 4. The reaction barriers (kcal/mol) for the cyclopropene addition to butadiene computed with *ab initio* and DFT methods using the 6-31G(d) basis set.

Theory	ΔE_{exo}	ΔE_{endo}	r_{exo}	r_{endo}	s
HF	36.3	34.3	0.0275	0.0291	0.0016
HF + ZPEC	38.4	36.5	0.0260	0.0273	0.0013
MP2/HF	6.8	3.9	0.1470	0.2564	0.1094
QCISD(T)/D96V/HF/6-31G(d)	15.8	14.0	0.0632	0.0714	0.0082
B3LYP	15.4	13.5	0.0649	0.0740	0.0091
B3LYP + ZPEC	17.2	15.2	0.0581	0.0657	0.0076
BLYP	14.0	12.3	0.0714	0.0813	0.0099
BLYP + ZPEC	15.6	14.0	0.0641	0.0714	0.0073

ZPEC = zero point energy correction; ΔE_{exo} and ΔE_{endo} = activation barrier for formation of *exo* and *endo* products (kcal/mol), respectively; r_{exo} and r_{endo} = reactivity of the same reactant in *exo* and *endo* reaction, respectively; s = selectivity for formation of the *endo* over *exo* product under kinetically controlled reaction conditions.

31G(d) with ZPEC computed the *exo* reactivity (0.0581) to be smaller than *endo* reactivity (0.0657) and subsequently, *endo* selectivity for the kinetically controlled addition of cyclopropene to butadiene was 0.0076. It was responsible for the

formation of the *endo* cycloadduct as a major product of the reaction which was also determined experimentally.

4. 2. Diels-Alder Reaction of Cyclopropene with Furan

Attention can now be turned to the cycloaddition of the same dienophile (cyclopropene) to furan. Furan is not as good a diene for the Diels-Alder reaction as is butadiene. It is reasonable to expect that the activation energy for this cycloaddition reaction should be slightly higher. As in the case of butadiene with furan, there are two possible pathways that will form *exo* or *endo* cycloadducts. There is no experimental activation barrier for the cyclopropene addition to furan. Binger and co-workers [75] observed a slight *exo* preference in the cyclopropene addition to 1,3-diphenylisobenzofuran; however, previous studies reported that the major product of this cycloaddition reaction was the *endo* cycloadduct [76-79]. From experimental data in the literature, it is obvious that the reaction, which is under a kinetically controlled environment (a low temperature reaction), will produce exclusively, or at least as a major product, the corresponding *exo* cycloadduct. This is contrary to the original cyclopropene addition to butadiene or cyclopentadiene where the exclusive products were the *endo* cycloadducts [53]. On the other hand, Breslow and Oda [80-81] isolated only the *exo* cycloadduct from the reaction between cyclopropenone and 1,3-diphenylisobenzofuran. This was confirmed by determining the crystal structure of the product [82].

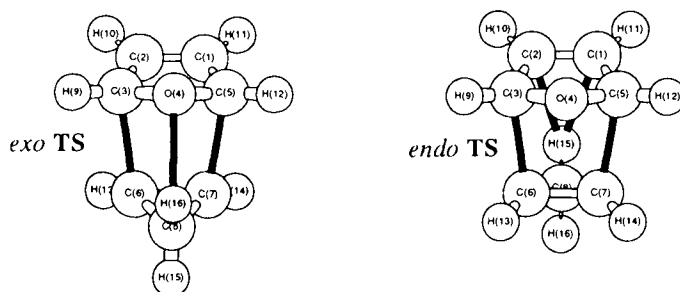


Table 5. Structural parameters for transition state structures for cyclopropene addition to furan computed with *ab initio* and DFT methods by using 6-31+G(d) basis set.

<i>Exo</i> Transition State Structure						
Theory Model	$r_{63}/\text{\AA}$	$r_{62}/\text{\AA}$	$r_{16,4}/\text{\AA}$	$r_{67}/\text{\AA}$	$r_{43}/\text{\AA}$	$a_{763}/^\circ$
HF	2.214	2.741	2.391	1.338	1.343	100.3
B3LYP	2.265	2.781	2.386	1.351	1.366	100.3
BLYP	2.281	2.805	2.274	1.365	1.385	100.3
<i>Endo</i> Transition State Structure						
Theory Model	$r_{63}/\text{\AA}$	$r_{62}/\text{\AA}$	$r_{15,2}/\text{\AA}$	$r_{67}/\text{\AA}$	$r_{43}/\text{\AA}$	$a_{763}/^\circ$
HF	2.193	2.786	2.595	1.340	1.348	100.4
B3LYP	2.234	2.814	2.576	1.354	1.370	100.5
BLYP	2.236	2.836	2.612	1.368	1.389	100.6

The computed geometries for two isomeric transition states are presented in Table 5. All computational methods predicted a synchronous formation of the two new C-C bonds and a concerted mechanism for the cycloaddition reaction. Due to the lack of an electron correlation in the HF ab initio computational approach, the computed bonds were substantially shorter when compared with experimental data or computational data obtained with correlational computational methods. If the two isomeric transition state structures obtained with DFT (B3LYP and BLYP) methods are taken into consideration, then it can be seen that the *exo* transition state structure is the closest structure to the reactants. According to the Hammond postulate [38], the *exo* transition state structure should then have a lower energy than the isomeric *endo* transition state structure.

Here, SOI can again be used to explain reaction preferences for formation of the *exo* cycloadduct. The DFT computational results were in full agreement with Apeloig and Matzner's ab initio calculations [58] which provided evidence for the interactions (p-p SOI) in the transition state structure between the diene and dienophile in the *exo* and *endo* transition state structures (Table 6). According to the previously determined true bond distances, the π - π SOI was dominant in the *exo* transition state structure (BOs were 0.00843 and 0.00663 in the *exo* and *endo* transition state structures, respectively) contrary to the majority of Diels-Alder reactions. This suggested that there might be other SOIs that were of a stronger nature than the π - π SOI. There were two SOIs that stabilized both transition state structures. They were the orbital interaction between the methylene hydrogen of cyclopropene and the lone pair of furan (H-n SOI) and the methylene hydrogen of cyclopropene and the π MO of furan (H- π SOI) in the *exo* and *endo* transition state structures, respectively. The H-n SOI in the *exo* transition state structure was stronger than the H- π SOI in the *endo* transition state structure. For instance, B3LYP computed 0.02282 for the H₁₆-O bond order in the *exo* transition state structure and 0.01717 for the H₁₅-C₂ bond order in the *endo* transition state

Table 6. Computed Mullikan bond orders (BO) and frontier orbital energies for two isomeric transition state structures for cyclopropene addition to furan

<i>Exo</i> Transition State Structures						
Theory Model	BO ₆₋₃	BO ₆₋₂	BO ₁₆₋₄	BO ₁₅₋₄	HOMO	LUMO
HF	0.36231	-0.00508	0.01234	0.00045	-0.30447	0.15431
B3LYP	0.33395	0.00843	0.02282	0.00089	-0.21432	0.00823
BLYP	0.34263	0.01281	0.02619	0.00104	-0.17656	-0.01646
<i>Endo</i> Transition State Structures						
Theory Model	BO ₆₋₃	BO ₆₋₂	BO ₁₅₋₂	BO ₁₆₋₂	HOMO	LUMO
HF	0.37688	-0.00269	0.01139	0.00012	-0.30938	0.15810
B3LYP	0.35389	0.00663	0.01717	0.00033	-0.21884	0.00947
BLYP	0.37017	0.00860	0.01910	0.00039	-0.18023	-0.01606

BO = bond order BO₆₋₃ is the bond order between carbon atoms 6 and 3 in the transition state structures.

structure. This indicated a greater stabilization of the *exo* transition state structure by SOI (Table 6). Furthermore, the B3LYP computed difference between the frontier orbitals in the *exo* transition state structure (0.22255 a.u.) was considerably lower than for the *endo* transition state structure (0.22831 a.u.). These results clearly indicated a stronger secondary orbital overlap in the *exo* transition state structure.

The computed activation barrier for the cyclopropene addition to furan was in full agreement with these observations (Table 7). If the HF computed energy were the activation barrier, then the reaction between furan and cyclopropene would not be possible. On the other hand, MP2 computed that the reaction was occurring almost without any barrier. When considering the experimental procedure conducted with 1,3-diphenylisobenzofuran and cyclopropene, the activation barrier should be around 16 kcal/mol with a slight preference (~0.5 kcal/mol) of the *exo* over the *endo* transition state structure. The B3LYP/6-31+G(d) theory model computed energies which one would expect. The B3LYP computed reactivity index indicated that the *exo* reaction path was slightly preferred (0.0617 over 0.0592). The selectivity index was very small (0.0025) which indicated a slight preference of formation of the *exo* over the *endo* cycloadduct. This observation was confirmed with experimental data [76-79].

Table 7. The reaction barriers (kcal/mol) for cyclopropene addition to furan computed with *ab initio* and DFT methods using the 6-31G(d) basis set.

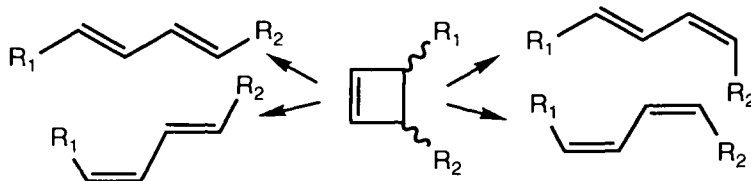
Theory	ΔE_{exo}	ΔE_{endo}	r_{exo}	r_{endo}	s
HF	33.5	34.4	0.0298	0.0291	0.0007
MP2	6.8	7.4	0.1271	0.1251	0.0020
B3LYP	16.2	16.9	0.0617	0.0592	0.0025
BLYP	15.4	16.2	0.0649	0.0617	0.0032

ZPVC = zero point energy correction; ΔE_{exo} and ΔE_{endo} = activation barrier for formation of *exo* and *endo* products (kcal/mol), respectively; r_{exo} and r_{endo} = reactivity of both reactants in *exo* and *endo* reaction, respectively; s = selectivity for formation of the *exo* cycloadduct under kinetically controlled reaction.

Now the possibility that a reaction mixture that contains one equivalent of all three components; cyclopropene, butadiene, and furan can be explored. As mentioned before, it is now possible to form four products (*exo* and *endo* cyclopropene adduct with butadiene [83], and *exo* and *endo* cyclopropene adduct with furan [84]). The reactivity indexes were 0.0740 (*endo* adduct with butadiene), 0.0649 (*exo* adduct with butadiene), 0.0617 (*exo* adduct with furan), and 0.0592 (*endo* adduct with furan). The selectivity index of formation of the *endo* cycloadduct with butadiene over the *exo* cycloadduct with furan was 0.0091 for cyclopropene as the dienophile. If higher selectivity favored the *endo* addition of cyclopropene to butadiene and if reaction were conducted under kinetic control, only the *endo* cyclopropene adduct with butadiene should be isolated as a product. This finding was also confirmed experimentally.

5. RING OPENING REACTIONS [85]

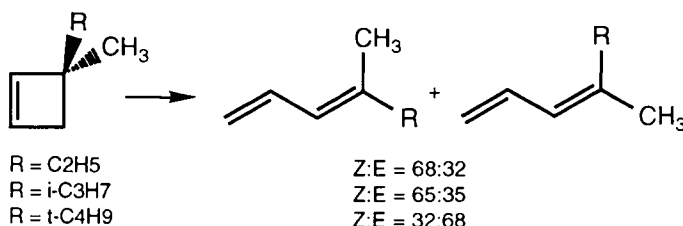
Understanding of the mechanism of a simple transformation like that of cyclobutene to 1,3-butadiene marks several major advances in the understanding of reactivity and selectivity for many other chemical reactions. When 3,4-disubstituted cyclobutene undergoes thermal ring opening to butadiene derivatives, four different products are possible. For example, by studying *cis*-3,4-dimethylcyclobutane, Srinivasan [86] depicted the mechanism as being concerted, involving



participation of the double bond in the breaking of the C₃-C₄ σ -bond with substituents on C₃ and C₄ being required to move either clockwise or counterclockwise with respect to the ring. The conservation of orbital symmetry, the basic concept of the Woodward-Hoffmann theory, provides a clear explanation for the stereochemistry of the electrocyclic cyclobutene-butadiene isomerization [56]. Since the reaction is carried out thermally, the HOMO (highest occupied molecular orbital) of the diene must correlate with an occupied orbital of the product. This approach was actually first applied by Longuet-Higgins and Abrahamson [87] to electrocyclic processes and then it was adapted by Woodward and Hoffmann for pericyclic reactions [56, 88]. The molecules have C₂ symmetry throughout the reaction. Preservation of orbital symmetry going from reactant to the product is only possible through a conrotatory manner. Therefore, the conrotatory motion preserves bonding of all the occupied orbitals. The transition state is aromatic with the Möbius array of four electrons as depicted by Dewar [89] and Zimmerman [90,91]. By contrast, the disrotatory process has the Hückel array of four electrons and is antiaromatic and thermally forbidden.

This explanation is very simple and can be applied to a symmetrically substituted cyclobutene where there is no difference present in the contribution of atomic orbitals in the corresponding frontier molecular orbitals. The situation is not as simple when a monosubstituted and asymmetric disubstituted cyclobutene is present. There is no simple principle that can select one direction over the other in two parallel conrotatory ring openings. Generally, such competitive processes are quite selective. Although there are two possible products of conrotatory 3-methylcyclobutene ring opening, only *trans*-1,3-pentadiene was isolated from the reaction mixture [92]. In practice, outward rotation of substituents with the formation of *trans*-products is selected as the general rule for the ring opening. This rule was rationalized on the basis of a minimization of steric effects in the corresponding transition state structure. Even though there are an abundant

number of examples that certainly support this rule, there are also examples that do not agree with this approach. It was in 1980 that Carry and Stevens [93] published results of thermal 3,3-disubstituted cyclobutane ring opening. The major products of their reactions were



achieved with inward rotation of larger substituents. They were the first to rationalize the substituted butane ring opening in terms other than the steric effect. Later Dolbier and co-workers [94] and Houk and co-workers [95] further developed this premise. Their approach took the previous idea a step further by demonstrating that the electronic effect and not the steric effect, was dominant for both kinetic and thermodynamic control of the 3-substituted cyclobutene ring opening. The mechanism, reactivity, and selectivity of these reactions can be perfectly described by using DFT computational studies.

5. 1. Cyclobutene ring opening

Before engaging in studying the reactivity of different derivatives of cyclobutadiene, the accuracy of DFT methods and DFT/AM1 computational approach for determining the reaction barrier for conrotatory cyclobutene ring opening must first be examined. Although the model compound, cyclobutadiene, was synthesized in 1905 by Willstätter and Schmaedel [96], it was Vogel who first observed the ring cycle equilibrium and had mechanistically investigated this transformation [97]. Goldstein and co-workers [98] determined experimentally the reaction barrier of 32.5 kcal/mol. Since then, the reaction has been extensively studied by both semiempirical [99] and *ab initio* methods [100-102]. The geometric parameters for cyclobutane ring opening generated with various computational methods are presented in Table 8 [103]. These results clearly demonstrated that for carbon and hydrogen organic molecules, all listed computational methods produced almost identical transition state structures.

If the activation barrier for this reaction were examined, it would seem obvious that B3LYP hybrid method produced an energy that was almost identical as the experimental value (Table 9), while HF computed an activation barrier that was too high. Considering the similarity in the transition state structures, it was not surprising that the B3LYP single point energy on AM1 geometry generated an activation barrier that was almost the same as to the full B3LYP computational studies. It also interesting to point out that local (SVWN) DFT method also generated reliable activation barriers for four-membered ring opening. This is not generally the case for many other reactions [52].

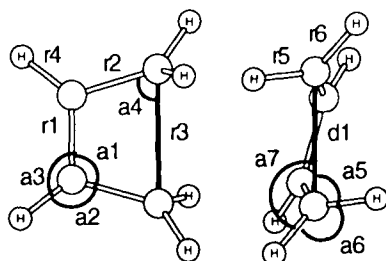


Table 8. The transition state parameters computed with semiempirical (AM1) [104], *ab initio** (HF and MP2), and DFT* (BLYP and SVWN) computational methods

	AM1	RHF	MP2	SVWN	BLYP
r1/Å	1.389	1.367	1.385	1.372	1.380
r2/Å	1.428	1.416	1.431	1.413	1.441
r3/Å	2.120	2.127	2.131	2.126	2.154
r4/Å	1.088	1.077	1.087	1.095	1.093
r5/Å	1.098	1.072	1.084	1.091	1.088
r6/Å	1.096	1.084	1.094	1.102	1.100
a1/Å	103.8	104.2	103.7	104.3	104.5
a2/°	126.6	126.2	126.2	126.0	126.0
a3/°	129.2	129.4	129.9	129.5	129.3
a4/°	74.5	73.6	74.0	73.8	73.7
a5/°	87.5	85.8	83.8	82.9	86.2
a6/°	115.2	113.9	114.6	114.2	113.8
a7/°	127.0	133.0	133.2	133.8	132.3
d1/°	-18.7	-21.8	-22.2	-20.7	-19.7

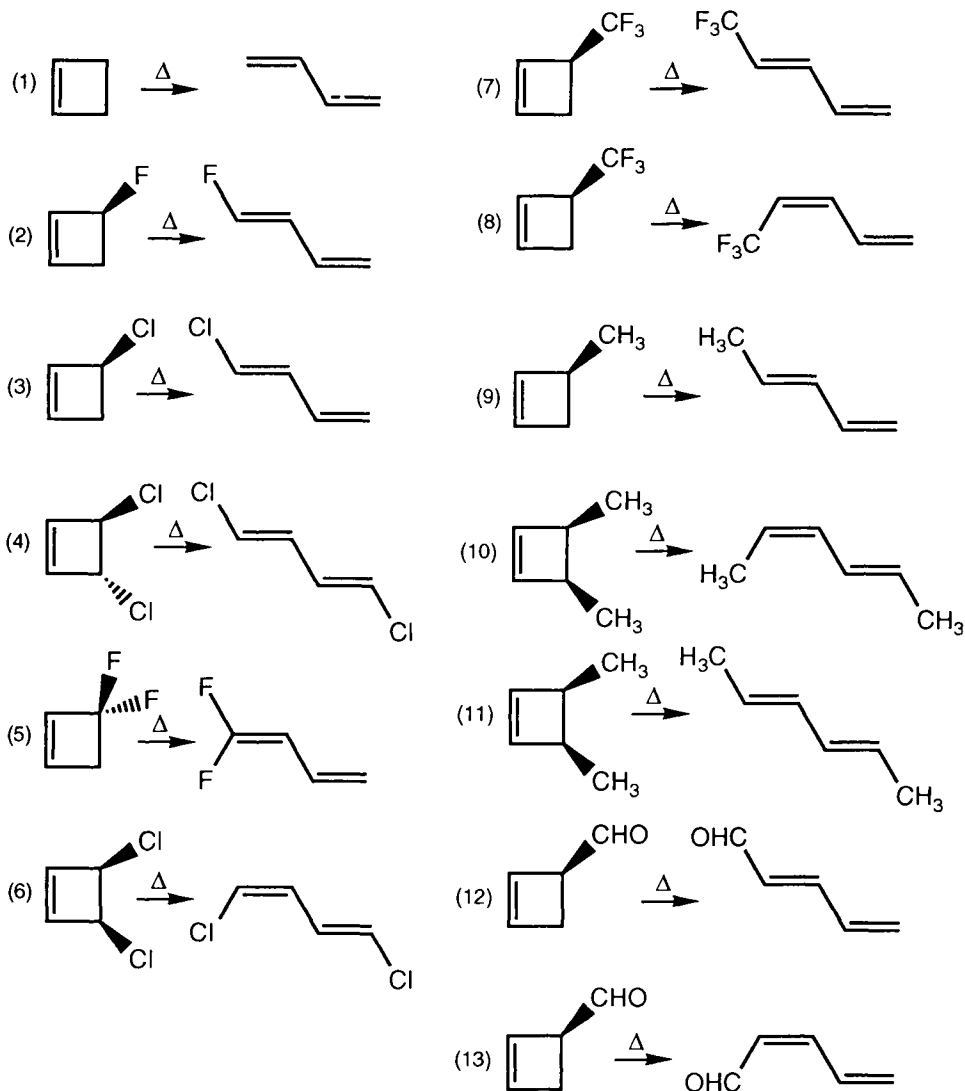
* The calculation was carried out with 6-311+G(d,p) gaussian type of basis set.

Table 9. Computed activation barriers (kcal/mol) for cyclobutene ring opening

Theory Model	without ZPEC	with ZPEC
RHF/6-311+G(d,p)	44.7	42.8
SVWN/6-311+G(d,p)	34.3	32.6
BLYP/6-311+G(d,p)	29.7	27.4
B3LYP/6-311+G(d,p)[52]	33.9	32.2
HF/6-311+G(d,p)/AM1[52]	44.5	
SVWN/6-311+G(d,p)/AM1[52]	34.4	
BLYP/6-311+G(d,p)/AM1[52]	29.4	
B3LYP/6-311+G(d,p)AM1[52]	33.5	
Experimental [98]		32.5

5. 2. Influence of substituents upon the reactivity of cyclobutene ring opening

After establishing both B3LYP and SVWN computational methods as reliable for computing reaction barriers for conrotatory cyclobutene ring opening, the reactivity of substituted cyclobutene with DFT methods [105] can now be investigated. The selected reactions for evaluation of reactivity of cyclobutane are presented in Scheme 1. For every ring opening reaction, a corresponding transition state is available. For many of the reactions, experimental values for reaction



Scheme 1. The reactions that were studied to demonstrate the reactivity-selectivity principle in conrotatory cyclobutene ring opening.

barrier are attainable. The reaction barriers computed with four computational methods are presented in Table 10. As it has been demonstrated previously, HF

Table 10. Activation barriers (kcal/mol) computed by using 6-31G(d) basis set

TS	E _I	E _{II}	E _{III}	E _{IV}	E _V	E _{VI}	E _{VII}	E _{VIII}	E _{IX}
TS1	46.9	45.1	35.6	33.9	31.2	29.6	36.4	34.8	32.5
TS2	42.6	40.8	29.1	27.6	24.2	22.8	27.8	26.5	28.1
TS3	43.7	41.8	30.7	29.0	26.1	24.5	30.4	28.8	29.4
TS4	40.5	38.6	26.3	24.7	21.5	20.1	25.0	23.5	25.7
TS5	57.8	55.8	40.5	38.8	34.4	32.7	38.5	36.9	45.0
TS6	51.8	49.8	36.3	34.5	30.9	29.2	35.1	33.4	35.6
TS7	48.7	46.7	36.2	34.4	31.4	29.6	37.2	35.5	36.3
TS8	51.5	49.7	38.5	36.8	33.7	32.0	38.5	36.8	
TS9	45.1	43.1	33.2	31.3	28.8	26.9	33.7	31.9	
ts10	48.2	46.2	34.9	32.9	30.0	28.1	34.4	32.6	34.3
ts11	82.4	79.0	63.8	60.2	56.8	53.1	61.3	57.6	
ts12	38.7	37.4	26.6	25.4	22.2	21.1	25.9	24.8	27.2
ts13	43.4	41.7	30.7	29.2	25.9	24.4	31.8	30.3	

E_I - computed with HF/6-31G(d) theory model; E_{II} - computed with HF/6-31G(d) + Zero Point Energy Correction (ZPEC); E_{III} - computed with B3LYP/6-31G(d); E_{IV} - computed with B3LYP/6-31G(d) + ZPEC; E_V - computed with BLYP/6-31G(d); E_{VI} - computed with BLYP/6-31G(d) + ZPEC; E_{VII} - computed with SVWN/6-31G(d); E_{VIII} - computed with SVWN/6-31G(d) + ZPEC; E_{IX} - experimental activation barriers.

ab initio computational method was not capable of coming close to the experimental activation energy, although the computed relative reactivity followed the experimentally observed order of reactivity (for all conrotatory ring openings the most reactive was *trans*-3,4-dichlorocyclobutene ring opening, and the least reactive was *cis*-3,4-dimethylcyclobutene ring opening). The closest agreement between computed and experimental activation barriers was obtained with B3LYP and with SVWN DFT (Table 10).

The computed reactivity for substituted cyclobutene ring opening is the next topic of our study. The B3LYP computed reactivity will only be considered, because all other computational methods predicted the same order of reactivity. The most reactive of all studied cyclobutenes was *trans*-3,4-dichlorocyclobutene ($r=0.0405$) [92] and the least reactive was "inward" conrotatory 3-trifluoromethylcyclobutene ring opening ($r=0.0272$). Naturally, the lowest reactivity (0.0166) was computed for thermally forbidden disrotatory *cis*-3,4-dimethylcyclobutene ring opening (Table 1). If only steric effect on the course of the ring opening of *cis*-3,4-dimethylcyclobutene was considered, then disrotatory motions (both methyl groups moves outward) should be preferable to conrotatory motions (one methyl group moves inward). This further supports the concept of electron "control" over the stereochemistry of ring opening.

Table 11. Estimated reactivity of substituted cyclobutene using various computational methods

A	r _I	r _{II}	r _{III}	r _{IV}	r _V	r _{VI}	r _{VII}	r _{VIII}	r _{IX}
(1)	0.0213	0.0222	0.0281	0.0295	0.0321	0.0338	0.0275	0.0287	0.0308
(2)	0.0235	0.0245	0.0344	0.0362	0.0413	0.0439	0.0360	0.0377	0.0356
(3)	0.0229	0.0239	0.0326	0.0344	0.0383	0.0408	0.0329	0.0347	0.0340
(4)	0.0247	0.0259	0.0380	0.0405	0.0465	0.0498	0.0400	0.0426	0.0389
(5)	0.0173	0.0179	0.0246	0.0258	0.0291	0.0306	0.0260	0.0271	0.0222
(6)	0.0193	0.0201	0.0275	0.0290	0.0324	0.0342	0.0285	0.0299	0.0281
(7)	0.0205	0.0214	0.0276	0.0291	0.0318	0.0338	0.0269	0.0282	0.0275
(8)	0.0194	0.0201	0.0260	0.0272	0.0297	0.0313	0.0260	0.0272	
(9)	0.0222	0.0232	0.0301	0.0319	0.0347	0.0372	0.0297	0.0313	
(10)	0.0207	0.0216	0.0287	0.0304	0.0333	0.0356	0.0291	0.0307	0.0292
(11)	0.0121	0.0127	0.0157	0.0166	0.0176	0.0188	0.0163	0.0174	
(12)	0.0258	0.0267	0.0376	0.0394	0.0450	0.0474	0.0386	0.0403	0.0368
(13)	0.0230	0.0240	0.0326	0.0342	0.0386	0.0410	0.0314	0.0330	

A = reactions as denoted in Scheme 1. r_I - computed with HF/6-31G(d) theory model; r_{II} - computed with HF/6-31G(d) + Zero Point Energy Correction (ZPEC); r_{III} - computed with B3LYP/6-31G(d); r_{IV} - computed with B3LYP/6-31G(d) + ZPEC; r_V - computed with BLYP/6-31G(d); r_{VI} - computed with BLYP/6-31G(d) + ZPEC; r_{VII} - computed with SVWN/6-31G(d); r_{VIII} - computed with SVWN/6-31G(d) + ZPEC; r_{IX} - experimental activation barriers.

To further explore the influence of the molecular orbital energies in the transition state structure, the frontier orbital energy gaps changes were computed (Table 12). It was already made apparent that the lower energy gap change between frontier orbitals in the transition state structure for two isomeric reactions suggested that one reaction would progress more easily over the other. This information, in turn, can be used for prediction of selectivity of the reaction. The approach is not reliable when frontier orbital energy gaps of two different molecules are compared. For example, B3LYP/6-31G(d) computed that the lowest frontier orbital energy gap was for the conrotatory 3-fluorobutene ring opening, while both computed and experimental reaction barriers selected *trans*-3,4-dichlorobutene as the most reactive (Table 10). This approach was used to explore the selectivity of *cis*-3,4-dimethylcyclobutene and 3-formylcyclobutene ring opening (Table 12). There were two isomers that could be formed from both reactions. For example, (*Z,E*)-2,4-hexadiene and (*E,E*)-2,4-hexadiene can be formed by the *cis*-3,4-dimethylcyclobutene ring opening through **TS10** and **TS11**, respectively (Reactions 10 and 11, Scheme 1). It was obvious that the path through transition state **TS10** should have had a considerably lower frontier orbital energy change (0.06673) than the reaction through transition state structure **TS11** (0.12545) computed by the B3LYP/6-31G(d) DFT theory model (Table 12). Therefore, the conrotatory ring opening that formed a less stable product and had considerably higher steric interactions in **TS10** would be a more feasible reaction to achieve than would

disrotatory ring opening which formed a more stable product through **TS11** and had lower steric interactions.

Table 12. Frontier orbital energy gap change (a.u.) in the transformation of reactants to corresponding transition state structure

	$\Delta\Delta E_I$	$\Delta\Delta E_{II}$	$\Delta\Delta E_{III}$	$\Delta\Delta E_{IV}$
TS1	0.07123	0.04808	0.04320	0.04353
TS2	0.05369	0.03760	0.03284	0.02882
TS3	0.07154	0.04522	0.03583	0.03720
TS4	0.07904	0.05188	0.03997	0.04169
TS5	0.07803	0.06591	0.06261	0.05468
TS6	0.10469	0.07574	0.06517	0.06146
TS7	0.07793	0.05556	0.05121	0.05134
TS8	0.0825	0.06216	0.05789	0.05294
TS9	0.07459	0.05220	0.04691	0.04173
TS10	0.09079	0.06673	0.06009	0.05665
TS11	0.14724	0.12545	0.12131	0.10219
TS12	0.07477	0.08872	0.03002	0.03021
TS13	0.09701	0.10797	0.04674	0.04826

$\Delta\Delta E_I$ - Computed with HF/6-31G(d) theory model; $\Delta\Delta E_{II}$ - Computed with B3LYP/6-31G(d) theory model; $\Delta\Delta E_{III}$ - Computed with BLYP/6-31G(d) theory model; $\Delta\Delta E_{IV}$ - Computed with SVWN/6-31G(d) theory model.

A second example with which the usefulness of the frontier orbital energy gap change approach to determine stereochemistry of the reaction outcome can be seen is that of 3-formylcyclobutene ring opening. As mentioned above, 3-monosubstituted cyclobutenes usually open thermally and in conrotatory fashion with any substituent going outwards. There are two products that can be formed in this reaction: one which is more stable and goes through the usual outward formyl group motion (reaction 12, Scheme 1), and the other, that forms a less stable product through inward formyl group motion (reaction 13, Scheme 1). The corresponding transition state structures for these isomeric reactions were **TS12** and **TS13**. The frontier orbital energy change again selected the reaction that produced less stable and more sterically demanding products through the transition state structure **TS12** (0.08872 eV 0.10797, Table 12).

It can be observed from the presented results that the electronic factor dominated over the steric factor in determining the reaction outcome (selectivity) as well as reactivity. Chemists have used atomic charges for a long time to determine reactivity as well as selectivity of a reaction. Today, there is a better way to determine that information - through electronic potential surfaces which select "hat spats" in the molecules [106]. Nevertheless, the "old" approach of estimating reactivity through atomic charges in combination with a strong foundation in organic chemistry is still a very viable approach. Every computational method employed, regardless of the theory level, predicted an increase in negative charge on C-3 and C-4 going from reactant to the transition

state structure (Table 13). Considering the fact that the ring opening reaction is an exothermic reaction, according to the Hammond postulate [38], the transition state that experiences the least increase in negative charge on the C-3 substituted carbon center should be the most reactive cyclobutene. Every organic chemist should resolve to put electron withdrawing substituents in the C-3 position of the cyclobutene ring. The computed and experimental activation barriers presented in Table 10 certainly coincided with this observation. Both 3-chloro and 3-fluorobutane had lower activation barriers for ring opening than cyclobutene. On the other hand, 3-methylcyclobutene also had a slightly lower activation barrier than cyclobutene (Table 10). Hence, there should be some interactions other than the simple electron donation or electron withdrawing capabilities of substituents present.

Table 13. The difference in atomic Muliken's charges on two carbons involved in C-C bond breaking for monosubstituted cyclobutene between the corresponding transition state and reactant

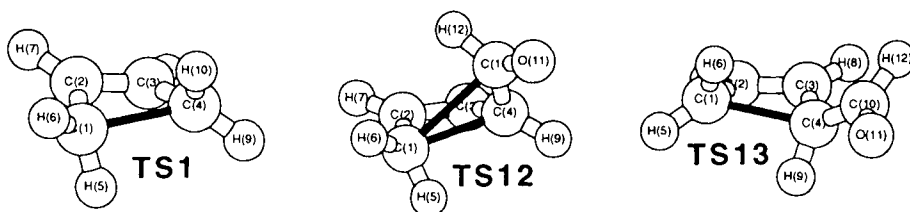
		ΔAC_I	ΔAC_{II}	ΔAC_{III}	ΔAC_{IV}
TS1	A	-0.063953	-0.045882	-0.042818	-0.035175
	B	-0.063953	-0.045882	-0.042818	-0.035175
TS2	A	-0.057089	-0.054644	-0.054295	-0.047497
	B	-0.082267	-0.053146	-0.047857	-0.040892
TS3	A	-0.049161	-0.032928	-0.031481	-0.025248
	B	-0.074873	-0.059661	-0.056431	-0.048619
TS5	A	-0.057110	-0.051858	-0.050386	-0.038257
	B	-0.064851	-0.035345	-0.033232	-0.022813
TS7	A	-0.077442	-0.051660	-0.043402	-0.039739
	B	-0.023390	-0.022926	-0.023349	-0.009391
TS12	A	-0.030535	-0.019645	-0.021217	-0.008089
	B	-0.041837	-0.020605	-0.015474	-0.005764

A - difference of atomic charges for carbon atom (C-3) that has substituents and is involved in bond breaking; B - The difference of atomic charges for carbon atom (C-4) involved in bond breaking; ΔAC_I - Computed with HF/6-31G(d) theory model; ΔAC_{II} - Computed with B3LYP/6-31G(d) theory model; ΔAC_{III} - Computed with BLYP/6-31G(d) theory model; ΔAC_{IV} - Computed with SVWN/6-31G(d) theory model.

The difference of atomic charges on C-3 computed with B3LYP/6-31G(d) theory level (ΔAC_{II} , Table 13) as relative reactivity of a similar series of compounds showed that the most reactive species of the six chosen reactions (Scheme 1) was the ring opening of 3-formylcyclobutene, while the slowest one was for 3-fluorocyclobutene ring opening. While 3-formylcyclobutane was the one that most readily engaged in the ring opening, it was 3-trifluoromethylcyclobutene that proved to be the least reactive in the chosen series of compounds (Table 10 and 13). Both of the substituents showed stabilization of the negative charge, formyl

through resonance, trifluoromethyl through an inductive effect. These findings, in addition to the higher reactivity of 3-methylcyclobutene and cyclobutene, suggested that besides the electronic factor, there must be some extra interactions that stabilized the transition state in the 3-substituted cyclobutene ring opening.

It was demonstrated that Secondary Orbital Interactions (SOIs) were responsible for the reaction outcome for the cycloaddition reaction between cyclopropene and furan and cyclopropene and butadiene. To reiterate the importance of SOI in reference to reactivity, the reaction outcome of 3-formylcyclobutene opening through bond orders in corresponding transition state structures (Table 14) will be taken into account. For simplicity, only the Löwdin bond orders for **TS1**, **TS12**, and **TS13** will be compared. From C(1)-C(2) and C(1)-(4), Löwdin bond orders showed that the transition state for cyclobutene ring opening (**TS1**) was around 64% more advanced in the direction of the formation of product while it was 55% for **TS12** and 59% for **TS13** (Table 14). According to the Hammond postulate, **TS12** was closer to the reactant than isomeric **TS13**. Therefore, the isomer formed from the former should have been the dominant product of the reaction. Secondary orbital interactions were between C(1) of the cyclobutene ring and C(10) of the formyl carbon. These low energy interactions are responsible for selectivity of many organic reactions. It was evident that these interactions were stronger (0.09719) in **TS12** than in **TS13** (0.03159). So, although steric interactions were certainly higher in **TS12**, secondary orbital interactions compensate for these interactions and favor **TS12** over **TS13**, thereby increasing the selectivity of the ring opening.



Tables 14. Bond orders computed on fully B3LYP/6-31G(d) optimized structures of reactants and transition states.

Bonds	TS1A	TS1B	TS12A	TS12B	TS13A	TS13B
C(1)-C(2)	1.35145	1.45132	1.28633	1.38437	1.29560	1.39993
C(2)-C(3)	1.59193	1.65112	1.62445	1.68752	1.62431	1.69545
C(3)-C(4)	1.35147	1.45170	1.21812	1.32737	1.22773	1.31769
C(1)-C(4)	0.63821	0.64042	0.56829	0.54706	0.59953	0.59106
C(1)-C(10)			0.05549	0.09719	0.02082	0.03159
C(1)-C(11)			0.06543	0.07367	0.06375	0.06585

A = Mulliken Bond Order; B = Löwdin Bond Order.

The computed reactivities for the substituted cyclobutenes (Table 11) can now be reinvestigated. To ensure simplicity, this principle will be explored using results computed with B3LYP/6-31G(d), although a similar conclusion could be reached *via*

other computational methods. There are three pairs of reactions that will demonstrate the reactivity-selectivity principle (7-8,10-11, and 12-13). Computed reactivity for reactions (7) and (8) were 0.0291 and 0.0272, respectively (Table 11). The corresponding selectivity factor of $s=0.019$ seemed to be sufficient to produce the *E* isomer through reaction (7) (Scheme 1) in 95 % [95]. Reactivities for conrotatory and disrotatory *cis*-dimethylcyclobutene ring opening (reactions 10 and 11, Scheme 1) were 0.0304 and 0.0166, respectively. The selectivity of the reaction was very high ($s=0.0138$) indicating an exclusive formation of product through reaction (10). For the third pair of reactions, the B3LYP computed selectivity was 0.0052 which insured 100% formation of the less stable *Z* isomer [107]. Selectivity factors of 0.0025 or 0.25% should be sufficient for formation of a single product. With a lower selectivity for such reactions (7), only a small percentage of the other product should be formed.

The principle presented here on two pericyclic reactions (Diels-Alder cycloaddition and cyclobutene ring opening) can be successfully applied to studies of the reaction outcome for many other pericyclic reactions.

6. RADICAL REACTIONS

The salient feature of the homolytic process is the presence of the radical intermediate. The lack of charge of radicals and the highest reactivity of nearly all of those that become involved in a typical organic reaction lead to important differences between homolytic and heterolytic processes. In many cases the odd electron is associated with extremely high reactivity and therefore they occurs as a reactive intermediate in many chemical transformations. In fact, a radical reaction is only practical way to functionalized saturated hydrocarbons with low reactivity. Therefore there is considerable interest to predict selectivity in those radical reactions.

6.1 Trichloromethyl radical proton abstraction reaction

Numerous references are available regarding the relative reactivity of different kinds of hydrocarbons toward free radicals and the variations of behavior observed among different abstracting radicals [108-112]. It is common knowledge that for alkanes, hydrogen becomes more easily abstracted by proceeding along the path of primary, secondary and tertiary. The magnitudes of the differences depend on the radical removing the hydrogen; the more reactive ones are less selective and those less reactive are more selective. Radical selectivity can be modified by the conditions; aromatic solvents, for example, apparently complex with the chlorine radical and increase its selectivity [113]. The magnitude of differences on the radical removing the hydrogen is contingent on the nature of the abstracting radical reagent; the more reactive ones (F and Cl radicals) are also less selective than those of lower reactivity (for example the Br radical). The trichloromethyl radical has the highest selectivity for the hydrogen abstraction reaction [114-115].

There is a considerable amount of interest in computational chemistry involving free radical behavior [116-117]. With this in mind, B3LYP/6-31G(d) computational studies of trimethylradical hydrogen abstraction reaction from methane, ethane, propane, and 2-methyl propane are presented. The geometries of transition state structure for those isolated reactions are presented in Table 15. The transition state structures were quite similar for all four reactions. The major difference was observed for C-H bond breaking and H-C bond making (Table 1). Going from methane to 2-methylpropane, the C-H bond distance of bond breaking was continuously shorter. On the contrary, the C-H bond distance of the bond in formation was continuously longer for the transition state, proceeding from methane to 2-methylpropane. These two structural changes indicated that the transition state between methane and the trimethyl radical was closest to the product while the transition state between 2-methylpropane and trichloromethyl radical was closest to the reactant. Knowing that hydrogen abstraction reactions are exothermic reactions, according to Hammond's postulate, the transition state that is closest to the reactants should have the lowest activation barrier. Combining the structural characteristic of the transition state structures and the Hammond postulate, the reactivity of the alkanes in hydrogen abstraction reactions with radicals (in this case toward trichloromethyl radical) was 2-methylpropane, propane, ethane, and methane or tertiary, secondary, primary, and methyl radical.

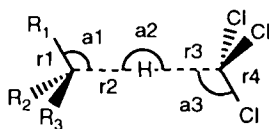


Table 15. The B3LYP/6-31G(d) computed parameters for transition state structures of hydrogen abstraction with the trichloromethyl radical [118]

Substituents	r1/Å	r2/Å	r3/Å	r4/Å	a1/°	a2/°	a3/°
R ₁ =R ₂ =R ₃ =H	1.088	1.407	1.287	1.778	103.2	180.0	106.6
R ₁ =CH ₃ , R ₂ =R ₃ =H	1.508	1.371	1.325	1.779	107.7	180.2	106.5
R ₁ =R ₂ =CH ₃ ; R ₃ =H	1.513	1.344	1.359	1.781	105.4	179.2	106.8
R ₁ =R ₂ =R ₃ =CH ₃	1.519	1.327	1.386	1.780	103.7	180.0	106.8

This information is, of course, only of a qualitative nature. To obtain a better picture of alkane reactivity in radical abstraction reactions, the activation barrier was computed for the reaction between alkanes and radical reactant. The example used was the reaction between 2-methylpropane, propane, ethane, and methane as alkanes and the trichloromethyl radical as a radical reactant (Table 16). The B3LYP computed activation barriers were not corrected of zero point energy, which is usually 1-2 kcal/mol. With this correction computed, experimental [119] values should be in excellent agreement. As expected, 2-methylpropane was the most susceptible in the hydrogen radical abstraction reaction. With the activation barrier around 8 kcal/mol, it was possible to perform the reaction at a

Table 16. The B3LYP/6-31G(d) computed activation barriers, reactivity, and selectivity for selected trichloromethyl hydrogen abstraction reactions

Reaction	$\Delta E_{\text{comp.}}$	ΔE_{exp}	$r_{\text{comp.}}$	$s_{\text{comp.}}$	$s_{\text{corr.}}$
$\text{CH}_4 + \text{CCl}_3$	19.3	17.9 ± 1.2	0.0515	0.0000	0.0000
$\text{CH}_3\text{CH}_3 + \text{CCl}_3$	15.7	14.2 ± 0.6	0.0637	0.0122	0.0183
$(\text{CH}_3)_2\text{CH}_2 + \text{CCl}_3$	13.1	10.6 ± 1.6	0.0763	0.0126	0.0042
$(\text{CH}_3)_3\text{CH} + \text{CCl}_3$	11.0	7.7 ± 0.5	0.0909	0.0146	0.0073

lower temperature with very high selectivity (selectivity index of 0.0335 in regard to abstraction of the hydrogen attached to the secondary carbon atom). It should be noted that in considering reactivity-selectivity indexes of different kinds of hydrogens, a statistical correction is necessary. For example, in the reaction of the chlorine radical with propane, if primary and secondary hydrogens were of equal reactivity, 1-chloropropane and 2-chloropropane would have formed in a ratio of 3:1 because there were six methyl protons and only two methylene protons. The ratios of the products formed at 25°C were 1.41:1, indicating relative reactivity of the primary towards the secondary hydrogen of 1:2.13 [120]. Therefore, selectivity presented in Table 16 should be corrected by number of protons (s_{corr} , Table 16). As a result, the computed values suggested a higher selectivity of the hydrogen abstraction reaction. For example, the corrected selectivity index for tertiary hydrogen abstraction in regard to tertiary hydrogen abstraction reaction with trichloromethyl radical was 0.0073 (Table 16) If the reaction were carried out at a low temperature (due to high reactivity), only one product should have been formed [119].

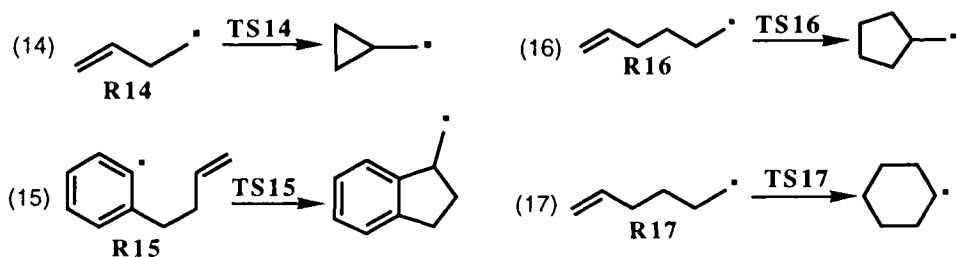
6.2 Intramolecular radical addition to carbon-carbon double bond

Addition of a radical to a carbon-carbon double bond in the same molecule occurs easily if a five or six-membered ring can be formed. Formation of the five-membered ring is faster than formation of the six-membered ring, but the six-membered ring is thermodynamically more stable. In the presence of good hydrogen donor solvents, and if the original radical center is not stabilized, the addition does not reverse and the reaction is kinetically controlled. The more rapidly formed five-membered ring is trapped by hydrogen abstraction. On the other hand, if the original radical is well stabilized and good hydrogen donors are absent, the cyclization is reversible; then the reaction is thermodynamically controlled. The equilibrium will favor the six-membered ring radical, and the cyclohexane product will dominate [121]. Ring formation by intramolecular addition to a double bond has proven to be a versatile synthetic technique. The reaction is usually highly regio- and stereoselective, and other types of functional groups may be present in the molecule without having to be protected [122-123].

The cause for the regio- and stereoselectivity has been traced to stereoelectronic effects. In order for a bonding interaction to occur, the radical center must interact with the π^* orbital of the alkene (LUMO). According to semiempirical and *ab initio* calculations, the preferred direction of attack was from an angle of about 70° with respect to the plane of the double bond [124-126]. The

obtained results showed an accurate qualitative trend for the radical cyclization. Much better agreement between computed and experimental data was obtained by B3LYP hybrid DFT computational studies [105].

The reactions studied with B3LYP/6-31G(d) theory model are presented in Scheme 2. This reaction has been thoroughly studied experimentally. The reaction outcome, as well as reaction barriers for these transformations, are easily accessed [127-129]. Let it first be shown how well the DFT computed reaction



Scheme 2. Radical ring closure reactions studied with B3LYP/6-31G(d) method

barriers agreed with experimentally determined values (Table 17). There were some deviations from the experimental values; nevertheless, the computed values were very close to the experimental values. What is more important is that the computed energy followed the experimental reactivity trend. It clear to see that by comparing any available energies, the most reactive was the cyclization of the phenyl radical (reaction 15, Scheme 2). The B3LYP computed activation barrier was 2.9 kcal/mol which happened to be in excellent agreement with the experimental activation barrier of 3.6 kcal/mol [129]. The calculated B3LYP reactivity index was 0.3448, which was very high. This was not so surprising because the phenyl radical has a higher energy than both secondary and primary radicals due to negligible interactions with neighboring orbitals. On the other hand, the strained transition state structure for reaction 14 slightly increased the reaction barrier in comparison with reaction 16.

The most interesting studies included the evaluation of the reactivity of the same radicals for two different reaction paths, 16 and 17. The computational studies correctly selected path 16 as the most reactive one. However, the question remains: is there sufficient selectivity to accomplish only transformation 16. The computed selectivity index was 0.0336 (0.1216-0.0980). These values assured the formation of products only through a five-membered ring formation. This was in full agreement with experimental results [128]. It is well known that a secondary radical is more stable than a primary radical. The calculations supported this by favoring the cyclohexyl radical by 7.7 kcal/mol over the cyclopentylmethyl radical. Therefore, formation of the six-membered ring product was thermodynamically controlled.

Many experimental chemists use model transition state structures to predict the outcome of organic reactions. If, in the transition state, a six-membered ring is formed, then there are two possible transition state structures, one that is chair-

like and the other that is boat-like. Generally, the chair-like transition state structure will have a lower energy and the products are usually formed through this transition state structure as predicted by our calculations. But sometimes, there are nonbonding interactions between substituents that might stabilize the boat-like transition state structure. As demonstrated above, these interactions can be predicted by DFT and *ab initio* calculations.

Table 17. Computed and experimental activation barriers (kcal/mol) for radical olephin cyclization computed with B3LYP/6-31G(d) hybrid DFT theory model

Reaction	ΔE_I	ΔE_{II}	$\Delta E_{exp.}$	r_I	r_{II}	$r_{exp.}$
(14)	10.6	10.7	7.6	0.0943	0.0934	0.1316
(15)	3.3	2.9	3.7	0.3030	0.3448	0.2703
(16)	6.9	7.6	6.1	0.1449	0.1316	0.1639
(17) ^a chair	9.3	10.2		0.1075	0.0980	
(17) ^b boat	11.8	12.6		0.0847	0.0794	

^a Through chair-like transition state; ^bthrough boat-like transition state; ΔE_I = reaction barrier without zero point energy correction, ΔE_{II} = reaction barrier with zero point energy correction; $\Delta E_{exp.}$ = experimental activation barriers; r_I = reactivity calculated from ΔE_I ; r_{II} = reactivity calculated from ΔE_{II} ; $r_{exp.}$ = reactivity calculated from $\Delta E_{exp.}$

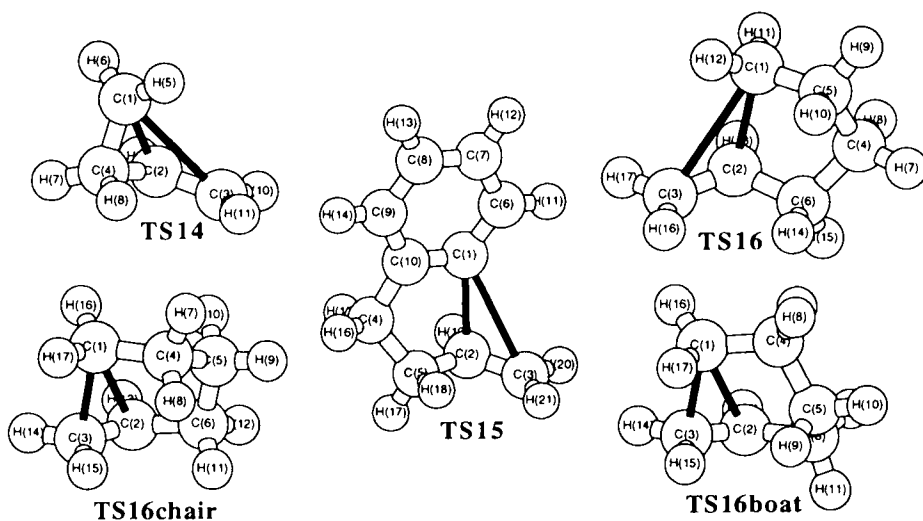
Nonbonding interactions in the transition states for radical cyclizations can now be brought to attention. The computed spin distribution and difference of the frontier orbital energies for transition state structures of reactions 14-17 are presented in Table 18. There are two groups of transition state structures that must be studied separately. They include one that forms primary radicals (**TS14**, **TS15**, and **TS16**) and other that forms secondary radicals (**TS17chair** and **TS17boat**). If spin distribution in the transition state structure is used as a measure of transition state structure on the potential energy surface, then **TS15** with higher spin density (SDr) would be closest to the starting radical, and **TS14** closest to the product radical. In accordance with the Hammond postulate, the most reactive was **R15** and the least reactive was **R14**. In the other group, **TS17chair** had a slightly higher spin density on the carbon that was the radical center in the reactant in comparison to **TSchair**. The path through the chair transition state structure should be preferable, as was determined by the computation of the activation barriers. Increment of the frontier orbital difference going from reactant to the transition state structure as measure of the reactivity was also applied. The smallest increment change ($\Delta\Delta FOE$) in the transition state in regard to the reactants means that the transition state was closer to the reactant on the reaction potential energy surface and the Hammond postulate deemed the reaction as being more feasible. The same observation was made with this approach as was made with spin distribution. **TS15** was the closest one to the reactants. It was then, necessarily, the most reactive of the three transition state structures of the first group. In the second group, transformation of radical **R14** to

the cyclohexene radical was facilitated *via* the chair-like transition state, **TS17chair**, because a smaller frontier orbital energy change was necessary (Table 18).

Table 18. Total spin density (SD) and frontier orbital energies for transition state structures of reactions presented in Scheme 2

species	SDr	SDp	SOMO	LUMO	Δ FOE	$\Delta\Delta$ FOE
TS14	0.6489	0.6693	-0.1600	-0.0414	0.1185	0.0404
TS15	0.7982	0.4227	-0.1891	-0.0619	0.1271	0.0331
TS16	0.7740	0.5396	-0.1654	-0.0426	0.1227	0.0372
TS17chair	0.8393	0.4284	-0.1557	-0.0323	0.1234	0.0365
TS17boat	0.8253	0.4580	-0.1552	-0.0327	0.1225	0.0375

SDr = spin density on the carbon that was the radical center in the reactant; SDp = spin density on carbon that will be the radical center in the product; Δ FOE frontier orbital difference between LUMO and SOMO of transition state structure; $\Delta\Delta$ FOE is the energy difference between Δ FOEs of transition state and corresponding reactant.



Tables 19. Bond orders computed on fully B3LYP/6-31G(d) optimized transition state structures.

Bonds	TS14	TS15	TS16	TS17chair	TS17boat
A-C(1)-C(2)	0.44325	0.19710	0.30707	-0.00988	-0.00898
A-C(1)-C(3)	-0.00511	-0.00120	-0.00526	0.24681	0.26335
B-C(1)-C(2)	0.51286	0.21587	0.32224	0.02473	0.02377
B-C(1)-C(3)	0.03445	0.02451	0.02410	0.27852	0.29842

A = Mulliken Bond Order; B = Löwdin Bond Order.

Calculated bond orders that determined interactions of reactive places in corresponding transition state structures are presented in Table 19. These results demonstrated the ways that computational chemists have to determine the reactivity of chemical structures in chemical reactions. Here, the same conclusions that have already been made on the basis of reaction barriers, spin distribution and change in frontier orbital energy difference were estimated. The smallest bond order for the new forming bond of C(1)-C(2) of the first group of transition state structures was **TS15**, indicating its closeness on the potential energy surface to radical **R15**. According to Löwdin's bond order, all of these reactions have very early transition state structures which is a sign of low activation barriers. This observation is absolutely correct if the fact that the highest reaction barrier in this group was only 7.6 kcal/mol is known. Because the C(1)-C(2) bond order was the lowest in **TS15**, the most readily engaged radical in the radical cyclization should be **R15** (Scheme 2). Of the two isomeric transition state structures, the chair transition state **TS17chair** has lower C(1)-C(3) bond order, making a reaction through it more easily accomplished.

7. REACTIVITY AND STABILITY OF CARBOCATIONS

There are many organic reactions that are widely used in preparation of desirable organic compounds that include formation of carbocations [130]. A critical step in these reactions, aside from nucleophilic and electrophilic substitution and electrophilic addition reactions, is the generation of the tricoordinated carbocation intermediate. For a mechanism to operate, it is essential that this species does not achieve an unreachably high energy. Carbocations are inherently high-energy species. For instance, the ionization of 2-chloro-3-methylpropane is endothermic by 153 kcal/mol in the gas phase [131]. A reaction with an activation energy of this magnitude would have an unobservable, slow rate of reaction at room temperature. It is known that the stability of carbocations greatly increases with the solvent; therefore, one should exclude the usefulness of computational results obtained in the gas phase for determination of the reactivity of the carbocations in everyday synthetic chemistry laboratories. Certainly computational modeling in solvent media should be preferable. Today DFT methods are showing very encouraging results in this direction [132], but carefully gathered results based on studies of carbocations in the gas phase can give us very useful results of their reactivity as well.

7.1 Hydride affinity as measure of carbocation reactivity

One of the most important yet common trends in organic chemistry is the increase in carbocation stability with additional alkyl substituents. This stability relationship is fundamental to understanding many aspects of reactivity, especially of nucleophilic substitution. In recent years, it has become possible to put the stabilization effect on a quantitative basis. One approach incorporates gas phase measurements, which determine the proton affinity of alkenes leading to

carbocation formation. From these data, the hydride affinity of the carbocation can be obtained. There are experimental data available for many carbocations. Thus, they represent a perfect choice to show how reliable DFT methods are for computation of carbocation hydride affinity.



Table 20. The B3LYP/6-31+G(d) computed hydride affinities (eV) for some carbocations

Species	E_c	E_{nm}	HA	HA_{corr}	HA_{exp}
Hydride anion	-0.461815				
Methyl cation	-39.480389	-40.520608	15.7	13.2	13.6
Ethyl cation	-78.856103	-79.833584	14.0	11.8	11.9
2-Propyl cation	-118.212607	-119.148439	12.9	10.8	10.7
2-Methyl-2-propyl cation	-157.555660	-158.464119	12.2	10.2	10.0
Allyl cation	-116.973489	-117.913921	13.0	10.9	11.1
3-Penten-2-yl cation	-195.659246	-196.548186	11.6	9.7	9.8
2-Methyl-3-butene-2-yl cation	-195.650979	-196.543890	11.7	9.8	9.8
Benzyl cation	-270.666651	-271.577272	12.2	10.2	10.1
1-Phenylethyl cation	-310.001476	-310.889986	11.6	9.7	9.8
2-Phenyl-2-propyl cation	-349.327257	-350.205631	11.3	9.5	9.5
Vinyl cation	-77.586838	-78.593267	14.8	12.4	12.4
Phenyl cation	-231.266401	-232.258927	14.4	12.1	12.9
Cyclopropenyl cation	-115.734475	-116.624840	11.7	9.8	9.6
Cyclopentadienyl cation	-193.154578	-194.110325	13.4	11.2	11.1
Cycloheptatrienyl cation	-270.682615	-271.512791	10.0	8.4	8.7

E_c = total energy for carbocation or hydrogen hydride (a.u.); E_{nm} = total energy for corresponding neutral hydrocarbon (a.u.); HA = hydride affinity (eV); HA_{corr} = 0.84HA; HA_{exp} = experimental hydride affinity (eV).

Comparison of the B3LYP/6-31+G(d) computed carbocations hydride affinities with experimental values [133-135] suggested that the computed values were on average 2 eV higher. Certainly very accurate hydride affinities can be obtained by multi correlational *ab initio* computational methods such as Petersson's quadratic complete basis set (CBS-Q) [136] computational approach (Table 21) [137]. Similar agreement between the B3LYP/6-31+G(d) computed and experimental values can be obtained by using a scaling factor (0.84). This is a common approach in computational chemistry as can be seen in the example of the use of a scaling factor for harmonic frequencies [138, 139].

Here, reactivity indexes cannot be computed in the way defined previously (as a reciprocal value of the activation barrier for certain reactions), but the relative reactivity of carbocations or reactions that have those carbocations as intermediates can be determined. The carbocations are a very reactive species

Table 21. Hydride affinities (eV) computed by Quadratic Complete Basis Set (CBS-Q) *ab initio* computational approach

Species	E_c	E_{nm}	HA	HA_{ex}
Hydrogen hydride	-0.519356			
Methyl cation	-39.384101	-40.405771	13.7	13.6
Ethyl cation	-78.671813	-79.629734	11.9	11.9
Vinyl cation	-77.423248	-78.415678	12.9	12.4
Cyclopropenyl cation	-115.498956	-116.381854	9.9	9.8
Allyl cation	-116.707910	-117.645037	11.4	11.1

E_c = total energy (a.u.); E_{nm} = total energy for corresponding neutral hydrocarbon (a.u.); HA = hydride affinity (eV); HA_{exp} = experimental hydride affinity (eV).

that occupy high energy local minima in many chemical reactions. As a result, they are much closer in energy to the rate determining transition state structure than to both reactants or products of the reactions. It is a valid approximation to say that the reactivity of the reactant will follow the stability order of the corresponding carbocations that are associated with their hydride affinities. According to this approach, a carbocation with a higher hydride affinity is more reactive and consequently less stable. If the hydride affinities in Table 20 are examined, then it is apparent that the order of carbocation stability is methyl<ethyl<2-propyl<2-methyl-2-propyl.

It is known that for S_N1 reaction (substitution-nucleophilic-unimolecular) [140-141] a key intermediate is a carbocation, therefore the more reactive substrate will be the one that can produce the most stable carbocation. The reactivity of methyl, ethyl, 2-propyl, and 2-methyl-2-propyl tosylates under S_N1 reaction conditions is inversely proportional to the calculated hydride affinity of the corresponding carbocations. The calculated values were in agreement with the experimental findings which were obtained through solvolysis rate measurement of these tosylates under S_N1 conditions [142, 143]. Correlation of the cation stability-hydride and affinity-solvolytic rate of the reaction under S_N1 reaction conditions was observed for the allyl cation (allyl, 3-penten-2-yl, and 2-methyl-3-butene-2-yl cations)[144] and the benzyl cation (benzyl, 1-phenylethyl, and 2-phenyl-2-propyl cations) [145] series. The most reactive substrates were the ones that formed the carbocations with the lowest hydride affinity.

Aromaticity [146] as a companion property of the reactivity of molecules can also be viewed through the hydride affinity of the carbocation. If a cyclic carbocation has a very low hydride affinity it should be aromatic and, subsequently, a chemical reaction that goes through this carbocation should have an aromatic-like transition state structure [147,148]. Consequently, reactivity of the reactant should be very high. There are four carbocations that can be examined using hydrogen affinities. Both vinyl and phenyl cations have carbons with positive charges involved in the C-C double bond and an "empty" π orbital perpendicular to the π orbitals of C-C double bond. The cation stabilization through molecular orbital overlap is not possible. Due to a positive inductive effect of π bonds, these two carbocations are destabilized, their hydrogen affinities being

higher than, for instance, for normal cations such as ethyl or propyl cations. This was observed experimentally as well as determined by computation of B3LYP/6-31+G(d) hydrogen affinities. By a simple examination of the structures of cyclopropyl, cyclopentadienyl, and cycloheptatrienyl cations, one might, by using the $4n + 2$ and $4n$ electron Hückel rules, determine that cyclopropyl and cycloheptatrienyl cations are aromatic while the cyclopentadienyl cation is antiaromatic. Naturally, this approach cannot give the answer about the difference in the stability or reactivity of the substrates that form these cations as intermediates in chemical reactions. The hydride affinities of these and similar cations can again serve as a measure of their relative aromaticity. According to the electron affinities (Table 20), the most aromatic was cycloheptatrienyl cation, followed by the cyclopropenyl cation and then the antiaromatic cyclopentadienyl cation. In fact, the cycloheptatrienyl cation was the most stable carbocation of all cations studied by hydride affinity (Table 20), although the secondary carbocation, cyclopentadienyl cation, as an antiaromatic molecular system it has higher hydride affinity than a normal secondary carbocation such as 2-propyl cation or even the 1-phenylethyl cation.

7.2 Strain energies as a measure of reactivity

Tertiary substrates which have the leaving group at the junction of two rings (the bridgehead position) [149-152] tend to be very inert in solvolysis, although bridgehead carbocations will form in superacid solutions [151]. The principle cause is the additional strain energy accompanying ionization due to deformation of the

Table 22. These values demonstrated a high level of correspondence with some of

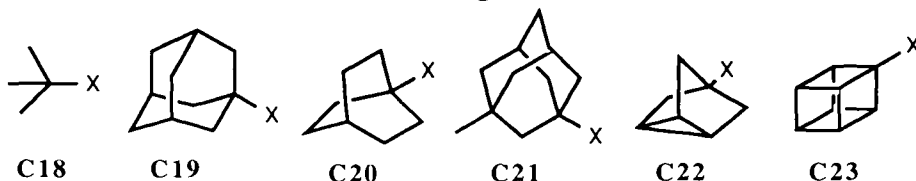


Table 22. Strain energy (kcal/mol) for accommodation of planar geometry of some bridgehead carbonium cations computed by B3LYP/6-31+G(d) on am1 geometries

Cation	E_c	E_{pc}	SE	SE_{exp}
C18			0.0	0.0
C19	-389.832050082	-389.811519663	12.9	12.0
C20	-312.380975227	-312.356573178	15.3	16.0
C21	-429.150656206	-429.115268503	22.2	20.0
C22	-271.771588349	-271.711849818	37.5	~35.0
C23	-308.537615136	-308.320836585	136.0	

E_c = total energy for carbocation (a.u.); E_{nm} = total energy for planar carbocation (a.u.); SE = computed strain energy (kcal/mol); SE_{exp} = experimental estimation of strain energy (kcal/mol).

carbocation from the preferred planar conformation, analogous to that in bridgehead alkenes. Rates of solvolyses of bridgehead tosylates, **C19-C23**, therefore, depend on the energy required for the carbocation to adapt to the planar structure.

The reactivity of these tosylates can be determined by computing the energy difference between the corresponding fully optimized carbonium cation and planar carbonium cation. The computed B3LYP/6-31+G(d) energies are presented in the experimental estimates. According to this approach, the carbocation that has the smallest strain energy will be the most reactive. Keeping this in mind, substrate **C18** would form a carbocation with no strain energy and would be most reactive while cubane substrate, **C23**, has substantial strain energy in formation of planar carbocation and will be the least reactive of all studied compounds. These computational studies are in full compliance with experimentally determined rates of solvolyses of some of the tosylates. This is, of course, only one of many ways of determining the reactivity of these compounds

8. CONCLUSION

Undoubtedly, there are many ways by which one can determine reactivity and selectivity of chemical transformations, one of them being through the use of computational techniques. This approach will play an increasingly more important role in the future. New computational methods that will be more cost effective and have a greater applicability to large chemical systems with higher accuracy will be developed. Nevertheless, some of the DFT computational methods have been shown to reproduce experimental activation energies and energy differences between reactants and key intermediates accurately. Therefore, it is time to introduce reactivity and selectivity in terms of the reciprocal values of the activation energies. Many reactants can produce different products through various reaction pathways, thus they will have different reactivities as well. Selectivity, on the other hand, can be defined as a difference of reactivity for the same reactant for two concurrent reactions. Today these values are readily available because computation of the activation energies for many chemical reactions is becoming quite a routine process.

This principle has been demonstrated here on only a limited number of examples. Certainly, this approach can be applied to any kind of chemical interaction. We hope that we have also demonstrated that frontier orbital interactions, spin and charge distribution, as well as bond orders, are very useful properties for evaluation of chemical reactivity and selectivity. In view of the advantages that they afford, more experimental organic chemists are encouraged to use them to design synthetic schemes for the preparation of new and valuable organic compounds as we are applying these approaches to our everyday organic synthetic projects.

REFERENCES

1. M. L. Casey, D. S. Kemp, K. Paul and D. Cox, *J. Org. Chem.* 38 (1973) 2294.
2. T. J. Gilbert and C. D. Johnson, *J. Am. Chem. Soc.* 96 (1974) 5846.
3. C. D. Johnson, *Chem. Rev.* 75 (1975) 755.
4. J. M. Seminario and P. Politzer (Eds.), *Modern Density Functional Theory: A Tool for Chemistry*, Elsevier, Amsterdam, 1995.
5. J. M. Seminario (ed), *Recent Developments and Applications of Modern Density Functional Theory*, Elsevier, Amsterdam, 1996.
6. J. Labanowski and J. Andzelm (eds.), *Density Functional Methods in Chemistry*, Springer, New York, 1991.
7. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University, Oxford, 1989.
8. Gaussian 94, Revision B.3, M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1995.
9. D. R. Hartree, *Proc. Cambridge Phil. Soc.* 24 (1928) 89.
10. V. Fock, *Z. Physik*, 61 (1930) 126.
11. C. C. J. Roothan, *Rev. Mod. Phys.* 23 (1951) 69.
12. J. A. Pople and R. K. Nesbet, *J. Chem. Phys.* 22 (1959) 571.
13. R. McWeeny and G. Dirksen, *J. Chem. Phys.* 49 (1968) 4852.
14. C. Møller and M. S. Plesset, *Phys. Rev.* 46 (1934) 618.
15. M. J. Frisch, M. Head-Gordon and J. A. Pople, *Chem. Phys. Lett.* 166 (1990) 281.
16. J. A. Pople, R. Seeger and R. Krishnan, *Int. J. Quantum Chem. Symp.* 11 (1977) 149.
17. R. Krishnan and J. Pople, *Int. J. Quantum Chem.* 14 (1978) 91.
18. R. Krishnan, M. J. Frisch and J. A. Pople, *J. Chem. Phys.* 72 (1980) 4244.
19. K. Raghavachari, J. A. Pople, E. S. Replogle and M. Head-Gordon, *J. Phys. Chem.* 94 (1990) 5579.
20. J. A. Pople, M. Head-Gordon and K. Raghavachari, *J. Chem. Phys.*, 87 (1987) 5968.
21. J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari and L. A. Curtiss, *J. Phys. Chem.* 90 (1989) 5622.
22. L. A. Curtiss, C. Jones, G. W. Trucks, K. Raghavachari and J. A. Pople, *J. Chem. Phys.* 93 (1990) 2537.
23. L. A. Curtiss, K. Raghavachari, G. W. Trucks and J. A. Pople, *J. Chem. Phys.* 94 (1991).
24. L. A. Curtiss, K. Raghavachari and J. A. Pople, *J. Chem. Phys.* 98 (1993) 1293.

25. A. D. Becke, *J. Chem. Phys.* 98 (1993) 5648.
26. C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 37 (1988) 785.
27. B. Miehlich, A. Savin, H. Stoll and H. Preuss, *Chem. Phys. Lett* 157 (1989) 200.
28. J. P. Perdew, *Phys. Rev. B* 33 (1986) 8822.
29. J. P. Perdew and Y. Wang, *Y. Phys. Rev. B* 45 (1992) 13244.
30. J. C. Slater, *Quantum Theory of Molecular and Solids. Vol. 4: Self-Consistent Field for Molecular and Solids*, McGraw-Hill, New York, NY, 1974.
31. A. D. Becke, *Phys. Rev. A*, 38 (1988) 3098.
32. J. P. Perdew and A. Zunger, *Phys. Rev. B*, 23 (1981) 5048.
33. S. H. Vosko, L. Wilk and M. Nusair, *Can. J. Phys.*, 58 (1980) 1200.
34. M. J. Frisch, \AA . Frisch and J. B. Foresman, *Gaussian 94 User's Reference*, Gaussian, Inc. Pittsburgh, PA, 1995.
35. J. B. Foresman and \AA . Frisch, *Exploring Chemistry with Electronic Structure Methods: A Guide to Using Gaussian*, Gaussian, Inc. Pittsburgh, PA, 1993.
36. J. C. Polanyi and A. H. Zewail, *Acc. Chem. Res.* 28 (1995) 119.
37. S. Pedersen, J. L. Herek and A. H. Zewail, *Science* 266 (1994) 1293.
38. G. S. Hammond, *J. Am. Chem. Soc.* 77 (1955) 334.
39. M. G. Evans and M. Polanyi, *Trans. Faraday Soc.* 32 (1936) 1340.
40. R. P. Bell, *Acid-Base Catalysis*, Oxford University Press, London, 1941, p. 85
41. M. J. S. Dewar, *The Molecular Orbital Theory of Organic Chemistry*, McGraw-Hill Inc., New York, NY, 1969, p. 284
42. A. Pross, *Adv. Phys. Org. chem.* 14 (1977) 69.
43. A. Pross, *Adv. Phys. Org. Chem.* 21 (1985) 99.
44. C. D. Ritchie, *Acc. Chem. Res.* 5 (1972) 348.
45. R. Alexander, E. C. F. Ko, A. J. Parker, T. J. Broxton, *J. Am. Chem. Soc.* 90 (1968) 5049.
46. D. E. Sunko, B. S. Jursic, M. Ladika, *J. Org. Chem.* 52 (1987) 2299.
47. Z. Rappoport, A. Pross, Y. Apeloig, *Tetrahedron Lett.* (1973) 2015.
48. L. P. Hammett, *Physical Organic Chemistry*, 2nd Ed., McGraw-Hill, New York, 1970, p. 355.
49. E. Grunwald and S. Winstain, *J. Am. Chem. Soc.* 70 (1948) 846.
50. C. G. Swain and C. B. Scott, *J. Am. Chem. Soc.* 75 (1953) 141.
51. J. N. Brønsted and E. A. Guggenheim, *J. Am. Chem. Soc.* 49 (1927) 2554.
52. B. S. Jursic, *Computing Transition State Structures with Density Functional Theory Methods*, IN: *Recent Developments and Applications of Modern Density Functional Theory*, J. M. Seminario, (ed.), Elsevier, Amsterdam, 1996, p. 709.
53. K. B. Wiberg and W. J. Barley, *J. Am. Chem. Soc.* 82 (1960) 6375.
54. J. E. Baldwin and V. P. Reddy, *J. Org. Chem.* 54 (1989) 5264.
55. K. Alder and G. Stein, *Angew. Chem.* 50 (1937) 510.
56. R. B. Woodward and R. Hoffmann, *The Conservation of Orbital Symmetry*, Verlag Chemie, Weinheim, 1971
57. R. Gleiter and M. C. Böhm, *Pure Appl. Chem.* 55 (1983) 237.

58. Y. Apeloig and E. Matzner, *J. Am. Chem. Soc.* 117 (1995) 5375.
59. P. Binger, P. Wedemann, R. Goddard and U. H. Brinker, *J. Org. Chem.* 61 (1996) 6462.
60. W. C. Hearndom and L. H. Hall, *Tetrahedron Lett.* 8 (1967) 3095.
61. J. J. Gajewski, *J. Org. Chem.* 57 (1992) 5500.
62. M. F. Ruiz-Lopez, X. Assfeld, J. I. Garc'a, J. A. Mayoral and L. Salvatella, *J. Am. Chem. Soc.* 115 (1993) 8780.
63. B. S. Jursic, *J. Mol. Struct. (Theochem)* 358 (1995) 139.
64. B. S. Jursic, *J. Mol. Struct. (Theochem)* 365 (1996) 55.
65. B. S. Jursic and B. LeBlanc, *J. Heterocyclic Chem.* 33 (1996) 1389.
66. B. S. Jursic and Z. Zdravkovski, *J. Chem. Soc. Perkin Trans 2* 1995, 1223 and references therein.
67. The bond orders are computed by computational routine included in SPARTAN. SPARTAN version 4.0. Wavefunction, Inc., 18401 Von Karman Ave., #370, Irvine, CA92715 U.S.A.
68. For an excellent discussion on the use of bond orders, see: G. Lendvay, *J. Phys. Chem.* 98 (1994) 6098.
69. K. Fukui and H. Fujimoto, *Bull. Chem. Soc. Jpn.* 40 (1967) 2018.
70. K. Fukui and H. Fujimoto, *Bull. Chem. Soc. Jpn.* 42 (1969) 2018.
71. K. Fukui, *Angew. Chem. Int. Ed. Engl.* 21 (1982) 801.
72. K. N. Houk, *Acc. Chem. Res.* 8 (1975) 361.
73. K. N. Houk, *J. Am. Chem. Soc.* 95 (1973) 492.
74. K. N. Houk, *Top. Curr. Chem.* 79 (1979) 1.
75. P. Binger, P. Wedemann, R. Goddard and U. H. Brinker, *U. H. J. Org. Chem.* 61 (1996) 61, 6462.
76. K. Geibel and J. Heindl, *Tetrahedron Lett.* 11 (1970) 2133
77. M. A. Battiste and C. T. Sprouse Jr., *Tetrahedron Lett.* 11 (1970) 3165.
78. M. A. Battiste and C. T. Sprouse Jr., *Tetrahedron Lett.* 11 (1970) 4661.
79. M. P. Cava and K. Narasimhan, *J. Org. Chem.* 36 (1971) 1419.
80. R. Breslow and M. Oda, *J. Am. Chem. Soc.* 94 (1972) 4787.
81. M. Oda, R. Breslow and J. Pecoraro, *Tetrahedron Lett.* 13 (1972) 4419.
82. M. H. J. Cordes, S. deGala and J. A. Berson, *J. A. J. Am. Chem. Soc.* 116 (1994) 11161.
83. B. S. Jursic, *J. Org. Chem.*, in press.
84. B. S. Jursic, *Tetrahedron Lett.*, 38 (1997) 1305.
85. For an excellent review of the application of theory for study of cyclobutene ring opening, see: W. R. Dolbier, Jr., H. Koroniak, K. N. Houk and C. Sueu, *Acc. Chem. Res.* 29 (1996) 471.
86. R. Srinivasan, *J. Am. Chem. Soc.*, 91 (1969) 7557.
87. H. C. Longuet-Higgins and E. W. Abrahamson, *J. Am. Chem. Soc.*, 87 (1965) 2045.
88. R. B. Woodward and R. Hoffmann, *J. Am. Chem. Soc.*, 87 (1965) 395.
89. M. J. S. Dewar, *Tetrahedron, Suppl. B*, (1966), 76.
90. H. E. Zimmerman, *J. Am. Chem. Soc.*, 88 (1966) 1563.
91. H. E. Zimmerman, *J. Am. Chem. Soc.*, 88 (1966) 1566.

92. H. M. Frey, *Trans. Faraday Soc.*, 58 (1962) 957.
93. M. J. Curry and I. D. R. Stevens, *J. Chem. Soc., Perkin Trans. 2*, (1980) 1391.
94. W. R. Dolbier, Jr., H. Korniak, D. J. Burton, P. L. Heinze, A. R. Baily, G. S. Shaw and S. W. Hansen, *J. Am. Chem. Soc.*, 106 (1984) 1871.
95. D. Dickens, H. M. Frey and J. Mercalf, *Trans. Faraday Soc.* 67 (1971) 2328.
96. R. Willstätter and W. von Schmaedel, *Ber. Deutsch. Chem. Ges.* 38 (1905) 1992.
97. E. Vogel, *Angew. Chem.* 66 (1954) 640.
98. M. J. Goldstein, R. S. Leight and M. S. Lipton, *J. Am. Chem. Soc.* 98 (1976) 5717.
99. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.* 107 (1985) 3902.
100. J. Breulet and H. F. Schaefer III, *J. Am. Chem. Soc.* 106 (1984) 1221.
101. E. A. Kallel, Y. Wang and K. N. Houk, *J. Org. Chem.* 54 (1989) 6006.
102. F. Bernardi, S. De, M. Olivucci and M. A. Robb, *J. Am. Chem. Soc.* 112 (1990) 1737
103. B. S. Jursic and Z. Zdravkovski, *Int. J. Quant. Chem.* 56 (1995) 115.
104. M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.* 99 (1977) 4499.
105. B. S. Jursic, to be published.
106. J. S. Murray and K. Sen, (eds.), *Molecular Electronic Potentials: Concepts and Applications*, Elsevier, Amsterdam, 1996.
107. S. F. Sarnar, D. M. Gale, H. K. Hall. Jr. and A. B. Richmond, *J. Phys. Chem.* 76 (1972) 2817.
108. M. L. Poutsman, in: *Free Radicals*; Vol. II; J. E. Kochi, (ed.), J. Wiley, New York, NY, 1973.
109. C. Rüchardt, *Angew. Chem. Int. Ed. Engl.* 7 (1970) 830.
110. A. F. Trotman-Dickenson, *Adv. Free Radical Chem.* 1 (1965) 1.
111. J. M. Tedder, *Angew. Chem. Int. Ed. Engl.* 21 (1982) 401.
112. A. A. Zavitsas and A. A. Melilin, *J. Am. Chem. Soc.* 97 (1975) 2757.
113. J. C. Martin, IN: "Free Radicals"; Vol II; J. K. Kochi, Ed. (Wiley, New York, 1973, p.493).
114. W. H. S. Yu and M. H. Wijnen, *J. Chem. Phys.* 52 (1970) 4166.
115. G. A. Russell and C. De Boer, *J. Am. Chem. Soc.* 85 (1963) 3136 and references therein.
116. M. L. McKee, *J. Am. Chem. Soc.* 112 (1990) 7957.
117. D. J. Pasto, R. Krasnansky and C. Zercher, *J. Org. Chem.* 52 (1987) 3062 and references therein.
118. B. S. Jursic, *J. Mol. Struct. (Theochem)*, 365 (1996) 75.
119. I. Matheson, J. Tedder and H. Siedebottom, *Int. J. Chem. Kinet.* 14 (1982) 1033.
120. J. H. Knox and R. L. Nelson, *Trans. Faraday Soc.* 55 (1959) 937.
121. J. W. Wilt, in: *Free Radicals Vol I*, J. K. Kochi, (ed.), J. Wiley, New York, NY, 1973, p 432.
122. P. Gottschalk and D. C. Neckers, *J. Org. Chem.* 50 (1985) 3498.

123. G. Stork and R. Mook, Jr. *J. Am. Chem. Soc.* 105 (1983) 3720.
124. M. J. S. Dewar and S. Olivella, *J. Am. Chem. Soc.* 100 (1978) 5290.
125. D. C. Spellmeyer and K. N. Houk, *J. Org. Chem.* 52 (1987) 959.
126. A. L. J. Beckwith and C. H. Schiesser, *Tetrahedron* 41 (1985) 3925.
127. L. Mathew and J. Warkentin, *J. Am. Chem. Soc.* 108 (1986).
128. D. Griller and K. Ingold, *Acc. Chem. Res.* 13 (1980) 317.
129. L. J. Johnson, J. Lusztyk, D. D. M. Wayner, A. N. Abeywickreyman, A. L. Beckwith, J. C. Scaiano, and K. U. Ingold, *J. Am. Chem. Soc.* 107 (1985) 4594.
130. G. A. Olah and P. v. R. Schleyer. (eds.), *Carbonium Ions*, J. Wiley, New York,, NY, 1973.
131. D. W. Berman, V. Anicich and J. L. Beauchamp, *J. Am. Chem. Soc.* 101 (1979) 1239.
132. K. B. Wiber, T. A. Keith, M. J. Frish and M. Murcko, *J. Phys. Chem.* 99 (1995) 9072.
133. D. H. Aue and M. T. Bowers, In *Gass Phase Ion Chemistry*, M. T. Bowers,, (ed.), Academic Press, New York, NY, 1979.
134. F. A. Houle and J. L. Beauchamp, *J. Am. Chem. Soc.* 101 (1979) 4067.
135. D. W. Berman, V. Anichich and J. L. Beauchamp, *J. Am. Chem. Soc.* 101 (1979) 1239.
136. G. A. Petersson, T. G. Tensfeldt and J. A. Montgomery. Jr., *J. Chem. Phys.*, 101 (1991) 6091.
137. B. S. Jursic, to be published.
138. W. J. Hehre, L. Radom, P. v. R. Schleyer and J. A. Pople, *Ab Initio Molecular Orbital Theory*, J. Wiley, New York, NY, 1986.
139. J. A. Pople, R. Krishnan, H. B. Schlegel, D. DeFrees, J. S. Binkley, M. J. Frish, R. F. Whiteside and W. J. Hehre, *Int. J. Quantum Chem.*, Symp. 15 (1981) 269.
140. C. K. Ingold, *Structure and Mechanism in Organic Chemistry*, 2nd Ed., Cornell University Press, Ithaca, New York, NY, 1969.
141. G. A. Olah, *J. Am. Chem. Soc.* 94 (1972) 808.
142. T. W. Bentley and P. v. R. Schleyer, *Adv. Phys. Org. Chem.* 14 (1977) 1.
143. F. L. Schadt, T. W. Bentley and P. v. R. Schleyer, *J. Am. Chem. Soc.* 98 (1976) 7667.
144. S. S. Kanter, K. Humski and H. L. Goering, *J. Am. Chem. Soc.* 104 (1982) 1693.
145. P. K. Norris, S. D. Barker, P. Neta, *J. Am. Chem. Soc.* 106 (1984) 3140.
146. J. P. Snyder, *Nonbenzoid Aromatics*, Vol 1, Academic Press, New York, NY, 1969.
147. M. J. S. Dewar, *The Molecular Orbital Theory of Organic Chemistry*, McGraw-Hill, New York, NY, 1969.
148. H. E. Zimmerman, *Acc. Chem. Res.* 4 (1971) 272.
149. P. v. R. Schleyer, *Adv. Alicyc. Chem.* 1 (1966) 283.
150. R. C. Bingham and P. v. R. Schleyer, *J. Am. Chem. Soc.* 93 (1971) 3189.
151. P. v. Schleyer and R. D. Nicholas, *J. Am. Chem. Soc.* 83 (1961) 2700.

152. G. A. Olah, *J. Am. Chem. Soc.* 107 (1985) 2764.
153. W. H. Watson. (ed.), *Stereochemistry and Reactivity of Systems Containing π Electrons*, Verlag Chemie, Deerfield Beach, FL, 1983.

This Page Intentionally Left Blank

A HARDNESS AND SOFTNESS THEORY OF BOND ENERGIES AND CHEMICAL REACTIVITY

José L. Gázquez

Departamento de Química, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana-Iztapalapa, A.P. 55-534, México, D.F. 09340, México

1. INTRODUCTION

The determination of the specific sites at which the interaction between two chemical species is going to occur, is of fundamental importance to determine the path and the products of a given reaction. In principle, from a theoretical viewpoint, one should calculate the potential energy surface associated with the interacting species, to obtain the reaction coordinate that allows one to establish the path followed by the reacting molecules to reach the transition state and the final products. However, in practice, this procedure may be very complicated and, in general, it may not necessarily lead one to obtain simple chemically significant information to establish the behavior of a family of molecules with respect to a family of reactants. Thus, over the years, chemists have developed intuitive concepts and simple theories that have allowed them to understand the behavior of molecules under different circumstances, their reactive sites, and possible reaction mechanisms.

Among these concepts, certainly, electronegativity [1], hardness and softness [2] on one hand, and frontier orbital theory [3] on the other hand, have played a major role to analyze the experimental evidence and to establish, a priori, the development of a wide variety of chemical reactions [4].

The objective of the present chapter is to make use of density functional theory [5] to develop a simple unified approach to bond energies, activation energies, and chemical reactivity, in order to show that the hardness and softness concepts provide fundamental information about the changes in energy associated with the interaction of chemical species. Thus, in the next section, some basic aspects of density functional theory will be reviewed, and it will be shown that within the framework provided by this theory, the fundamental equations for chemical events may be expressed in terms of several reactivity parameters, such as the chemical potential [6] (electronegativity), the hardness [7], the softness [8], the electronic density, and the Fukui function [9], which is closely related to the frontier orbitals. In Section 3, it will be shown that the total electronic energy may be expressed in terms of an effective valence energy contribution and an effective core contribution. The former depends mainly on the chemical potential and the hardness of the system, and accounts for the main energy changes that occur in a chemical interaction, while the latter depends mainly on the core densities of the interacting molecules, and it approximately cancels the nuclear-nuclear interaction energy. The energy expression will be then used to examine the principle of maximum hardness [10-13], and the principle of hard and soft acids and bases [14,15], and it

will also be used to analyze bond, activation, and reaction energies in terms of hardness differences [16,17]. The implications of the present approach in the cases of catalyzed reactions, and reactions in solution will be briefly discussed in Section 4 and, finally, in Section 5 the main conclusions of this work will be presented.

2. REACTIVITY PARAMETERS

In the last years, it has been found that the intuitive concepts of electronegativity, hardness, and softness are closely related to basic variables of density functional theory [5]. Thus, through the use of this theory it has been possible to establish expressions that allow one to quantify these concepts, and it has also been possible to demonstrate the principles associated with them. The purpose of the present section is to review some fundamental aspects of density functional theory, and to establish the expressions for several reactivity parameters.

2.1. The density functional theory framework

Density functional theory provides an alternative approach to the description of the electronic structure of atoms and molecules, in principle more simple than conventional wavefunction theory, because the total energy is determined directly from the electronic density distribution $\rho(\mathbf{r})$, through the expression

$$E[\rho] = F[\rho] + \int d\mathbf{r} \rho(\mathbf{r}) v(\mathbf{r}) \quad (1)$$

where $F[\rho]$ is the universal Hohenberg-Kohn functional, and $v(\mathbf{r})$ is the external potential generated by the nuclei. The functional $F[\rho]$ is given by the sum of the kinetic $T[\rho]$, the classical Coulomb interaction $J[\rho]$, and the exchange-correlation $E_{xc}[\rho]$ energy density functionals,

$$F[\rho] = T[\rho] + J[\rho] + E_{xc}[\rho] \quad (2)$$

The electronic density of the ground state is determined from the Euler-Lagrange equation that results from the variation of Eq. (1) with respect to $\rho(\mathbf{r})$, subject to the condition that the integral of the latter over the whole space must be equal to the total number of electrons N ,

$$\mu = \delta F[\rho] / \delta \rho(\mathbf{r}) + v(\mathbf{r}) \quad (3)$$

where μ , the Lagrange multiplier, is the chemical potential of the system, and it is a constant throughout the whole space. The solution of Eq. (3) leads to the ground state electronic density, that may be substituted in Eq. (1) to obtain the ground state energy.

The problem of a pure density functional theory lies in the fact that the exact forms of $T[\rho]$ and $E_{xc}[\rho]$ in terms of $\rho(\mathbf{r})$ are unknown at the moment [18,19], and the approximations, which are generally based on the generalized gradient expansion, do not provide the accuracy required for chemical properties, such as the bond energies, mainly

because the errors in the kinetic energy functional are at least of the same order of magnitude than most of the relevant energy differences required to study chemical problems. This is not the case in the Kohn-Sham version of density functional theory [5,18,19], that avoids the problem of the kinetic energy by introducing an orbital language in which the kinetic energy is given by the expression corresponding to a system of non interacting electrons, and the small difference between the latter and the exact kinetic energy is included in the exchange-correlation energy density functional. In this case, the accuracy provided by several approximations to $E_{xc}[\rho]$, allows one to determine energy differences that agree very well with the experimental values. This procedure has proven to be a very effective tool for the description of the electronic structure of atoms and molecules, and it is now extensively used to study a wide variety of chemical problems [18,19].

On the other hand, the pure density functional theory provides a conceptual framework that has proven to be very useful to establish expressions that are closely related with chemical concepts, and to rationalize, through the values associated with them, the behavior of a wide variety of chemical species under different circumstances [20]. In order to calculate the values of these quantities, it has become common practice to make use of conventional molecular orbital theory [21-26]. Thus, this procedure allows one to transform the relevant information contained in the wavefunction, into chemically meaningful results, through the bridge provided by pure density functional theory.

2.2. Fundamental concepts

Consider that the electronic energy is a function of the total number of electrons, and a functional of the external potential, then, using Eqs. (1) and (3), one can show that [5]

$$dE = \mu dN + \int d\mathbf{r} \rho(\mathbf{r}) \delta v(\mathbf{r}) \quad (4)$$

According to this expression one can see that the chemical potential is equal to the derivative of the electronic energy with respect to the total number of electrons, when the external potential is held fixed,

$$\mu = (\partial E / \partial N)_v = -\chi \quad (5)$$

where χ is the electronegativity. From Eq. (4) one can also see that the electronic density is equal to the functional derivative of the electronic energy with respect to the external potential, when the number of electrons is held fixed,

$$\rho(\mathbf{r}) = (\delta E / \delta v(\mathbf{r}))_N \quad (6)$$

Thus, from Eq. (5) one may identify that the chemical potential of density functional theory is equal to the negative of the electronegativity of chemistry. This identification allows one to understand the principle of electronegativity equalization [6], because, according to Eq. (3), when two species with different chemical potentials are placed together, their values must change to reach a common value. That is, the principle of electronegativity equalization is just a principle of chemical potential equalization. The equalization of μ is achieved through charge

transfer between the interacting species, and through the changes in μ produced by the changes in the external potentials of the isolated species when they are placed together.

Consider now that the chemical potential is a function of the total number of electrons, and a functional of the external potential, then, using Eq. (4), one can show that [5]

$$d\mu = \eta dN + \int d\mathbf{r} f(\mathbf{r}) \delta v(\mathbf{r}) \quad (7)$$

According to this expression one can see that η , which has been identified as the chemical hardness (the factor of 1/2 in the original definition [7] of the global hardness has been omitted here for convenience), is equal to the derivative of the chemical potential with respect to the total number of electrons, when the external potential is held fixed,

$$\eta = (\partial\mu / \partial N)_v = (\partial^2 E / \partial N^2)_v \quad (8)$$

where the second equality comes from Eq. (5). From Eq. (7) one can also see that $f(\mathbf{r})$, which has been identified as the Fukui function, is equal to the functional derivative of the chemical potential with respect to the external potential, when the number of electrons is held fixed,

$$f(\mathbf{r}) = (\delta\mu / \delta v(\mathbf{r}))_N = (\partial\rho(\mathbf{r}) / \partial N)_v \quad (9)$$

where the second equality comes from a Maxwell relation [27].

It is interesting to look at the finite differences approximations to the derivatives in Eqs. (5), (8), and (9). In the case of the first and the second derivatives of the total energy with respect to the number of electrons one finds that [7]

$$\mu \approx -\frac{1}{2}(I + A) \quad (10)$$

and

$$\eta \approx I - A \quad (11)$$

where I is the first ionization potential and A is the electron affinity of the reference system. These formulas provide a strong support for the interpretation of the first and second derivatives of the energy with respect to the number of electrons as the chemical potential (electronegativity) and the chemical hardness, respectively [28], because, in general, the behavior of μ and η , when they are calculated through Eqs. (10) and (11), agrees with empirical scales of these quantities.

In a finite differences approximation the Fukui function becomes [9]

$$f^+(\mathbf{r}) \approx \rho_{N+1}(\mathbf{r}) - \rho_N(\mathbf{r}) \quad \text{for nucleophilic attack} \quad (12)$$

$$f^-(\mathbf{r}) \approx \rho_N(\mathbf{r}) - \rho_{N-1}(\mathbf{r}) \quad \text{for electrophilic attack} \quad (13)$$

and

$$f^0(\mathbf{r}) \approx \frac{1}{2}(\rho_{N+1}(\mathbf{r}) - \rho_{N-1}(\mathbf{r})) \quad \text{for radical attack} \quad (14)$$

where $\rho_{N+1}(\mathbf{r})$, $\rho_N(\mathbf{r})$, and $\rho_{N-1}(\mathbf{r})$ are the electronic densities of the system with $N+1$, N , and $N-1$ electrons, respectively, all with the ground state geometry of the N electron system. The difference between $\rho_{N+1}(\mathbf{r})$ and $\rho_N(\mathbf{r})$ is closely related to the density of the lowest unoccupied molecular orbital (LUMO), while the difference between $\rho_N(\mathbf{r})$ and $\rho_{N-1}(\mathbf{r})$ is closely related to the density of the highest occupied molecular orbital (HOMO). That is, the fukui function is a quantity closely related to the frontier orbitals.

Thus, one can see that within the framework provided by density functional theory, the basic equations for the description of a chemical event, Eqs. (4) and (7), may be expressed in terms of basic variables such as the chemical potential (electronegativity), the chemical hardness and the fukui function (frontier orbitals). In fact, through this approach one may introduce a coherent quantitative language of hardness and softness functions which are non-local, local, and global [29]. The global softness is given by

$$S = (\partial N / \partial \mu)_v = 1 / \eta \quad (15)$$

the local softness is given by

$$S(\mathbf{r}) = (\partial \rho(\mathbf{r}) / \partial \mu)_v = (\partial N / \partial \mu)_v (\partial \rho(\mathbf{r}) / \partial N)_v = S f(\mathbf{r}) \quad (16)$$

and the softness kernel is given by

$$S(\mathbf{r}, \mathbf{r}') = -(\delta \rho(\mathbf{r}) / \delta u(\mathbf{r}')) \quad (17)$$

where $u(\mathbf{r}) = v(\mathbf{r}) - \mu = -(\delta F[\rho] / \delta \rho(\mathbf{r}))$ is the intrinsic potential. The inverse of the softness kernel is the hardness kernel,

$$\eta(\mathbf{r}, \mathbf{r}') = (\delta^2 F[\rho] / \delta \rho(\mathbf{r}') \delta \rho(\mathbf{r})) = -(\delta u(\mathbf{r}) / \delta \rho(\mathbf{r}')) \quad (18)$$

Now, in the case of the finite differences approximation to the derivative in the second equality of Eq. (9), because of the local dependence on the position within the molecule, instead of using $f(\mathbf{r})$ directly, it is more simple to condense its values around each atomic site into a single value that characterizes the atom in the molecule. This can be done by first condensing the electronic density to the charge of each atom in the molecule, and differentiating afterwards with respect to the total number of electrons in the system [30]. Thus, the finite differences approximation leads to three indexes known as the condensed fukui functions,

$$f_{A_i}^+ = q_{A_i}(N_A + 1) - q_{A_i}(N_A) \quad \text{for nucleophilic attack} \quad (19)$$

$$f_{Ai}^- = q_{Ai}(N_A) - q_{Ai}(N_A - 1) \text{ for electrophilic attack} \quad (20)$$

and

$$f_{Ai}^0 = \frac{1}{2} (q_{Ai}(N_A + 1) - q_{Ai}(N_A - 1)) \text{ for radical attack} \quad (21)$$

where q_{Ai} is the charge of the i -th atom in the molecule A , that may be determined by several procedures, the simplest one being the Mulliken population analysis. Similarly, because of the last equality in Eq. (16), one may also consider three additional indexes known as the condensed local softnesses

$$S_{Ai}^+ = S_A f_{Ai}^+ \text{ for nucleophilic attack} \quad (22)$$

$$S_{Ai}^- = S_A f_{Ai}^- \text{ for electrophilic attack} \quad (23)$$

and

$$S_{Ai}^0 = S_A f_{Ai}^0 \text{ for radical attack} \quad (24)$$

The fundamental concepts of chemistry just established above may be quantified, leading to a description of the inherent chemical reactivity of molecules, that may be used to describe their behavior with respect to nucleophilic, electrophilic, and free radical attacks, and to rationalize experimental information.

3. ENERGY AND HARDNESS DIFFERENCES

Consider a system that is composed of several molecules that interact with each other. The total energy difference between an initial state, that will be considered here as the state when the interacting molecules are very far apart from each other, and any other state, when all molecules are close to each other, is given by

$$\Delta E = E_f[\rho_f] - E_i[\rho_i] + V_{NN}^f - V_{NN}^i \quad (25)$$

where the final state is characterized by the electronic energy E_f , the external potential $v_f(\mathbf{r})$, the electronic density $\rho_f(\mathbf{r})$, and the chemical potential μ_f , while the initial state is characterized by E_i , $v_i(\mathbf{r})$, $\rho_i(\mathbf{r})$, and μ_i . The external potential $v_f(\mathbf{r})$ is the potential generated by the nuclei in the configuration corresponding to that of the final state, while $v_i(\mathbf{r})$ is the potential generated by the nuclei when all the reacting molecules are very far

away from each other. The quantities V_{NN}^f and V_{NN}^i represent the nuclear-nuclear repulsion energy in the final and in the initial states, respectively.

In order to evaluate the electronic energy difference one can make use of the expression [15]

$$E[\rho] = N_e \mu - \frac{1}{2} N_e^2 \eta + E_{\text{core}}[\rho] \quad (26)$$

where N_e represents an effective number of valence electrons, and

$$E_{\text{core}}[\rho] = \int d\mathbf{r} \rho_c(\mathbf{r}) v(\mathbf{r}) + \frac{1}{2} \iint d\mathbf{r} d\mathbf{r}' \frac{\rho_c(\mathbf{r}) \rho_c(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{1}{2} \iint d\mathbf{r} d\mathbf{r}' \rho_c(\mathbf{r}) \rho_c(\mathbf{r}') \frac{\delta^2 (T[\rho] + E_{\text{xc}}[\rho])}{\delta\rho(\mathbf{r}') \delta\rho(\mathbf{r})} \quad (27)$$

represents the core contribution to the total electronic energy. In the latter expression the second functional derivative of the classical Coulomb interaction energy density functional, $J[\rho]$, has been replaced by $1/|\mathbf{r}-\mathbf{r}'|$, and [13]

$$\rho_c(\mathbf{r}) = \rho(\mathbf{r}) - N_e f(\mathbf{r}) \quad (28)$$

Note that since the integral of the electronic density over the whole space is equal to the total number of electrons, and the integral of the fukui function over the whole space is equal to one, then $\rho_c(\mathbf{r})$ integrates to N_c ,

$$N_c = N - N_e \quad (29)$$

Equation (26) may be derived from Eqs. (1) and (3), the second order functional expansion [13,28,31,32] of $F[\rho]$ in terms of its functional derivatives,

$$F[\rho] = \int d\mathbf{r} \rho(\mathbf{r}) \frac{\delta F[\rho]}{\delta\rho(\mathbf{r})} - \frac{1}{2} \iint d\mathbf{r} d\mathbf{r}' \rho(\mathbf{r}) \rho(\mathbf{r}') \frac{\delta^2 F[\rho]}{\delta\rho(\mathbf{r}') \delta\rho(\mathbf{r})} + \frac{1}{6} \iiint d\mathbf{r} d\mathbf{r}' d\mathbf{r}'' \rho(\mathbf{r}) \rho(\mathbf{r}') \rho(\mathbf{r}'') \frac{\delta^3 F[\rho]}{\delta\rho(\mathbf{r}'') \delta\rho(\mathbf{r}') \delta\rho(\mathbf{r})} + \dots \quad (30)$$

the first order functional expansion of $\delta F[\rho]/\delta\rho(\mathbf{r})$ in terms of its functional derivatives,

$$\frac{\delta F[\rho]}{\delta\rho(\mathbf{r})} = \int d\mathbf{r}' \rho(\mathbf{r}') \frac{\delta^2 F[\rho]}{\delta\rho(\mathbf{r}') \delta\rho(\mathbf{r})} - \frac{1}{2} \iint d\mathbf{r}' d\mathbf{r}'' \rho(\mathbf{r}') \rho(\mathbf{r}'') \frac{\delta^3 F[\rho]}{\delta\rho(\mathbf{r}'') \delta\rho(\mathbf{r}') \delta\rho(\mathbf{r})} + \dots \quad (31)$$

and the inverse relationship between the hardness (Eq. (18)) and the softness (Eq. (17)) kernels [29]. The energy expression given by Eq. (26) contains only terms up to the second order in the functional derivative of $F[\rho]$.

Substituting Eq. (26) in Eq. (25) one finds that

$$\Delta E = N_e (\mu_f - \mu_i) - \frac{1}{2} N_e^2 (\eta_f - \eta_i) + E_{\text{core}}[\rho_f] - E_{\text{core}}[\rho_i] + \Delta V_{\text{NN}} \quad (32)$$

Now, by assuming that the core density of the system remains unchanged at any distance during the interaction, and that there is practically no overlap between the core densities of all the atoms that form part of the molecules that interact with each other, then one can show that if $N_e \ll N$, the sum of the terms in $(E_{\text{core}}[\rho_f] - E_{\text{core}}[\rho_i])$ associated with the first two terms in the right hand side of Eq. (27) is approximately equal to $-\Delta V_{\text{NN}}$. If it is further assumed that the core terms difference related with the second functional derivatives of the kinetic and the exchange-correlation energies cancel each other, one finds, from Eq. (32), if the chemical potential remains constant, that

$$\Delta E \approx -\frac{1}{2} N_e^2 (\eta_f - \eta_i) = -\frac{1}{2} N_e^2 \left(\frac{1}{S_f} - \frac{1}{S_i} \right) \quad (33)$$

where Eq. (15) has been used in the second equality.

This expression allows one to demonstrate the maximum hardness principle, because it establishes that for any process that occurs at constant chemical potential, as E increases, η must decrease, and as E decreases η must increase. A maximum of E corresponds to a minimum of η , and a minimum of E corresponds to a maximum in η (maximum hardness principle). It is important to note that in earlier demonstrations the maximum hardness principle was stated for the electronic energy [12,13,32]. Thus, in such context, an extremal in the hardness will coincide with an extremal in the total energy only when the nuclear-nuclear repulsion energy is also an extremal in the same point. However, in the present context, one can see that if there is a great cancellation between the core energy change $(E_{\text{core}}[\rho_f] - E_{\text{core}}[\rho_i])$, and the nuclear-nuclear repulsion energy change ΔV_{NN} , the maximum hardness principle holds for the total energy, independently of any extremal in V_{NN} or in the electronic energy. Numerical evidence supporting these statements has been reported in several works [12,33-39].

On the other hand, the numerical evidence has also shown that even if the chemical potential does not remain constant, the changes in μ are, in general, very small in comparison with the changes in η [12,16,33-39]. Thus, one may assume that for many cases the change in the chemical potential is negligible with respect to the change in the hardness, and therefore Eq. (33) may be used as an approximate expression for the energy changes, that may be applied to study different aspects of chemical reactivity.

Equation (26) has also been applied [15] to examine the hard and soft acids and bases principle, that establishes that hard acids prefer to coordinate to hard bases, and soft acids to soft bases. By determining the energy difference between two chemical species A and B through Eq. (26), and by approximating the equilibrium chemical potential, and the equilibrium

chemical hardness of AB, in terms of the chemical potentials and the hardnesses of the isolated systems, it has been shown that the interaction between species whose hardnesses are approximately equal is the one that leads to the greatest stabilization energy. Thus, even though there may be a favorable interaction between species whose hardnesses are very different from each other, the most favorable interaction occurs when the hardnesses of the two systems are approximately equal to each other.

An important aspect related with Eq. (33) comes from the fact that the hardness may be approximated in terms of the eigenvalues of the highest occupied (HOMO), and the lowest unoccupied molecular orbitals (LUMO),

$$\eta = \epsilon_{\text{LUMO}} - \epsilon_{\text{HOMO}} \quad (34)$$

Thus, substituting this expression in Eq. (33) one finds that

$$\Delta E \approx -\frac{1}{2} N_e^2 \left[(\epsilon_{\text{LUMO}}^f - \epsilon_{\text{HOMO}}^f) - (\epsilon_{\text{LUMO}}^i - \epsilon_{\text{HOMO}}^i) \right] \quad (35)$$

a result that shows, explicitly, that the frontier orbitals play a fundamental role in the description of a chemical event. Equation (35) establishes that the HOMO-LUMO gap will be at maximum when the energy is at minimum, and it will be at minimum when the energy is at maximum. These statements are in agreement with theoretical calculations [12,34-39].

3.1. Bond energies

In order to analyze the implications of Eq. (33) with respect to bond energies, first consider that the system consists of two species, A and B, that interact with each other forming a bond when they reach an equilibrium position. In this case, the energy change will depend on the difference between the hardness of the system AB in the equilibrium position, $\eta_{\text{AB}} = 1/S_{\text{AB}}$, and the hardness of the system when A and B are very far away from each other. In view of Eq. (15), it has been shown [14,16,40] that it seems reasonable to assume that the inverse of the latter, the softness of the system when A and B are very far away from each other, is roughly equal to the sum of the softnesses of the isolated species, because in the initial state there is practically no overlap between them. Thus, according to Eq. (33), the bond energy between A and B is approximately given by

$$\Delta E_{\text{bond}} \approx -\frac{1}{2} N_e^2 \left(\frac{1}{S_{\text{AB}}} - \frac{1}{S_{\text{A}} + S_{\text{B}}} \right) = -\frac{1}{2} N_e^2 \left(\eta_{\text{AB}} - \frac{\eta_{\text{A}} \eta_{\text{B}}}{\eta_{\text{A}} + \eta_{\text{B}}} \right) \quad (36)$$

If S_{AB} is smaller than $(S_{\text{A}} + S_{\text{B}})$, then ΔE_{bond} will be negative, a result that implies that the strength of the bond between A and B increases when the hardness increases.

Now, the bond energy given by Eq. (36) depends on the parameters associated with the isolated species A and B, and on the softness of the system AB in the equilibrium position S_{AB} . It would be interesting to express the latter also in terms of the isolated species parameters to obtain an expression for the bond energy just in terms of the properties of the interacting molecules. It has been shown [16] that this may be achieved by making use of the arithmetic average principle for molecular softness [41], that establishes that the softness of a

system in the equilibrium position may be approximated by the arithmetic average of the softnesses of the constitutive atoms. Thus, in the present case,

$$S_{AB} \approx (1/2)(S_A + S_B) \quad (37)$$

and therefore, substituting this expression in Eq. (36) one finds that

$$\Delta E_{\text{bond}} \approx -\frac{1}{2} N_e^2 \left(\frac{1}{S_A + S_B} \right) = -\frac{1}{2} N_e^2 \left(\frac{\eta_A \eta_B}{\eta_A + \eta_B} \right) \quad (38)$$

This approximation establishes that the harder the species that interact, the stronger the bonds that they may form.

Now, if one sets $N_e = 1$, and one makes use of the experimental values of I and A in Eq. (11), to determine the hardnesses in Eq. (38), one can calculate the bond energy. In Figure 1, one can see a comparison between the calculated and the experimental bond energies of 249 systems reported by Huheey, Keiter and Keiter [42]. Note that the majority of the points are clustered around a small region, indicating that, indeed, Eq. (38) provides the correct trends and reasonable estimates of the bond energy. On the other hand, if one determines the value of N_e that will reproduce the experimental bond energy, one finds for the same 249 cases an average value of 1.12, which is close to a value of one as expected.

It is interesting to consider the case when A or B, or both, correspond to molecular fragments, because in these cases the interaction occurs through specific atoms of A and B. In this context, it has also been shown [16] that it seems reasonable to assume that the interaction energy will be dominated by the softnesses of the specific atoms when they are placed in the chemical environments provided by fragments A and B respectively, rather than by the softnesses of A and B. This is equivalent to the assumption that only a specific atom of A and a specific atom of B participate in the interaction, and that the changes in all the other atoms of A and B can be neglected, which means that one should replace the global softnesses S_A and S_B by the condensed local softnesses $S_{A_i} = S_A f_{A_i}$ and $S_{B_j} = S_B f_{B_j}$, because these values characterize better than the global values the behavior of the site at which the interaction takes place [25,26]. Thus, the bond energy between the i -th atom in fragment A, and the j -th atom in fragment B, according to Eq. (38), is given by

$$\Delta E_{\text{bond}}^{A_i B_j} = -\frac{1}{2} N_e^2 \left(\frac{1}{S_A f_{A_i} + S_B f_{B_j}} \right) \quad (39)$$

This approximation establishes that the strongest bond in a molecule is the one formed by the adjacent atoms with the smallest values of the condensed fukui function, and that the weakest bond is the one formed by the adjacent atoms with the largest values of the condensed fukui function. Note that since the condensed fukui functions are different for nucleophilic, electrophilic, and free radical attacks, the weakest bond in a molecule, which may be associated with the most reactive site (this one may be either of the two atoms forming the bond or the bond itself), may be a different one, depending on the type of attack, in agreement

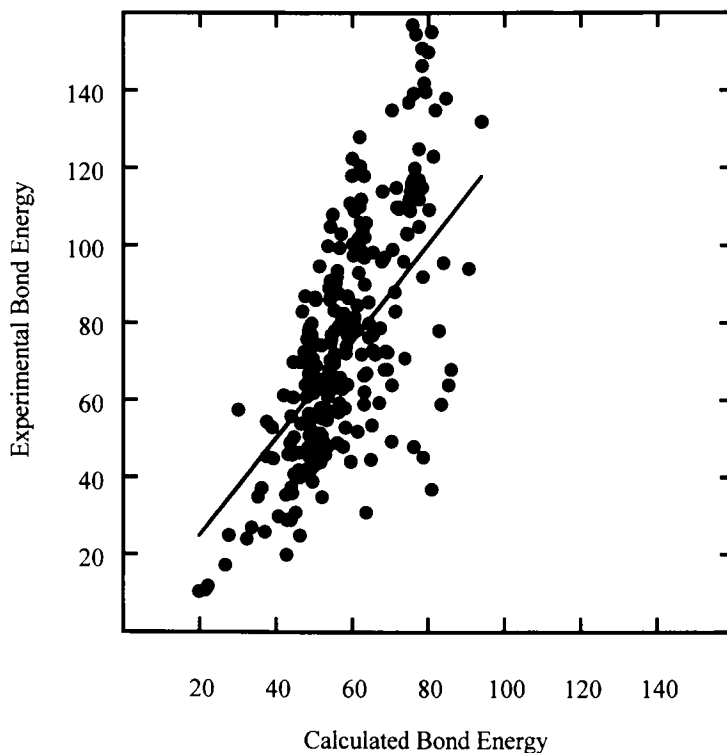


Figure 1. Plot of the experimental bond energy values of 249 systems versus the calculated values through Eq. (38) with $N_e = 1$ in kcal/mole. The straight line corresponds to Eq. (38) with $N_e = 1.12$ (the arithmetic average of the values of N_e obtained from Eq. (38) to reproduce the experimental bond energies).

with the experimental evidence. Equation (39) allows one to roughly determine which will be the weakest bond in a molecule, and therefore the most reactive site for each type of attack.

It is important to mention that if one makes use of the experimental values of I and A in Eq. (11), to determine the hardnesses in Eq. (39), and one makes use of molecular orbital theory to determine the values of the condensed fukui function, then, if one sets $N_e = 1$, one finds that this expression provides the correct trends, and reasonable estimates of the bond energies of a wide variety of molecular systems [16].

Thus, the numerical evidence shows that the main contribution to the bond energy is provided by the hardnesses difference between the initial state, when the atoms that are to be bound are very far away from each other, and the final state, when the bonded atoms arrive at the equilibrium distance.

Equation (39) may be particularly useful to describe the inherent chemical reactivity of a molecule, because it only depends on the isolated system values, and it allows one to establish which are the weakest bonds for different types of attacks.

3.2. Activation energies

In order to analyze the implications of Eq. (33) with respect to activation energies, let us consider a system that is composed of several molecules that react with each other. In order to determine the activation energy one needs to calculate the total energy difference between the initial state, when the interacting molecules are very far apart from each other, and the transition state, when all molecules are close to each other, and some bonds are being broken, while some new bonds are being formed. According to Eq. (33) this energy difference is given by [17]

$$\Delta E_{\text{act}} \approx \frac{1}{2} N_e^2 (\eta_i - \eta_{ts}) = \frac{1}{2} N_e^2 \left(\frac{S_{ts} - S_i}{S_i S_{ts}} \right) \quad (40)$$

Since $\Delta E_{\text{act}} \geq 0$, and E_{ts} is a maximum at the transition state, one can see that Eq. (40) implies that the softness of the system is also a maximum in the transition state, while the hardness is a minimum. This statement is in agreement with theoretical calculations that show that in the transition state, which corresponds to the less stable configuration of the system, the hardness attains its lowest value.

Now, assume that the softness of the initial state, when all the reactants are far away from each other, is roughly equal to the sum of the softnesses of the reactants when they are isolated from each other, and consider the case of a reaction between several molecules, in which only a specific atom of each one of the reactants participates directly in the bond breaking and bond forming processes. In this context, if it is assumed that the changes in all the other atoms of the reacting molecules can be neglected, one finds that

$$S_i \approx \sum S_r f_{rk} \quad (41)$$

where S_r is the softness of a given reactant, f_{rk} is the condensed fukui function of the k -th atom in the given reactant, and the sum is to be taken over all the reactants. Substituting Eq. (41) in Eq. (40), one can see that this approximation implies that the activation energy will be dominated by the softnesses of the atoms of the reactant molecules that participate directly in the reaction, when they are placed in the chemical environments provided by the reactant molecules, rather than by the global softnesses of the reactant molecules. In addition, one can see that this approximation also implies that the larger the values of the condensed fukui functions, the lower the activation energy barrier. Thus, one may conclude again that the most reactive sites of a molecule will be those with the largest values of the condensed fukui function, because they correspond to the weakest bond, and because they lead to the lowest activation energy barrier. It is important to remember that the largest values of the condensed

fukui function may be located at different sites for nucleophilic, electrophilic or free radical attacks.

In previous publications, it had already been inferred [8,9,21-26,43], that the larger the fukui function, the greater the reactivity, and this statement had already been successfully used to explain several aspects of the chemical reactivity of different systems. The present approach allows one to understand that this may be due to the fact that the sites with the largest values of the appropriate condensed fukui function may be associated with the weakest bonds, and with those reaction paths with the smallest activation energy barriers.

Now, in order to estimate the global softness of the transition state, one can make use of the arithmetic average principle for the softness of a system in terms of its constitutive parts, since it seems reasonable to assume that the softness of the transition state is proportional to the sum of the softnesses of all the molecular fragments that participate in the bond breaking and bond forming processes. However, in this case, the proportionality constant should take a value greater than the one corresponding to the arithmetic average principle, because in the transition state the molecular fragments are weakly bonded to each other, and therefore it represents a different situation to the one corresponding to the arithmetic average principle that describes the molecular softness in terms of the constitutive parts in the equilibrium position, when strong bonds have been formed. Thus, in general, the transition state softness may be written in the form [17]

$$S_{ts} \approx \alpha \sum S_{mf} f_{mfl} \quad (42)$$

where S_{mf} is the softness of a given molecular fragment, f_{mfl} is the condensed fukui function of the l -th atom in the given molecular fragment, and the sum is to be taken over all the molecular fragments in the transition state. The presence of the condensed fukui function in Eq. (42), implies that one only takes into account the changes in those atoms of the molecular fragments that participate directly in the bond breaking and bond forming processes. The proportionality constant α is expected to have a value that lies between one and one over the number of molecular fragments in the transition state (arithmetic average). A value of α around one is interpreted as if there were practically no bonding between the molecular fragments, while a value of α lower than one is interpreted as if there were a weak bonding between the molecular fragments. Thus, the proportionality constant provides a measure of the looseness of the transition state [44].

It is important to mention that if one makes use of the experimental values of I and A in Eq. (11), to determine the hardnesses in Eq. (40) in the form given by Eqs. (41) and (42), and if one sets $N_e = 1$ and $\alpha = 1$, one finds that this expression provides the correct trends, and reasonable estimates of the activation energies. Through this approach, one also finds that if α is determined to reproduce the experimental activation energy, one is led to values that correlate rather well with other theoretical estimates of the looseness of the transition state [17,44].

3.3. Reaction energies

The expressions derived for the bond energies and the activation energies may be used to analyze the behavior of reaction energies with respect to the changes in the hardness.

In the case of the bond energies, Eq. (36) may be applied to the calculation of reaction energies, if a chemical reaction is viewed as a bond breaking and bond formation process, because then one can determine the energy changes associated with the bonds broken, and the energy changes associated with the bonds formed. The summation of all the energy changes will be equal to the reaction energy. Thus, using Eq. (36), and assuming that $N_e \approx 1$, this procedure leads to

$$\Delta E_{\text{reac}} \approx -\frac{1}{2} (\Sigma \eta_p - \Sigma \eta_r) \quad (43)$$

where η_p is the hardness of a given product, and the sum is to be taken over all products, and η_r is the hardness of a given reactant, and the sum is to be taken over all reactants. In order to derive Eq. (43), it has been assumed that the sum of the terms that depend on the hardnesses of the molecular fragments that result from the bond breaking process, and the sum of the terms that depend on the hardnesses of the molecular fragments that give rise to the new bonds, approximately cancel each other, because the fragments for both cases are the same (see the example of an exchange reaction [16]).

Now, Eq. (43) implies that, $\Delta E_{\text{reac}} < 0$ if the sum of the hardness of the products is greater than the sum of the hardness of the reactants, and $\Delta E_{\text{reac}} > 0$ if it occurs the opposite. This statement is in complete agreement with the experimental evidence which shows that reactions almost always go in the direction that produces the hardest molecule, or the products of highest average hardness [45,46].

In the case of the activation energies, the reaction energy may be determined from the difference between the activation energy corresponding to the reaction in the direction of reactants to products, and the activation energy corresponding to the reaction in the direction of products to reactants. Thus, if $N_e \approx 1$, and the condensed Fukui functions are set equal to one, according to Eqs. (40)-(42), one finds that [17]

$$\Delta E_{\text{reac}} = \Delta E_{\text{act}}^{r \rightarrow p} - \Delta E_{\text{act}}^{p \rightarrow r} \approx 1 / \Sigma S_r - 1 / \Sigma S_p \quad (44)$$

Therefore, Eq. (44) implies that, $\Delta E_{\text{reac}} < 0$ if the sum of the softness of the products is lower than the sum of the softness of the reactants, and $\Delta E_{\text{reac}} > 0$ if it occurs the opposite. This conclusion is in agreement with the one derived from Eq. (43). However, in this case, since the sum is taken over the molecular softnesses, one is led to an harmonic average of the hardnesses, instead of the arithmetic average found in the bond energy analysis. In general, both average values will lead to the same results, and provide a strong support to the statement that reactions tend to go in the direction that produces the hardest possible species.

4. CATALYZED REACTIONS AND REACTIONS IN SOLUTION

It is interesting to analyze the implications of the relations derived in the previous section in the cases of catalyzed reactions, and of reactions in solution. In both cases it seems reasonable to assume that the main effect of the catalyst or the solvent is to modify the values

of the condensed Fukui functions of the molecules, at least in their initial state. In this situation, if the effect of the catalyst or the solvent is to increase the values of the condensed Fukui functions at the sites of the molecules where the reaction occurs, in the absence of the catalyst or the solvent, then, if the reacting site remains being the same one, the bond will be weakened, and the activation energy will be decreased, facilitating the reaction. On the other hand, if the effect of the catalyst or the solvent is to decrease the values of the condensed Fukui functions at the sites of the molecules where the reaction occurs, in the same circumstances as above, then the bond will be strengthened and the activation energy will be increased, complicating the reaction. If the effect of the catalyst or the solvent is to change the site at which the maximum values of the Fukui functions occur, then, the reaction path and the products will be different to those of the reaction in the absence of the catalyst or the solvent. These statements are in agreement with possible effects that have been experimentally observed when one compares a reaction in the gas phase, with the corresponding catalyzed reaction, or the corresponding reaction in solution. For example, it is known that in hydrogenation reactions over metal catalysts, the bond of the hydrogen molecule weakens when it is adsorbed on the surface, facilitating the reaction. In the present context, the weakening of the bond may be associated with an increment of the condensed Fukui function, that will imply a lowering of the activation energy barrier for the reaction with any molecule which is to be hydrogenated.

Thus, one can see that it seems reasonable to assume that the main effect of a catalyst or a solvent is to soften or to harden the reacting molecules, and because of Eqs. (40) and (41), these modifications have a direct effect on the activation energy barrier, even if the transition state adopts the same structure than the one adopted by the reacting molecules in the absence of a catalyst or a solvent.

Consider now the general effect of a catalyst in a reaction. In the case of homogeneous catalysis, the catalyst may be considered as an additional reactant that interacts with the rest of the reactants, but it remains unchanged after the reaction takes place, while in the case of heterogeneous catalysis, the catalyst is, in general, the surface over which the reaction takes place, and it remains basically unchanged along the reaction energy path. In both cases, one has to take into account the presence of the catalyst in the initial and in the transition states. In the initial state, when all the reactant molecules are very far away from each other, one may assume that the catalyst is also very far away from the rest of the reacting molecules, and therefore, according to the additivity of softness, one only needs to add the catalyst softness, S_{cat} , to determine the total softness of the initial state. In the transition state, the catalyst may be considered as an additional molecular fragment and, therefore, one needs to add the catalyst softness to determine the total softness of the transition state. This way, the activation energy in a catalyzed reaction will be given by

$$\Delta E_{\text{act}}^{\text{cat}} \approx -\frac{1}{2} N_e^2 \left(\frac{1}{S_{\text{ts}} + S_{\text{cat}}} - \frac{1}{S_i + S_{\text{cat}}} \right) \quad (45)$$

where it has been assumed that, in the transition state, the catalyst is weakly bonded to the reacting molecules, so that $\alpha \approx 1$. Equation (45) implies, in the present context, that the main effect of a catalyst is to soften the system along the reaction energy path, independently of the effect that it may have directly on the softness of the reacting molecules.

It is interesting to compare the activation energy of the uncatalyzed reaction, with respect to the activation energy of the catalyzed reaction. If one determines the difference between Eqs. (45) and (40), one finds that

$$\Delta E_{\text{act}}^{\text{cat}} - \Delta E_{\text{act}} \approx \frac{1}{2} N_e^2 \left[\frac{S_{\text{cat}} (S_{\text{ts}} + S_i + S_{\text{cat}}) (S_i - S_{\text{ts}})}{(S_{\text{ts}} + S_{\text{cat}}) (S_i + S_{\text{cat}}) S_{\text{ts}} S_i} \right] \quad (46)$$

First, one may note that since $S \geq 0$, and since the softness of the transition state will be, in general, greater than the softness of the initial state (because lower softness implies greater hardness), this, in turn, implies greater stability [47]. This means that the system in the initial state is more stable than in the transition state, then $S_i - S_{\text{ts}} \leq 0$, and therefore, $\Delta E_{\text{act}}^{\text{cat}} - \Delta E_{\text{act}} \leq 0$, that implies that the consequence of the softening effect of the catalyst is to lower the activation energy barrier. Secondly, one can see from Eq. (46) that the lowering of the activation energy increases when the softness of the catalyst increases, approaching a constant value when S_{cat} is very large, a result that implies that the softer the catalyst, the better. Both statements are in agreement with the experimental evidence, catalysts lower the activation energy, and noble metals, that are rather soft, are good catalysts.

5. CONCLUDING REMARKS

The overall analysis presented in this work shows that through the use of the density functional theory framework one may establish a simple unified approach to bond energies, activation energies, and reaction energies, in terms of hardness differences.

According to the analysis, and the results presented here, one may conclude that Eq. (33) for the energy changes, accounts for the main contributions that are involved in the interaction between different chemical species, and therefore it may be used to study trends and to rationalize the behavior of molecules under different circumstances.

On the other hand, it seems that the additivity of softness assumption may provide important information about the transition state structure, and about the reaction mechanism, because if the experimental activation energy is known, one can determine the looseness of the transition state, and also, among several possibilities, one may select the molecular fragments that provide the best description.

In particular, I believe that the present approach may be very useful in organic reactions, because the differences in the chemical potential (electronegativity) of many organic compounds are very small. This situation implies that, in general, one may expect that the chemical potential difference term will be negligible, in comparison with the hardness difference term, and therefore, Eq. (33) may be particularly appropriate to understand the chemical reactivity of organic molecules.

Acknowledgments. I would like to thank A. Vela and L.I. Rangel for many valuable discussions.

REFERENCES

1. R.T. Sanderson, *Chemical Bonds and Bond Energy*, 2nd ed., Academic Press, New York, NY, 1976.
2. R.G. Pearson, *Hard and Soft Acids and Bases*, Dowden, Hutchinson and Ross, Stroudsville, PA, 1973.
3. K. Fukui, *Theory of Orientation and Stereoselection*, Springer, Berlin, 1973.
4. G. Klopman (ed.), *Chemical Reactivity and Reaction Paths*, Wiley, New York, NY, 1974.
5. R.G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford, New York, NY, 1989.
6. R.G. Parr, R.A. Donnelly, M. Levy and W.E. Palke, *J. Chem. Phys.*, 68 (1978) 3801.
7. R.G. Parr and R.G. Pearson, *J. Am. Chem. Soc.*, 105 (1983) 7512.
8. W. Yang and R.G. Parr, *Proc. Natl. Acad. Sci. USA*, 82 (1985) 6723.
9. R.G. Parr and W. Yang, *J. Am. Chem. Soc.*, 106 (1984) 4049.
10. R.G. Pearson, *J. Chem. Ed.*, 64 (1987) 561.
11. R. G. Parr and P.K. Chattaraj, *J. Am. Chem. Soc.*, 113 (1991) 1854.
12. J.L. Gázquez, A. Martínez and F. Méndez, *J. Phys. Chem.*, 97 (1993) 4059.
13. R.G. Parr and J.L. Gázquez, *J. Phys. Chem.*, 97 (1993) 3939.
14. P.K. Chattaraj, H. Lee and R.G. Parr, *J. Am. Chem. Soc.*, 113 (1991) 1855.
15. J.L. Gázquez, *J. Phys. Chem.*, in press (1997).
16. J.L. Gázquez, *J. Phys. Chem.*, submitted (1997).
17. J.L. Gázquez, *J. Phys. Chem.*, submitted (1997).
18. R.G. Parr and W. Yang, *Annu. Rev. Phys. Chem.*, 46 (1995) 701, and references therein.
19. W. Kohn, A.D. Becke and R.G. Parr, *J. Phys. Chem.*, 100 (1996) 12974, and references therein.
20. R.G. Pearson, *Coord. Chem. Rev.*, 100 (1990) 403, and references therein.
21. C. Lee, W. Yang and R.G. Parr, *J. Mol. Struct. (Theochem)*, 163 (1988) 305.
22. W. Langenaeker, M. De Decker and P. Geerlings, *J. Mol. Struct. (Theochem)*, 207 (1990) 115.
23. W. Langenaeker, K. Demel and P. Geerlings, *J. Mol. Struct. (Theochem)*, 234 (1991) 329.
24. W. Langenaeker, K. Demel and P. Geerlings, *J. Mol. Struct. (Theochem)*, 259 (1992) 317.
25. F. Méndez and J.L. Gázquez, *J. Am. Chem. Soc.*, 116 (1994) 9298.
26. F. Méndez and J.L. Gázquez in *Theoretical Models for Structure, Properties and Dynamics in Chemistry*, S.R. Gadre (ed.), *Proceedings of Indian Academy of Sciences (Chem. Sci.)*, 106 (1994) 183.
27. R.F. Nalewajski, *J. Chem. Phys.*, 78 (1983) 6112.
28. J.L. Gázquez in *Chemical Hardness*, K.D. Sen (ed.), *Structure and Bonding*, 80 (1993) 27.
29. M. Berkowitz and R.G. Parr, *J. Chem. Phys.*, 88 (1988) 2554.
30. W. Yang and W.J. Mortier, *J. Am. Chem. Soc.*, 108 (1986) 5708.
31. R.G. Parr, S. Liu, A.A. Kugler and A. Nagy, *Phys. Rev. A*, 52 (1995) 969.
32. S. Liu and R.G. Parr, *J. Chem. Phys.*, 106 (1997) 5578.
33. R.G. Pearson and W.E. Palke, *J. Phys. Chem.*, 96 (1992) 3283.
34. D. Datta, *J. Phys. Chem.*, 96 (1992) 2409.
35. P.K. Chattaraj, S. Nath and A.B. Sannigrahi, *Chem. Phys. Lett.*, 212 (1993) 223.
36. S. Pal, N. Vaval and R.K. Roy, *J. Phys. Chem.*, 97 (1993) 4404.

37. P.K. Chattaraj, S. Nath and A.B. Sannigrahi, *J. Phys. Chem.*, 98 (1994) 9143.
38. S. Pal, A.K. Chandra and R.K. Roy, *J. Mol. Struct. (Theochem)*, 361 (1996) 57.
39. S. Pal, R.K. Roy and A.K. Chandra, *J. Phys. Chem.*, 98 (1994) 2314.
40. R.F. Nalewajski, J. Korchowiec and Z. Zhou, *Int. J. Quantum Chem.*, S22 (1988) 349.
41. W. Yang, C. Lee and S.K. Ghosh, *J. Phys. Chem.*, 89 (1985) 5412.
42. J.E. Huheey, E.A. Keiter and R.L. Keiter, *Inorganic Chemistry Principles of Structure and Reactivity*, 4th ed., Harper Collins College, New York, NY, 1993.
43. J.L. Gázquez and F. Méndez, *J. Phys. Chem.*, 98 (1994) 4591.
44. S.S. Shaik, H.B. Schlegel and S. Wolfe, *Theoretical Aspects of Physical Organic Chemistry*, Wiley, New York, NY, 1992.
45. D. Datta, *Inorg. Chem.*, 27 (1992) 2797.
46. R.G. Pearson, *Inorganica Chimica Acta*, 198-200 (1992) 781.
47. T.K. Ghanty and S.K. Ghosh, *J. Phys. Chem.*, 100 (1996) 12295.

Molecular geometry as a source of chemical information for π - electron compounds

Tadeusz Marek Krygowski and Michał Ksawery Cyrański*

Department of Chemistry, University of Warsaw,
L. Pasteura 1, 02-093 Warsaw, Poland

Abstract

After a short introduction presenting a general view on the title problem and after information about more general approaches such as e.g. the Bürgi and Dunitz Principle of structural correlation a series of empirical models applying the molecular geometry are shown as methods allowing us to read the molecular geometry in the language of chemistry. First, a model enabling estimation of the heat of formation from CC bond lengths for hydrocarbons is presented. The ring energy content calculated by this model shows that it may vary considerably depending on different reasons: the topological embedding in benzenoid hydrocarbons, topology of substitution and the nature of substituents in polysubstituted benzene derivatives, and intermolecular interactions in the crystalline state. Secondly, the model enabling estimation of the canonical structure weights is presented (HOSE model). Based on this model it is shown how geometry of the benzene ring may reflect the substituent effect in terms of varying weights of canonical structure. The non-mesomeric interaction of the nitro group with the ring in nitrobenzene is shown and carefully documented. A new substituent effect due to angular groups (AGIBA-effect) is best illustrated by use of the canonical structure weights calculated from molecular geometry. Next we present the application of the Bent-Walsh rule which shows how intramolecular and intermolecular interactions associated with the charge transfer cause deviations from this rule. Finally, application of the molecular geometry to determine indices of aromaticity is presented. The most recent and interesting result in this field is that the index HOMA based solely on the experimental (or computed) bond lengths may be divided into two independent terms, of which one accounts for the energetic and the other for the geometric contributions to aromaticity. The first one is related to resonance energy of the ring in question, the second one - to the bond length alternation. A few illustrative applications are presented.

* Stipendiarius of the Foundation for Polish Science

INTRODUCTION

The importance of the information hidden in a molecular geometry is best expressed by R. Hoffmann [1] who wrote in the foreword to the monograph on determination of molecular geometry:

“There is no more basic enterprise in chemistry than the determination of the geometrical structure of a molecule. Such a determination, when it is well done, ends all speculation as to the structure and provides us with the starting point for understanding of every physical, chemical and biological property of the molecule.”

Undoubtedly, the molecular geometry itself does not always provide all this information as such. Most often it must be transformed into the proper models which translate geometry parameters into the appropriate language explaining a given physical, chemical or biological property.

Molecular geometry is accessible either by use of experimental techniques of measurements or by applying theoretical methods of calculations. Most popular are X-ray diffraction measurements which at present are being done routinely and the results of which are archived in Cambridge Structural Database (*circa* 160,000 crystal and molecular structures of organic compounds in 1996) [2] and Inorganic Structural Data Base (*circa* 50,000 structures in 1996) [3]. For simpler and symmetrical molecules, also electron diffraction studies are very useful (e.g. [1]), as well as very precise but time-consuming measurements by use of microwave spectroscopy [4]. A very nice review on successful applications of *ab initio* calculations in chemistry is given in the monograph [5]. However it should be mentioned that, as any other methods, *ab initio* techniques in calculating optimized molecular geometry are biased by a systematic error which depends on bond length [6].

Molecular geometry has a countless number of individual applications in interpretation of the chemical and physicochemical properties of chemical species. In this review we will deal only with interpretations which result from some general models based on molecular geometry of π -electron systems and which can be easily applied in other cases.

It should be emphasized here that a very fruitful and powerful interpretation of chemical properties deduced from the molecular geometry results from the application of the so-called Principle of Structural Correlation invented and developed some 25 years ago by Hans-Beat Bürgi, Jack Dunitz and their colleagues [7-11]. The main idea of this principle is as follows: **if some geometric parameters of the molecule (or its fragment) in question are subjected to some perturbations which may be caused by either intermolecular or intramolecular interactions due to the varying molecular or crystal environment they may be mutually interrelated.** The principle allows one to show that there is a way from the crystal structure data to the chemical reaction path. Since these ideas have recently been reviewed in detail [12] by the original

authors and their colleagues who have also made a further development in the field, we are not going to repeat those reviews but we refer our reader directly there.

1. HEAT OF FORMATION DERIVED FROM THE MOLECULAR GEOMETRY: THE BOND ENERGY DERIVED FROM CC BOND LENGTHS

It is a trivial observation that the bond energy depends on the value of bond length. This fact was used in semiempirical models of quantum chemistry to estimate the value of resonance integral [13]. Recently the empirical model of estimating bond energy from bond length was presented for the systems built up of CC bonds [14].

Pauling proposed [15] a fractional bond number, n , defined as:

$$R(n) - R(1) = -c \cdot \ln(n) \quad (1)$$

where $R(n)$ and $R(1)$ are the bond lengths for which the bond numbers are equal to n and 1, respectively, whereas c is an empirical constant. This idea has been used in structural chemistry many times and proved extremely successful in interpretation of various chemical problems [8,11,16]. Another empirical rule relates the bond energy $E(n)$ to the bond number n [17]:

$$E(n) = E(1) \cdot n^p \quad (2)$$

where $E(1)$ and $E(n)$ represent the energies of bonds with bond numbers equal to 1 and n , respectively. Combination of (1) and (2) leads to the expression for bond energy dependent solely on the bond length $R(n)$ [14]:

$$E(n) = E(1) \cdot \exp\left\{\alpha \cdot [R(1) - R(n)]\right\} \quad (3)$$

where $\alpha = p/c$. The parameters for estimation of bond energy for any system built up of the CC bonds are given below:

$$R(1) = 1.533 \text{ \AA} \quad [18] \quad (4a)$$

$$R(2) = 1.337 \text{ \AA} \quad [19] \quad (4b)$$

$$E(1) = 94.66 \text{ kcal/mol} \quad [20] \quad (4c)$$

$$E(2) = 131.91 \text{ kcal/mol} \quad [20] \quad (4d)$$

Applying the precise geometry of nine benzenoid hydrocarbons, eq. (3) and parameters defined by eqs. (4a-d) allowed us to obtain the quantities which are comparable with the experimental values of heats of formation, hereafter abbreviated to HF. In order to do so, it was necessary to take into account the C-H bond energies. Their bond lengths are not accessible from X-ray diffraction measurements, thus we assumed that each bond might be treated as having an additive energy equal to 100.53

kcal/mol [20]. Addition of CH bond energies to the energy calculated by summing up the bond energies of all CC bonds calculated by use of equation (3) leads to the value of energy which may be compared with the experimental value of the heat of formation. Then we scaled the bond energy, $E(\text{C-C})$, in equation (3) in order to reproduce the heat of formation of benzene, and obtained $E(\text{C-C})$ equal to 87.99 kcal/mol. Thus the final formula for calculating heat of formation (HF) may be written as follows:

$$\text{HF} = 100.53 \cdot n - 87.99 \sum_{i=1}^N \exp\left[2.255 \cdot (1.533 - R(i))\right] \quad (5)$$

where N and n are numbers of CC and CH bonds, respectively.

Table I presents a comparison of experimental HF's and the values calculated from bond lengths and formula (4). The small differences are at the level of error of experiment.

Table I. Estimated and experimental values of heat of formation form atoms in kcal/mol [14].

Compound	estimated HF (eq. 5)	experimental HF [21]
Benzene	1320.6	1320.6
Naphthalene	2100.8	2093.8
Anthracene	2868.4	2863.9
Phenanthrene	2861.8	2869.5
Tetracene	3592.4	3638.8
Chrysene	3664.4	3643.9
Triphenylene	3647.3	3641.2
3,4-Benzophenanthrene	3646.6	3638.8
Pyrene	3227.8	3207.7

1.1. Energy content of individual phenyl rings in various topological and chemical embedding

It is intuitively obvious that benzene rings may have different energy content depending on the kind of the closest environment. It is well known that the resonance energies of benzene, naphthalene and anthracene (and other benzenoid hydrocarbons) do not follow any additivity rule. Thus the individual rings in these systems may have lower resonance energy than in benzene itself. Moreover, it cannot be excluded that the different rings in the same molecule may differ in their energy content. Problems of this kind have so far been rather difficult to tackle. Since the method proposed above provides encouraging agreement between the estimated and experimental heats of formation, it seems to be ideally suited to be applied to study these kind of problems. For each benzene ring the above-presented equations (taking only CC bonds but not

the CH bonds) may be used to calculate the sum of all CC bonds building up a given ring in question, giving the quantity hereafter called Ring Energy Content, abbreviated to REC:

$$\text{REC} = 87.99 \sum \exp[2.255 \cdot (1.533 - R(i))] \quad (5a)$$

1.2. Ring Energy Content of benzene rings in benzenoid hydrocarbons

The above-presented treatment of experimentally accessible bond lengths allows us to estimate bond energies for any molecule or its fragment, provided it is built up of CC bonds and their lengths are precisely measured. Figure 1 presents a few benzenoid hydrocarbons in which REC values are calculated for each individual ring [14]. It is very striking that the REC values vary considerably: from 724.6 kcal/mol for benzene to 668.9 kcal/mol for the central ring in triphenylene or even 648.1 kcal/mol for the central ring in perylene. It is worth mentioning that more reactive central rings of phenanthrene and pyrene have also low REC values: 690.4 and 696.5 kcal/mol, respectively.

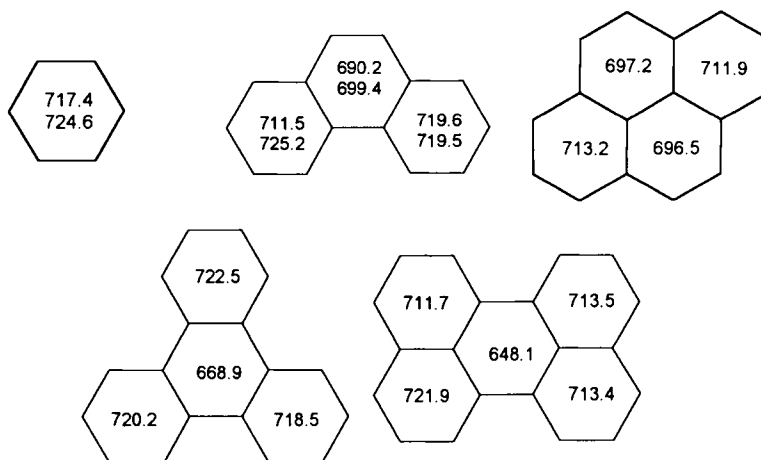


Fig. 1. REC values of individual benzene rings in some benzenoid hydrocarbons

Two different values of REC for particular rings of benzene and phenanthrene result from different measurements of their molecular geometry. The different REC values for symmetrically equivalent rings in other benzenoid hydrocarbons result from the method of X-ray structure determination. If the symmetrical molecule does not lie at the symmetry element in the crystal lattice, each of the rings is measured independently and its geometry (and in turn its REC value) is biased by the another error of measurement. The differences in REC values for symmetrically equivalent

rings in the molecules of phenanthrene, pyrene, triphenylene and perylene in Fig. 1 may serve as a natural visualization of the precision of estimating the REC values. The mean deviation from the calculated means for all symmetrically equivalent rings is 2.7%.

An interesting observation is that benzene rings which are terminals in benzenoid hydrocarbons fused by one or two joint CC bonds to the rest of the molecule, exhibit REC values close to that found for benzene itself. Moreover, it is apparent that the REC values of the ring depend strongly on the topological environment of the ring in question. This finding is in line with the conclusions drawn in the aromaticity studies of benzenoid hydrocarbons [22] which evidenced that the topology is the very (even the most) important factor in determining the aromatic character of the ring in question. The low values of REC are equivalent to the low values of resonance energy. Figure 2 presents REC values for individual rings in a complex benzenoid hydrocarbon: tribenzopnenenthrapentaphene [23]. The great diversity of REC values and strong dependence of their values on the topological embedding is well illustrated there and shows that the phenyl rings in the same molecule may differ to a large extent according to its energy content.

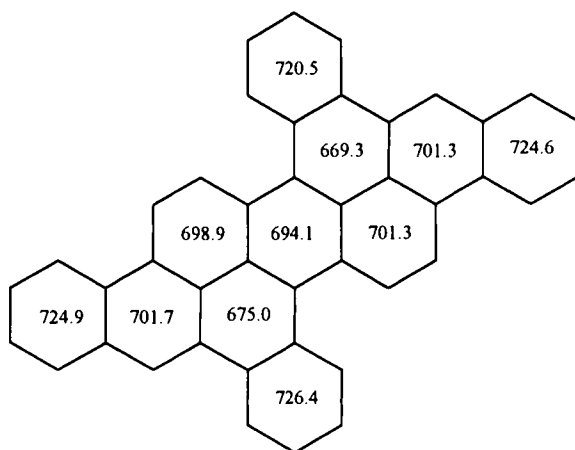


Fig. 2. The Ring Energy Contents for particular benzene rings in tribenzopnenenthrapentaphene.

The strong dependence of REC values of the ring on the topological environment in the molecule is presented in Fig. 3a as a distribution of REC values for 169 rings of 26 benzenoid hydrocarbons (data taken from [22]). This distribution may be compared with another one, which, from the topological point of view, is very similar. This is a case of polysubstituted benzene derivatives. Instead of CC or CH bonds which form the closest environment of a given ring in the benzenoid hydrocarbon, one has CX or CH bonds, where X is a substituent. In order to study how these two chemically different situations may affect REC values we have presented an analogical distribution for 2045 polysubstituted benzene derivatives (from mono- to hexa-

substituted benzene derivatives). As it is clearly seen it gives a much less dispersed distribution (Fig. 3b).

The mean REC value for the rings in benzenoid hydrocarbons differs markedly from that for polysubstituted benzenes, which are 700.1 kcal/mol and 729.8 kcal/mol, respectively. It should be mentioned here that this large difference may be, at least partly, due to the systematic error in determining the geometry of substituted benzene derivatives, which may exhibit considerable thermal motion leading to the shortening of bond lengths and, in turn, to an increase in their energies [4]. The shape of their distributions also differs quite strongly: the variances are 399.4 and 285.6 (kcal/mol)². Since the distributions in question, at least the one in Fig. 3a, do not seem to be normal, the more proper way of describing dispersion of the data is the interquartile range (see caption to the Fig.3), which is much greater for the benzenoid hydrocarbons histogram than for the other one.

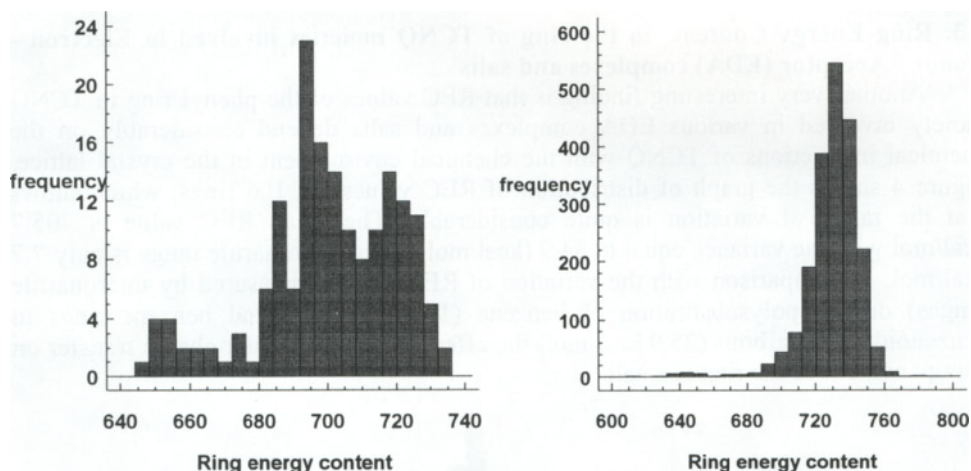


Fig. 3a-b. Distribution of Ring Energy Content for benzenoid hydrocarbons and polysubstituted benzene derivatives. The interquartile ranges for the distribution are 25.9 kcal/mol and 16.1 kcal/mol, respectively.

These two situations differ in the kind of interactions between the ring in question and its closest environment. In the case of benzenoid hydrocarbons the possible bonds linked to this ring are either CH or CC. In the case of polysubstituted benzene derivatives they are either CH or CX, where X is the first atom of the substituent. Chemically these situations differ considerably. The variation in electronegativity of atoms (either carbon or hydrogen) in benzenoid hydrocarbons is negligible in comparison with that which is found in the polysubstituted derivatives of benzene. In this case a variety of substituents attached to the ring may differ in their electronegativities and hence be able to induce strong mesomeric interactions with or even through the ring. In these cases, this justifies us to apply the term chemically perturbed benzene rings whereas in the case of benzenoid hydrocarbons we accept the term topologically perturbed benzene rings.

From the above results it may be concluded that the topological effects of the environment on the geometry of a given ring affect the REC values more strongly than the chemical ones - which is in line with the previously observed changes of aromaticity indices HOMA [24, 25] and its independent components GEO and EN [26] accounting for the geometric and energetic contribution to the aromatic character of the ring, respectively. It was found [22] that the EN term varies much less in the case of chemically perturbed rings than in those which were perturbed topologically and this observation is in line with what we observe for distributions of the REC values.

Another conclusion may also be drawn. This strong dependence of REC values (as well as formerly stated EN and HOMA values) of benzene rings in the benzenoid hydrocarbons on topology may be of a key importance for understanding of the successful treatments of benzenoid hydrocarbons by the so-called graph-topological models [27].

1.3. Ring Energy Content in the ring of TCNQ moieties involved in Electron - Donor - Acceptor (EDA) complexes and salts

Another very interesting finding is that REC values of the phenyl ring in TCNQ moiety involved in various EDA complexes and salts depend considerably on the chemical interactions of TCNQ with the chemical environment in the crystal lattice. Figure 4 shows the graph of distribution of REC values for 106 rings, which shows that the range of variation is quite considerable. The mean REC value is 705.7 kcal/mol with the variance equal to $34.9 \text{ (kcal/mol)}^2$. The interquartile range is only 7.7 kcal/mol. In comparison with the variation of REC values (measured by interquartile ranges) due to polysubstitution of benzene (16.1 kcal/mol) and benzene rings in benzenoid hydrocarbons (25.9 kcal/mol) the effect of intermolecular charge transfer on energy content of the ring is small.

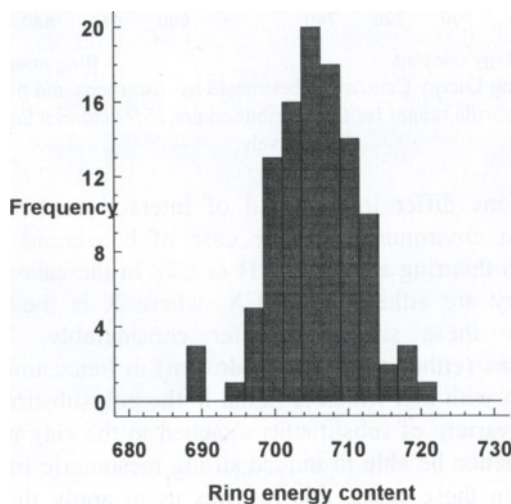


Fig. 4. Distribution of Ring Energy Content for TCNQ moieties in 106 EDA complexes and salts. The interquartile range is 7.7 kcal/mol.

Figure 5 presents the dependence of REC values on the charge transferred from the donating component of the complex and/or salt to the TCNQ moiety (estimated by use of the procedure described in [28]). It is clear that the increase of charge transfer

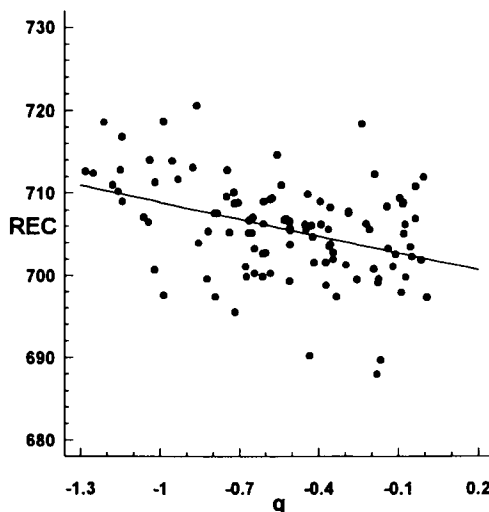


Fig. 5. Dependence of REC values for phenylic ring in TCNQ involved EDA complexes and in salts plotted against the charge transferred onto the TCNQ moiety.

from the donating component to TCNQ moiety results in an increase of energy content of the ring in question. The regression line is

$$\text{REC} = -6.85q + 702.1 \quad (6)$$

and from the slope we find that the increase of charge of the TCNQ moiety by one electron leads to the gain of energy approximately equal to 6.8 kcal/mol. Due to a rather low value of the correlation coefficient, $r = -0.4$ (significant at 0.001) this quantity has only a very qualitative meaning. Nevertheless it may be said that this increase of energy is equivalent to an increase of resonance energy of the ring in question.

1.4. Ring Energy Content depending on the intermolecular H-bonding: the case of p-nitrosophenolate anion

An example of p-nitrosophenolate anion is very spectacular indeed. In various salts this anion is differently hydrated and this is a cause of variation of the ring geometry and consequently of the energy content of the ring. Figure 6 shows the relevant data for p-nitrosophenolates of sodium (trihydrate) [29], magnesium (hexahydrate) [30], and lithium (dihydrate) [31].

Apart from REC values of the ring of p-nitrosophenolate moiety, there are given short contacts (in A) between the terminal basic atoms of nitroso- and oxo- groups and

H-atoms of the surrounding water molecules, and additionally, the H...O interaction energies estimated by use of a simple formula based on the same principles as equations (3-5) and presented in [32,33]. See also paragraph 1.6.

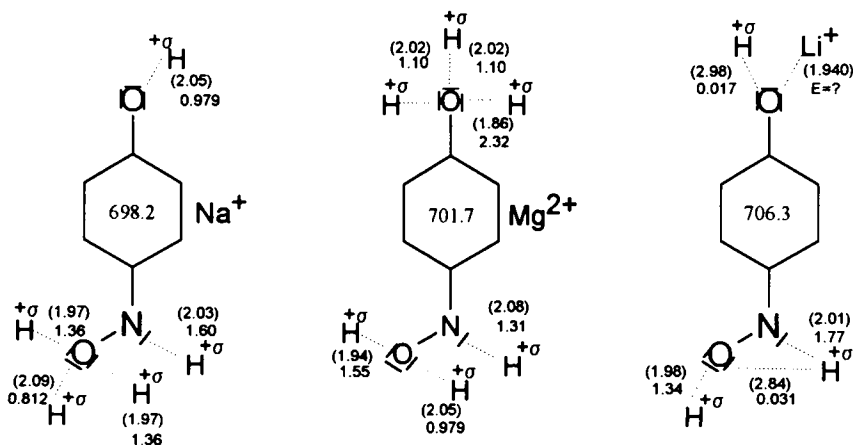


Fig. 6. Structures of p-nitrosophenolate derivatives with depicted REC values, the approximate energies of H...O interactions, and the closest contacts (in parentheses).

1.5. Ring Energy Content as a quantitative measure of fulfilling the Hückel $4n+2$ rule for derivatives of fulvene and heptafulvene

Quite a dramatic change in energy content of the ring is observed due to a substituent effect in exocyclically substituted derivatives of fulvene and heptafulvene. Figure 7 presents relevant data. Energy content of the ring increases from 841 kcal/mol for heptafulvene to 849.1 kcal/mol for its 8,8-diformyl derivative in which the Hückel rule requires charge transfer from the ring onto the electron-accepting formyl groups. In the case of fulvene, there is no experimental geometry of it, but from the mean geometry of the ring calculated for 11 derivatives substituted by groups not donating electrons the REC value is 586.7 kcal/mol compared to the 6,6-tetramethylamino derivative with REC = 597.3 kcal/mol. In line with the Hückel rule, the substituents which interact with the ring *via* charge transfer stabilize it, and REC values may represent this stabilization numerically. It is worth mentioning that the above-shown changes of REC values are well related with the variation of aromaticity of the nonalternant rings in questions [34].

How strong can be the dependence of the REC values of a given chemical moiety on the chemical interactions with the closest environment is shown in Fig. 8a-b presenting the histogram of REC values for cyclopentadienyl ring in its complex compounds with Fe (ferrocenes and related compounds) [35]. The two cases represent histograms of complexes with Cr (104 entries) and Fe (770 entries). The mean REC

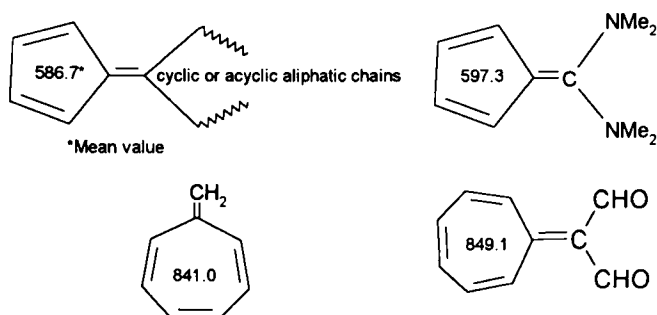


Fig. 7. Ring energy contents of fulvene, heptafulvene and their derivatives.

values differ significantly, being 602.0 and 592.2 kcal/mol, respectively. Very large values of standard deviations, 26.6 and 24.3 kcal/mol, respectively, make it clear that the interactions between cyclopentadienyl rings and the central metal ion may be very differentiated.

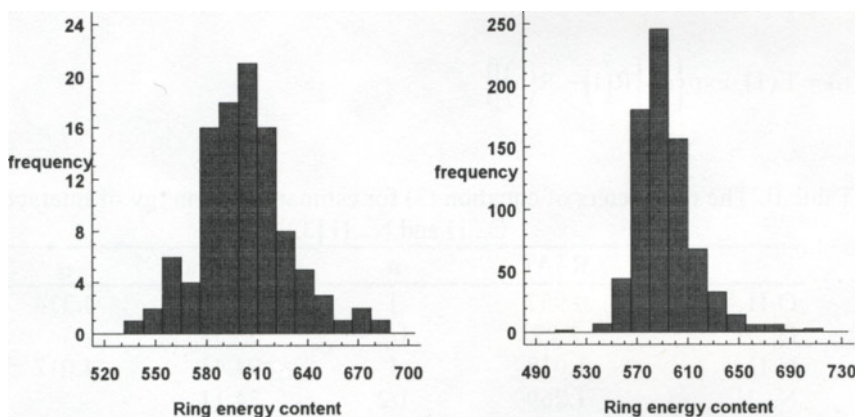


Fig. 8a-b. Histograms of REC values for cyclopentadienyl complexes with Cr and Fe, respectively. The interquartile ranges are 30.95 and 26.3 kcal/mol, respectively (see text).

The above-presented illustration shows how much one can do by going deeper inside with a much detailed structural analysis trying to rationalize these kinds of effects which are by no means very significant. Evidently, for each individual situation with a differently bound ring the resulting REC value may be very informative.

1.6. Estimation of H...O and H...N energy of interactions in H-bonds

Crystallographic literature is full of information on the so-called close interatomic contacts, which in the case of distances between H-atoms and electron donating atoms are used as a criterion of H-bond existence. If this distance is shorter than the sum of van der Waals radii - then the H-bond type interaction is assumed

[36,37]. Geometric criteria are very simple and convenient, but only very approximately informative on the energetics of interactions, particularly in more complex situations (bent H-bonding, bifurcating or other complex situations). There exist well accepted theories of H-bonds based on their geometry patterns, as the best known Lippincott-Schröder model [38] with parametrization for OH...O bonds by Derrisen and Smit [39]. However, they are not often applied because a considerable amount of detailed structural information is necessary to do so. Therefore it was suggested [32] to estimate the H...O and H...N energy of interaction directly from the respective interatomic distances using a similar procedure as that for estimating the CC bond energy (equations 1-3) [32]. The Pauling concept of bond number (equation 1) was applied to O-H bond distance in water as for a single OH bond with $n=1$, and to the O...H interatomic distance in $(\text{H}_2\text{O}\dots\text{H}^+\dots\text{OH}_2)$ as for a OH bond with bond number $1/2$. The N-H in ammonia and NH interatomic distance in $\text{H}_3\text{N}\dots\text{H}^+\dots\text{NH}_3$ were taken as bond lengths for NH with the bond number equal to 1, and $1/2$, respectively. These interatomic distances are associated with the respective bond energies. Table II presents the numerical data. The final formula (3) when used with the data from Table II, allows us to estimate the energy of interaction from the respective O...H and N...H interatomic distances.

$$E(n) = E(1) \cdot \exp\left\{\alpha \cdot [R(1) - R(n)]\right\} \quad (3)$$

Table II. The parameters of equation (3) for estimation of energy of interaction for O...H and N...H [32]

	R [Å]	n	E [kcal/mol]	α
O-H	0.957	1	110.6	4.334
O...H	1.22	1/2	35.38	
N-H	1.018	1	94.41	4.012
N...H	1.269	1/2	34.11	

The energies of H...O and H...N interaction in Figs 9 and 10 were estimated in this way. It is worth mentioning that the energies of O...H interactions estimated from the geometry of neutron diffraction measurements for 43 H-bonded systems obtained by use of the Lippincott and Schröder model and by eq. (3) lead to the qualitatively equivalent results; the correlation coefficient for this dependence is $r = 0.998$ [33].

This model turned out to be very useful for interpretation of the properties of the systems with intramolecular O-H ...O bonds. Both the location of the IR $\nu(\text{OH})$ band in IR and the chemical shift $\delta(\text{H})$ in H NMR are often used as empirical indicators of the hydrogen bond. Both of them correlate well with the O...H interaction energy, as shown in Figs 9 and 10.

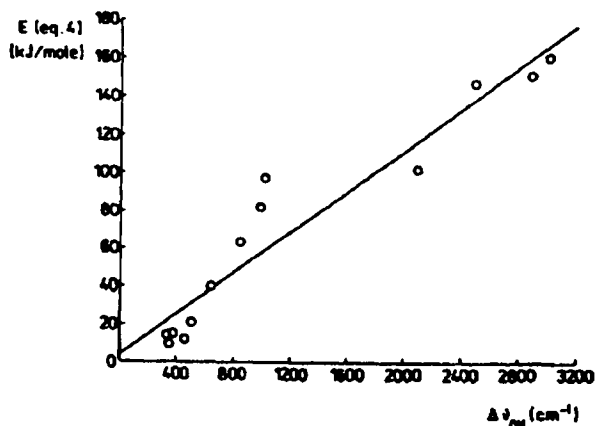


Fig. 9. Relationship between the energies of O...H interaction estimated using equation (3) and IR $\nu(\text{OH})$ band location. $1J = 0.2389$ cal.

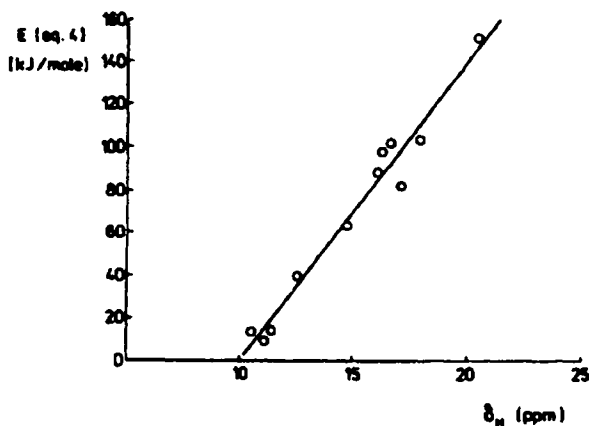


Fig. 10. Relationship between the energies of O...H interactions estimated using equation (3) and proton NMR chemical shifts, δ_H . $1J = 0.2389$ cal. Correlation coefficient $r = 0.983$.

2. CANONICAL STRUCTURE WEIGHTS DERIVED FROM THE MOLECULAR GEOMETRY

Energetics is one of the fundamental ways which helps to understand chemistry. Another way, very often used in organic chemistry, is the so-called resonance theory [40]. Its originates from *Valence Bond* (VB) theory, which in turn may be a very useful tool for qualitative and quantitative understanding of chemistry, particularly organic reactivity [41] or molecular energetics [42]. It considers description of a molecule (or its fragment) in terms of canonical (resonance) structures, and chemical or physicochemical properties are interpreted in terms of weights of a particular

structure(s) determining a given property in question. The problem however arises, that while drawing these canonical structures we may follow some rules (cf. e.g. [43]), but the weights with which these structure come into description of the molecule (or its fragment) in question are unknown. Very often they are rather tentatively established *ex post*, just for rationalization of some chemical facts.

It is possible to apply theoretical models working exactly in the frame of the *Valence Bond* theory. Another way is to employ empirical models for the above mentioned purpose. The aim of this review is to show how molecular geometry may be employed to solve these kinds of problems. As a useful method, an empirical model called Harmonic Oscillator Stabilization Energy (HOSE) can be used. Its the chief aim is to serve as a tool for determining the weights of canonical structures in π -electron systems [44,45], which may be either molecules or their fragments.

2.1. Principles of the HOSE model [44,45]

The HOSE model has already been reviewed [46,47], but on the basis of its original presentation, without practical hints for further applications. We make an attempt to present the HOSE model in a way which approaches closer the way of thinking used in everyday practice of organic chemistry. Moreover, we wish to present it in such a way that will give a better insight into how the model may be applied to some new problems. Within the frame work of the HOSE model, the following assumptions are made:

1. Geometry of the molecule or its fragment is realized as a result of optimization of all intramolecular interactions (forces) present in it as well as all intermolecular interactions in which this molecule or its fragment is involved. In most cases the intermolecular interactions may be neglected.
2. Geometry of a given, *i*-th, canonical structure assumes that some bonds are purely single, some others are purely double. Hence some values of single X-Y bonds and double X=Y bonds are used in the model as references.
3. The real geometry of the molecule or its fragment is a weighting sum of a few (in principle many) canonical structures. Thus their proper blend allows us to obtain the real geometry.
4. Deformations of bond lengths in the real molecular geometry from some reference lengths (cf. point 2 above) may be approximately described in terms of the harmonic potential.
5. CC bonds from the acyclic polyenes and XY bonds from their heteroanalogues are used as the reference single and double bond lengths are taken
6. Other (e.g. angular) deformations are of lesser importance and are neglected in this model.

The formula for calculating energy by which the real molecule (or its fragment) is more stable than the *i*-th canonical structure (or in other words, by which the *i*-th canonical structure is less stable than the real molecule) is given by eq. (7) below:

$$\text{HOSE}_i = 301.15 \cdot \left[\sum_{r=1}^{n_1} (R_r' - R_o^s)^2 \cdot k_r' + \sum_{r=1}^{n_2} (R_r'' - R_o^d)^2 \cdot k_r'' \right] \quad (7)$$

where R_r' and R_r'' stand for the lengths of π bonds in the real molecule, whereas n_1 and n_2 are the numbers of the corresponding formal single and double bonds in the i -th canonical structure, respectively. In the process of deformation the n_1 bonds corresponding to the single bonds in the i -th canonical structure are lengthened, whereas the n_2 bonds corresponding to the double bonds in the i -th canonical structure are shortened to the bond lengths R_s^o and R_d^o , respectively. 301.15 is the normalization constant for the units used in calculations: bonds lengths in angstrom units, force constants k in $(\text{dynes/cm})10^5$ then the energy is in kJ/mol The force constants k_r in equation 5 are estimated empirically by using the formulas (eq. 8):

$$k_r = a + bR_r \quad (8)$$

Applying k_r and R_r for purely single and double bonds, constants a and b are estimated. This formula works well for the bonds between atoms of elements in the second row of the Periodic Table. For longer bonds formula (9) is used [48].

$$\log k_r = 2.15 - 6.60R_r \quad (9)$$

Table III comprises parameters used in the HOSE model.

Table III. Reference bond lengths and empirical parameters for the HOSE model [44-45]

Type of bond	R_s^o	R_d^o	a	b
CC	1.467	1.349	44.390	-26.020
CN	1.474	1.247	43.180	-25.730
CO	1.428	1.209	52.350	-32.880
NN	1.420	1.254	78.920	-52.410
NO	1.415	1.164	33.187	-19.924

Applying the precise geometry (bond lengths) of a molecule or its fragment and data of Table III and equations (7-9), one can calculate the value of HOSE_i , i.e. the approximate energy by which the i -th canonical structure (with bond lengths R_s^o and R_d^o) is less stable than the real molecule with its own geometry. Obviously, each canonical structure may have quite different HOSE value. The total number of HOSE values which may be calculated for a given molecule or its fragment depends on the number and kind of the canonical structures which we have chosen for further studies.

Following chemical intuition and basic ideas of VB theory, two additional assumptions are made:

- 7) All the most important canonical structures have to be taken into consideration in calculating total HOSE, for a given molecule, and the formula for the overall HOSE is a weighting sum of all of them:

$$\text{HOSE} = \sum_{i=1}^N C_i \cdot \text{HOSE}_i \quad (10)$$

where summation runs over all canonical structures.

- 8) The weight of the *i*-th canonical structure, C_i , in the description of geometry of the real molecule is **inversely proportional** to its destabilization energy (i.e. HOSE_i) the energy by which the *i*-th canonical structure is less stable than the real molecule:

$$C_i = \frac{[\text{HOSE}_i]^{-1}}{\sum_{j=1}^N [\text{HOSE}_j]^{-1}} \quad (11)$$

where N is the number of resonance structures taken into consideration.

Some comments are needed to explain the choice of assumptions made in the HOSE model. It is also useful to take into account experience gathered after about of 15 years of its application.

(i) First of all it should be said that the application of the HOSE model is most efficient if one takes into account a reasonable (i.e. not too large) number of canonical structures. If too many canonical structures are used, the resulting data are “flattened”, which means that the differentiation of the weights may be too small to be decisive.

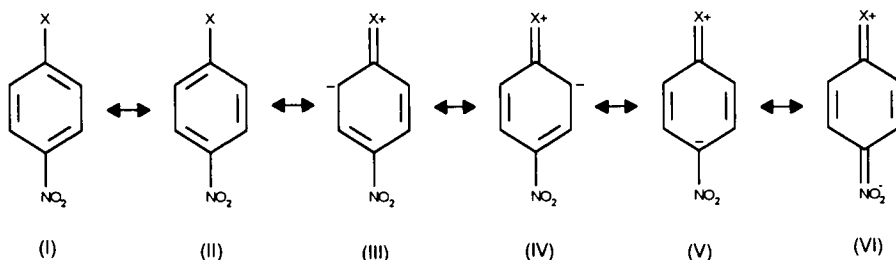
(ii) If the problem in question involves the molecules (or their fragments) which differ widely in bonds constituting them, it is better to choose a fragment which is fixed and possibly the same for all systems taken into consideration. A very useful illustration of these two points is presented in Fig. 11 where the canonical structures taken into calculation were built up only from the phenyl ring and the CN bond of the nitro group. Neither the NO bonds, nor the CX bond were taken into account. The former for the reason mentioned in (i), the latter because X changes over a large variety of heteroatoms.

(iii) Only precise geometric data should be applied. Bond lengths in formula estimating HOSE (7) are taken in quadratic terms, and taking into account the estimation force constant by eq. (8) results a the cubic term ! Thus any error put into the calculation of the HOSE value may be enlarged. Fortunately, in estimation of weights this effect may be markedly diminished.

2.2. Substituent effect illustrated by use of the HOSE model

Most of the chemical and physicochemical properties dependent on the substituent effect have been successfully interpreted in terms of the Hammett equation

or its extensions [49-54]. Bond lengths had not been used successfully in this way until the HOSE model was applied, with the help of which bond lengths were translated into the language of resonance theory and the weights of the canonical structures were computed. However this kind of treatment needs precise molecular geometry of about ten Y-Ph-X systems with a wide spectrum of substituent effects of the X-group. For the first time it was found for p-derivatives of nitrobenzene [55] for which the following canonical structures were taken into account :



Scheme 1. Considered canonical structures of p-nitrobenzene derivatives.

In these structures only the ring CC bond lengths and the CN bond length were taken into calculation of the canonical structure weights. The resulting HOSE weights for canonical structure follow the dependence on substituent constant σ_p or σ^+ . The σ^+ constants were used since the nitro group is the electronaccepting one, and these constants account for the contribution of through resonance effect (canonical structure VI) when the counter substituent is strongly electron donating [56].

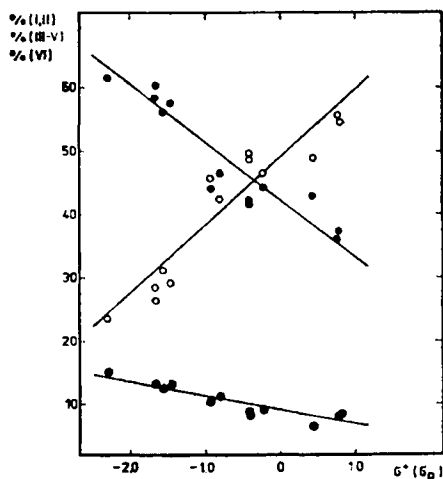


Fig. 11. The dependence of % (I,II) (empty circles), % (III, IV) (full circles) and % (VI) (crosses) calculated for the traditional resonance theory scheme of substituent effect on σ^+ for electron-donating and σ_p for electron-accepting substituents.

Figure 11 is a good illustration of the improvement of the final result owing to the fact that smaller number of the canonical structures were taken into account (comment i) instead of many additional possibilities. In this case a decrease of this number in the interpretation was achieved by blocking (summing up) the weights of canonical structures of similar meaning. If in the cases presented in Fig. 11 all possible canonical structures were taken into account, including mesomeric forms of the nitro group and also of the counter substituent (where it was possible), then the final picture might be completely ineligible. This would be also due to the by passing of the comment (ii).

Further studies of the application of the HOSE model to description of the substituent effect on benzene ring geometry revealed [57] that these kinds of correlations as presented in Figs. 11 and 12 appear only in a case when the fixed group Y in X-Ph-Y system is either strongly electron-donating (e.g. p-substituted anilines [57]), Fig. 12, or strongly electron-attracting as in the case of the nitrobenzene derivatives, as shown above. The dependence becomes worse in this part of the scatterplot where the mesomeric effects between the substituents are either weak or nonexistent (substituents with $\sigma < 0.1$.)

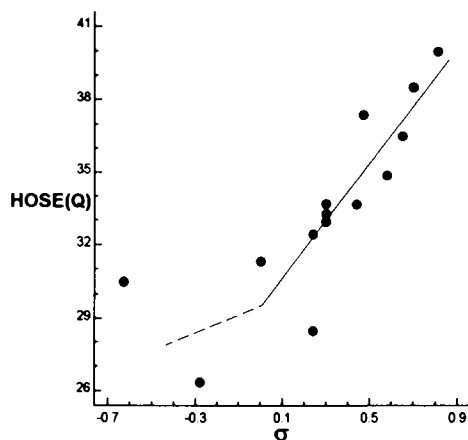


Fig. 12. Dependence of the weight of the quinoid canonical structure for para- substituted aniline derivatives on Hammett substituent constant. Correlation coefficient (in the case of electron-accepting substituents) equal to 0.861 (13 points).

2.3. Structural evidence against the classical through resonance concept in p-nitroaniline and its derivatives

For a long time it has been generally accepted that in p-nitroaniline (and related compounds) a strong through resonance effect operates between the nitro and the amino group or other electron donating groups. Moreover, it has been usually silently assumed that both groups operate in it with (almost) the same strength (but in opposite directions). As a result of this assumption, this effect is said to be best described by assuming a large weight of the canonical structure VI of scheme 1. Thorough X-ray

diffraction studies on *N,N*-dimethyl-4-nitro-3,5-xylidine [58] (I), *N,N*-dimethyl-4-nitro-2,6-xylidine [59] (II) and for comparison *N,N*-diethyl-4-nitroaniline [60] (III), depicted in Fig. 13, evidenced that the above presented view, almost fully accepted in organic chemistry, is not quite correct. Scheme 2 presents canonical structures from the original paper [58] which were used there for comparison with the Hiberty *et al.* results obtained by use of VB theory for *p*-nitroaniline [61].

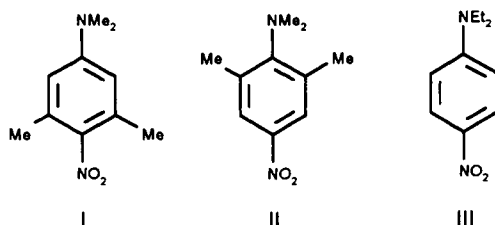


Fig. 13. Considered three *N,N*-dialkyl-*p*-nitrobenzene derivatives - see text.

Table IV shows how the weighting of the canonical structures, schematically presented in Scheme 2, is associated with ortho substitution by two methyl groups in relation to nitro- and *N,N*-alkyl-amino groups.

Scheme 2. Considered canonical structures

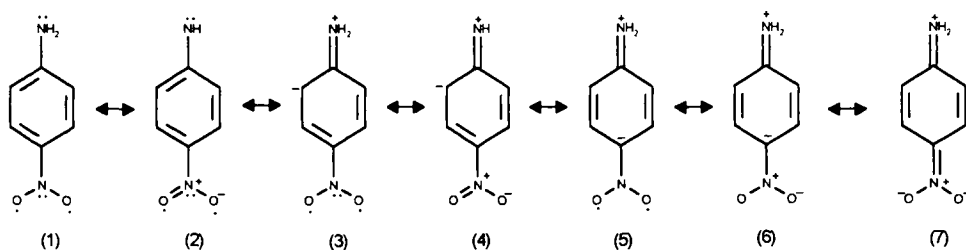


Table IV. Weighting (in %) of the canonical structures (1) - (7) calculated for structures I-III (see Fig. 13)

Canonical structures	I	II	III ^{*)}
(1) and (2)	31.5	41.1	25.2
			28.0
(3) and (4)	36.4	33.1	36.7
(symmetrized)			36.7
(5) and (6)	20.8	16.5	25.4
			22.7
(7)	11.4	9.3	12.7
			12.6

^{*)} Two independent molecules in the unit cell.

A consideration of the monoionic quinoid structure distribution for (3), (4) and (5), (6) reveals that the sum of their weightings is 57.2 and 60.7% for I and III, and only 49.6% for II. This means that noncoplanarity of the NO₂ group with the ring, caused by 3,5-dimethyl groups, does not disturb π -electron delocalization in such a way as to eliminate the similarity between molecule I with the nitro group twisted by ca 60° and molecule III in which the nitro group and the ring are coplanar. In contrast, the methyl groups in positions 2,6 in molecule II and in consequence noncoplanarity of the Nalk₂ group (by ca 60°) with the ring causes a dramatic effect in π -electron delocalization; a significant decrease of the weighting of monoionic structures (3-6) and also a significant increase of the weighting of nonionic structures (1 and 2). This means that the amino group is the main factor inducing the through resonance effect, since if it is twisted out of the coplanarity with the ring, the weights of the nonionic structures (in the ring) increase very strongly. The above conclusions are in line with the VB calculations made for p-nitroaniline and p-nitrophenol [61].

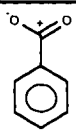
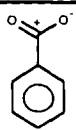
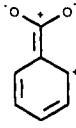
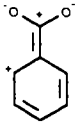
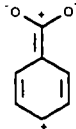
2.4. Does the nitro group interact mesomerically with the ring in nitrobenzene?

Nitrobenzene is a paradigm molecule in organic chemistry and its canonical structures are often presented in handbooks on organic chemistry [62] or in reviews [63]. They say that the nitro group interacts with the phenyl ring by resonance (i.e. the canonical structures with positive charges in the ring (depicted in Table V) should dominate) [64]. The problem arises as concerns the weights of particular canonical structures. In many cases the resonance structures in which the positive charges reside in ortho and para positions in the ring were used to interpret electrophilic substitution in meta positions [62]. Application of the HOSE model to precisely determined molecular geometry of nitrobenzene [64] revealed the weighting scheme, as shown in Table V. Apart from the experimental geometry the bond lengths from *ab initio* calculations (6-31G and 6-31G*) were used leading to equivalent results. It is apparent that the quinoid canonical structure is definitely less important than the two Kekulé structures.

The above results are in line with the picture obtained from the experimental charge density studies on nitrobenzene [65] made by use of low temperature X-ray diffraction studies [64]. When the perpendicular sections are carried out through the CN and CC bonds in the ring in their centers, the picture obtained is as in Fig. 15. To make the view more clear the additional operation is made: the difference between this map and the same map rotated by 90° is shown, indicating the π -bond ellipticity.

The following conclusion may be drawn from the above results: **the CN bond in nitrobenzene is almost cylindrical indicating a very low contribution of the π - electron component.** In contrast to them, the typical aromatic CC bonds in the ring are significantly elliptical - as expected from the chemical intuition and experience. Thus the results of experimental charge density studies are in line with the much simpler treatment based on the HOSE model and precisely measured bond lengths.

Table V. Weighting (in %) of the canonical structures of nitrobenzene from geometry estimated within various methods of *ab initio* calculations and methods of X-ray structure determination.

		Method of geometry determination				
		X-ray	Corrected for librations	6-31G	6-31G*	
	and		73.43	73.50	72.85	73.15
	and		17.73	17.61	18.05	17.87
			8.84	8.88	9.1	8.98

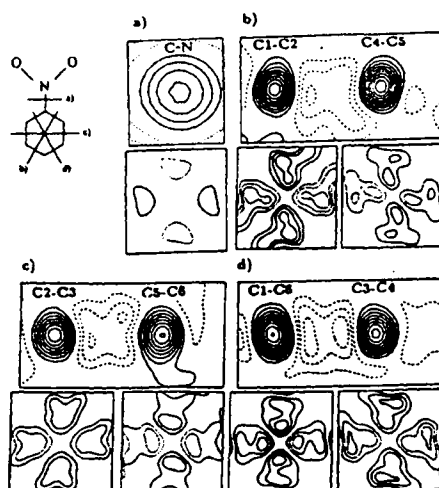


Fig. 15. Difference density (DD) maps of nitrobenzene [65] with respect to a procrystal of spherically averaged atoms, on cuts vertical to the molecular plane through the midpoints of bonds, as shown in the left top diagram. Under each DD map, the difference between this map and the same map rotated by 90° is shown ($\Delta\rho$), indicating the π -bond ellipticity. Contour line values are $\pm 0.05n \text{ e}/\text{\AA}^3$, $n=1,2,3,\dots$

From the above considerations a substantial conclusion may be drawn: the resonance effect of the nitro group in nitrobenzene is negligible in the ground state

of the nitrobenzene and the widely used canonical structures implying the resonance interactions between the NO_2 group and the ring are correct only for the reactivity problems when nitrobenzene is in strongly Lewis acidic/basic media.

2.5. Angular group induced bond alternation - a new substituent effect detected by molecular geometry

Recently it has been found that the angular substituent, the methoxy group, induces an imbalance of the Kekulé structures, based on the calculated (6-31G*) geometry for anisole [66]. The observed imbalance is presented in Fig. 16.

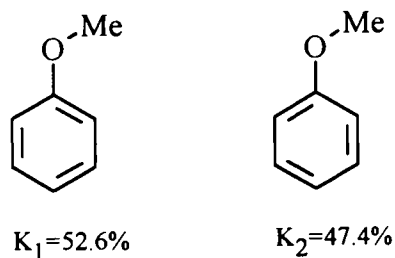


Fig. 16. Kekulé structures and their relative weights for anisole.

It indicates that the methoxy group induces a shorter CC bond *cis* to it and a longer one *trans* to it. Moreover, when the OMe group is made closer to the ring by bending the C-O-Me bond angle by 10° , the imbalance increases to the value 60.8:39.2. Deeper experimental studies of this effect evidenced its presence in many systems with angular substituents, e.g. in 1,3,5-trimethoxybenzene [67], 2,4,6-trimethoxy-*s*-triazine [68] and N-substituted derivatives of phenyldiazene [69]. The imbalance of Kekulé structures in the case of 2,4,6 trimethoxy-*s*-triazine [68] is as large as 57.8:42.2, for both, the experimental low temperature X-ray measured geometry and *ab initio* 6-311G (d,p) computed geometry (Fig. 17).

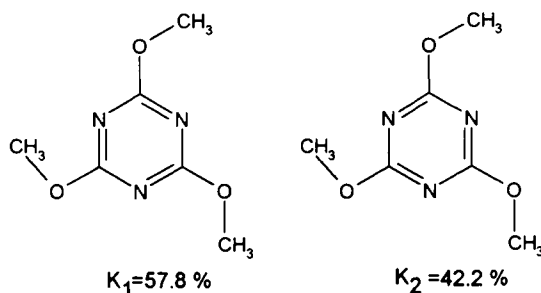


Fig. 17. Resonance structures of 2,4,6-trimethoxy-*s*-triazine and their weights computed from *ab initio* 6-311G (d,p) geometry by use of the HOSE model.

On the other hand, analysis of experimental geometry of 21 diazobenzene derivatives [69] has shown that the $N=NR$ group induces a longer CC bond in the ring *cis* to $N=N$ group, and that the difference between the mean bond lengths $C1C2$ and $C1C6$ (where $C1$ is the substituted carbon atom in all cases considered) is significant at a very high significance level. The calculated geometry of 1,3,5-triazabenzene with C_c symmetry by *ab initio* 6-31G* led to the imbalance of Kekulé structure 66.5:33.5, but in the opposite direction of the effect than that observed for 1,3,5-trimethoxybenzene [67]. These results are summarized in Fig. 18.

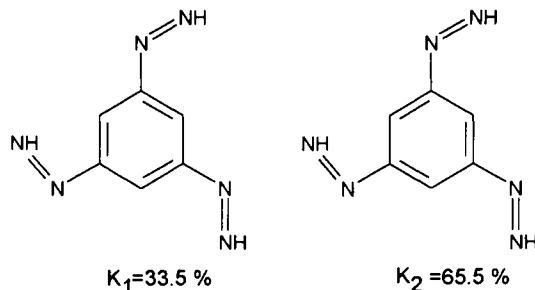


Fig. 18. Resonance structures of 1,3,5-triazabenzene and their weights computed from *ab initio* 6-31G* geometry by use of the HOSE model. Note the reverse weighting scheme as compared with that in Fig. 17.

Analysis of the above-mentioned experimental data enriched by *ab initio* calculations revealed that the following phenomenological rule works [67]: **the angular XY substituent with a single $X-Y$ bond induces more double CC bond in the ring *cis* to the XY bond and more single bond *trans* to the XY group. If the XY group has a double $X=Y$ bond then the effect is reverse.**

The above rule predicts the substantial difference in the Kekulé structure imbalance for two conformers of 1,4-dimethoxybenzene. Figure 19 shows this difference

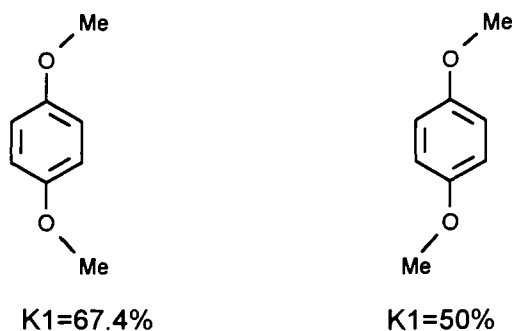


Fig. 19. Kekulé structure weights for *cis* and *trans* conformers of 1,4-dimethoxybenzene.

From the bending XYC bond angle as observed in the first paper [66] a conclusion may result that the effect is a through space effect of the substituent. In order to study this problem more thoroughly the following model calculations were carried out [69]. The π -electronic interactions between the $N=N$ groups and the ring in 1,3,5-triazobenzene with C_3 symmetry can be eliminated by rotating the $N=N$ groups by 90° around their $C-N$ bonds, leaving $C2C1N$ and $C1C2H$ angles unchanged (i.e. as in the planar, optimized conformation) and optimizing all other geometrical parameters. As a result, the ring still undergoes an important bond alternation ($C1-C2 = 1.3753\text{\AA}$, $C1-C6 = 1.3916\text{\AA}$), but in a way opposite to that observed in the planar triazolo derivative (where the observed bond lengths were $C1C2 = 1.3972\text{\AA}$ and $C1C6 = 1.3756\text{\AA}$). Figure 20 presents these results in terms of Kekulé structure imbalance. They may be compared with the weights for a planar optimized geometry (Fig. 18).

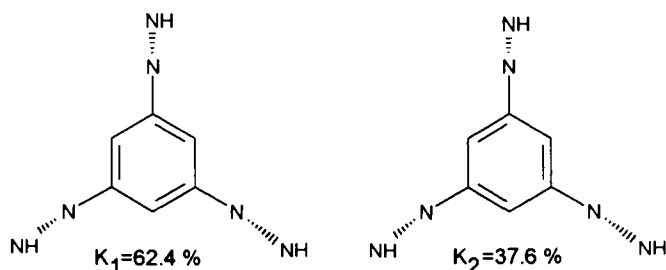


Fig. 20. Weighting scheme for 1,3,5-triazobenzene with $N=NH$ groups rotated perpendicularly to the ring but with a maintained $NC1C2$ bond angle as in the planar conformation. Note the reverse weighting scheme as compared with that in Fig. 18.

It is worth mentioning that in this case the changes are in the same direction as in 1,3,5-trimethoxybenzene of C_3 symmetry with a $C1C6$ bond longer than $C1C2$. A similar bond alternation has been observed by Stanger [70] in a constrained bent benzene with CCH bond angle equal to 90° . All these results point to the same conclusion: **the bond alternation induced by angular strain is opposite to that induced by direct π electronic interaction in triazobenzene, the latter being the strongest and imposing the direction of the overall effect of the bond fixation.**

It is also worth mentioning that the angular substituents by changing the ring geometry affect its aromaticity. Moreover, since aromaticity depends so strongly on the conformation of the molecule it may be concluded that the aromaticity may depend significantly on the conformation of angular groups. Figure 21 presents the aromaticity index HOMA [25] of the phenyl ring for two conformers of *p*-dinitroso and *p*-dimethoxy benzene [71]. The observed effect of a decrease in the aromatic character of the ring is small but the main reason of a decrease in aromaticity is in both cases geometric in nature [26,71,72].

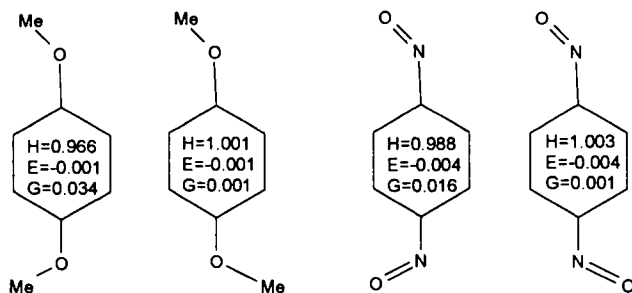


Fig. 21. Aromaticity index HOMA and its components EN and GEO of the phenyl ring for two conformers of p-dinitroso and p-dimethoxy benzene.

3. SUBSTITUENT EFFECT ON THE MOLECULAR GEOMETRY

As it is well known for almost a hundred years and what has become a subject of countless reports, substituents may affect significantly chemical reactivity and physicochemical properties of the molecules in question [73,74]. Their effect on molecular geometry became better known since X-ray diffraction techniques of structure determinations began to be a readily accessible source of molecular geometry and particularly the data bases have become well supplied with precise structural data.

The first effective and successful approach to the above problem was done by Domenicano, Vaciago and Coulson as early as in 1975 [75,76]. They found that the substituents in a monosubstituted benzene derivatives distort significantly the *ipso* bond angle in the ring, and the distortion depends well on the group electronegativity [77,78] of the substituent [79]. The idea can be visualized by using the Bent -Walsh rule [80,81] as shown in Fig. 22 [46]. The rule reads [80,81]: "if a group X1 attached to carbon is replaced by a more electronegative group X2, then the carbon valence toward X2 has more p character that it had toward X1". This, of course, implies a decrease of the p character in the other two hybrid orbitals constituting a bonds

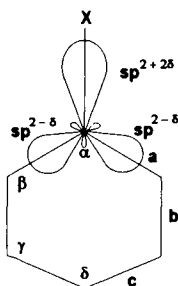


Fig. 22. A scheme presenting the action of the Walsh-Bent rule on sp^2 -hybridized orbitals at the substituted carbon atom in a benzene ring; X is a strongly electronegative substituent.

and leads to [79] an increase in the α value and a shortening of both bonds of the adjacent CC bonds a.

An obvious consequence of the Bent-Walsh rule is that bond lengths a and bond angle α should be correlated. Unfortunately this has never been found due to too low precision in bond lengths determination [82]. However, if instead of bonds lengths a the differences $\Delta = b - a$ are taken, then the mutual correlation appears [83] as shown in Fig. 23.

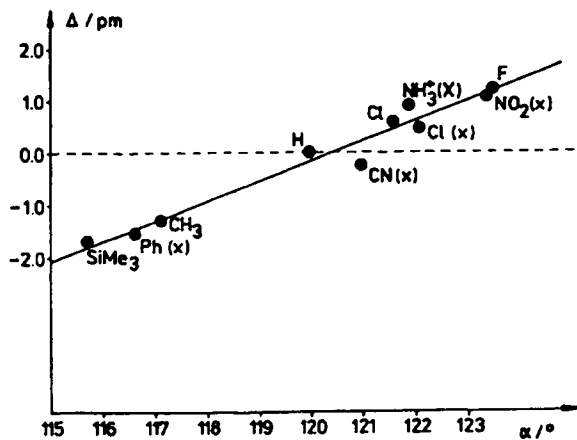


Fig. 23. A plot of Δ versus α for 10 symmetrically p-disubstituted derivatives of benzene. Data labeled by X stand for X-ray structure determination [83].

The dependence Δ vs α works since in the subtraction part of the experimental error cancel out, and hence the dependence predicted by the Bent-Walsh rule may be observed. It is worth mentioning that the line described by Δ vs α relationship for p-homo-disubstituted benzene derivatives may be a good reference line for the situations in which there is no π -electron cooperative effect. For p-X-Y-disubstituted derivatives, if X and Y differ significantly by their electron accepting/donating abilities, the appropriate points in this graph are shifted down, as shown in Fig. 24.

In the case of N,N-diethyl-p-nitroaniline as it is well visualized in the graph, the points (two points for each of these groups, since there are two independent molecules in the unit cell) which can be defined for the nitro group and N,N-diethylamino group are shifted in a quite different way. Two points for the NO₂ group are shifted down by less than 0.02 Å, whereas two points for the N,N-diethyl group are shifted by more than 0.04 Å. This effect may be rationalized by assuming that the substituent effect being a combination of the mesomeric and inductive ones, operates mesomerically stronger for the amino group than for the nitro group. The conclusion is in line with other results obtained by independent reasoning in the previous paragraphs.

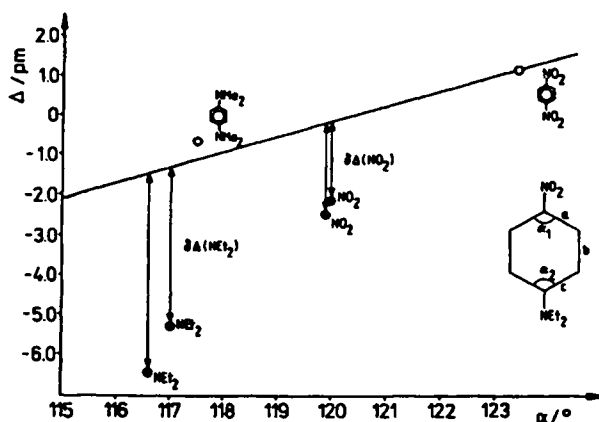


Fig. 24. A plot of Δ versus α for symmetrically p-disubstituted benzene derivatives (the line is as in Fig. 23) with two examples as open points: p-dinitrobenzene and p-N,N,N',N'-tetramethylphenylenediamine. Solid points are: Δ and α values for nitro group ($\Delta = b-a$) and NEt_2 ($\Delta = b-c$) in N,N-diethyl-p-nitroaniline (two independent molecules in an asymmetric unit of the crystal cell). The shift down from the line $\delta\Delta$ describes (quantitatively) a π -electron cooperative effect between the NO_2 and NEt_2 groups.

It is worth mentioning that a similar picture is obtained in the case of EDA-complexes of N,N,N',N'-tetramethylphenylenediamine, as shown in Fig. 25.

Again, the greater is the charge transfer, this time an intermolecular charge transfer from the phenylenediamine derivative onto the accepting molecule in the EDA complex or anion in the salt, the greater is the observed shift of Δ from the reference line obtained for the Δ vs α plot for para-homodisubstituted benzene derivatives.

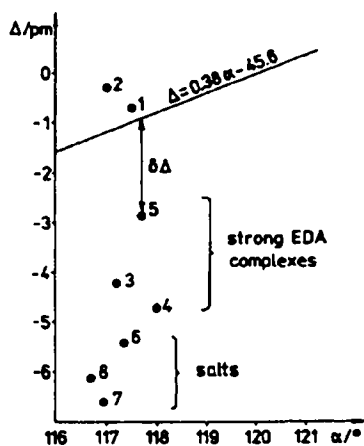


Fig. 25. A plot of Δ versus α for neutral N,N,N',N'-tetramethylphenylenediamine (1), its weak electron donor-acceptor complex (2), strong complexes (3-5) and salts (6-8).

Undoubtedly, the model presentation of fulfilling the Bent - Walsh rule may be a convenient way of a quantitative estimation of through resonance effect compared with the reference interactions which are inductive (or due to electronegativity) in nature.

4. AROMATIC CHARACTER DERIVED FROM MOLECULAR GEOMETRY

Molecular geometry is one of the most important sources of information on aromaticity of π - electron systems. Bond lengths have long been applied for defining indices of aromaticity. Firstly Julg *et al.* [84] applied a function of variance of bond lengths in their definition of aromaticity index A_j . Then this idea was changed, namely the mean bond length in their formula was replaced by an optimal bond length, leading to the formula (12) [24,25]. The physical meaning of the optimal bond length is that the energy of extension of a typical double bond to the optimal bond length is equal to the energy of compression of a typical single bond to the optimal bond. Application of the harmonic potential to extension and compression and appropriate reference bond lengths for typical double and single bonds including heteroatoms, the model, called HOMA (from harmonic oscillator model of aromaticity) could be extended for hetero- π -electron systems [25]. The general formula for HOMA for the systems built up of CC, CX, CY, and XY bonds reads:

$$\text{HOMA} = 1 - \left\{ \alpha(\text{CC}) \sum \left[R(\text{CC})_{\text{opt}} - R_i \right]^2 + \alpha(\text{CX}) \sum \left[R(\text{CX})_{\text{opt}} - R_i \right]^2 + \alpha(\text{CY}) \sum \left[R(\text{CY})_{\text{opt}} - R_i \right]^2 + \alpha(\text{XY}) \sum \left[R(\text{XY})_{\text{opt}} - R_i \right]^2 \right\} / n \quad (12)$$

where n is the number of bonds taken into summation, $\alpha(\text{XY})$ is an empirical constant which accounts for the ability of the specific bond $R(\text{XY})$ to undergo compression or expansion, as well as taking into account different ranges of bond length variability depending on the nature of the bond, fixed so that $\text{HOMA} = 0$ for the Kekulé structure of the typical aromatic system, and is equal to 1 for the system with all bonds equal to the optimal value $R(\text{XY})_{\text{opt}}$. R_i - stand for bond lengths of the considered molecular system. Table VI gives the parameters which are necessary for eq. (12).

Table VI. Structural parameters of the HOMA index

	R(s)	R(d)	R_{opt}	α
CC	1.467	1.349	1.388	257.7
CN	1.465	1.269	1.334	93.52
CO	1.367	1.217	1.265	157.38
CP	1.814	1.640	1.698	118.91
CS	1.807	1.611	1.677	94.09
NN	1.420	1.254	1.309	130.33
NO	1.415	1.164	1.248	57.21

A large field of applications of the HOMA index has recently been presented in two reviews [25,72], so there is no need to repeat them. The most important achievement is that the HOMA index can be analytically separated into two terms [26, 85] which account for the energetic and geometric contributions, EN and GEO, respectively. Originally this was done for carbocyclic systems [26], then the formula was modified to account for heterocyclic systems as well [85]. This more general formula is presented by eq. (13):

$$\text{HOMA} = 1 - \left[257.7 (1.388 - r_{\text{av}})^2 + \frac{257.7}{N} \sum (r_{\text{av}} - r_i)^2 \right] = 1 - \text{EN} - \text{GEO} \quad (13)$$

The values of r_{av} and r_i in eq. 11 are obtained for XY bond types using the Pauling concept of bond number [15]; all bonds with heteroatoms are in this way recalculated into virtual CC bonds. For CC bonds themselves the values of r_i are equal to the experimental bond lengths; r_{av} stands for an averaged value.

The two terms are, in general, independent [26,85] of each other and each of them determines the aromatic character in a different way. As far as the individual rings in benzenoid hydrocarbons are concerned in the central ring of perylene or triphenylene the deciding term is an energetic one, whereas for the central rings in phenanthrene or pyrene the geometric term is most important. An increase of the energetic term, EN, is equivalent to a decrease of resonance energy of the molecule or its fragment in question, whereas an increase of the geometric one, GEO, means that the system undergoes an increase of bond alternation. It is worth mentioning that both components of the aromatic character may be represented by other indices. The aromaticity indices I_5 and I_6 introduced by Bird [86] for 5- and 6-membered rings correlate well with the GEO terms [14,26,72]. This result is logical since the Bird indices depend linearly on the variance calculated for Gordy's [87] bond orders for the bonds in question. Another index, BAC [14,26,34-35,72], describes alternation too, but using quite a different model: it is a normalized sum of the differences between the consecutive bonds in the ring. It does not work for polycyclic systems. Nevertheless, when BAC is related to GEO term the correlation is excellent [26,72]. Finally, it should be mentioned here that the EN term is perfectly related to the REC values (equation 3) for phenylic rings in: benzenoid hydrocarbons, TCNQ moieties in the EDA complexes and salts, benzene moieties in p-disubstituted benzene derivatives and in cyclopentadienyl rings in complexes with Rh.[26].

Katritzky *et al.* [88] and Jug [89] and others [14,22,57] have shown recently that aromaticity is a multidimensional phenomenon. For this purpose they applied statistical methods (the principal component or factor analyses [51,90,91]) which need application of many indices of aromaticity estimated (experimentally or calculated theoretically) for many molecular systems. Thus their conclusion is of great general importance but no information may be extracted for any individual molecule or its fragment. The method presented above, the separation of HOMA value into EN and GEO terms, allows us to describe numerically which of these two factors is

responsible for the dearomatization of the molecule or its fragment in question. This treatment was used for benzenoid hydrocarbons [26, 22], polysubstituted benzene derivatives [92], heterocyclic systems [85, 93] and nonalternant systems including annulenes [26]. All these aspects are reviewed in [72].

An interesting aspect of aromaticity is its relation to the planarity of the π -electron system in question. As it was shown using the geometry of *m*-cyclophane [94] and of per-substituted naphthalenes, even a substantial folding of the ring in the former case or of the bicyclic system in the latter, does not lead to the loss of aromatic character. It happens rather suddenly when a considerable deviation from coplanarity

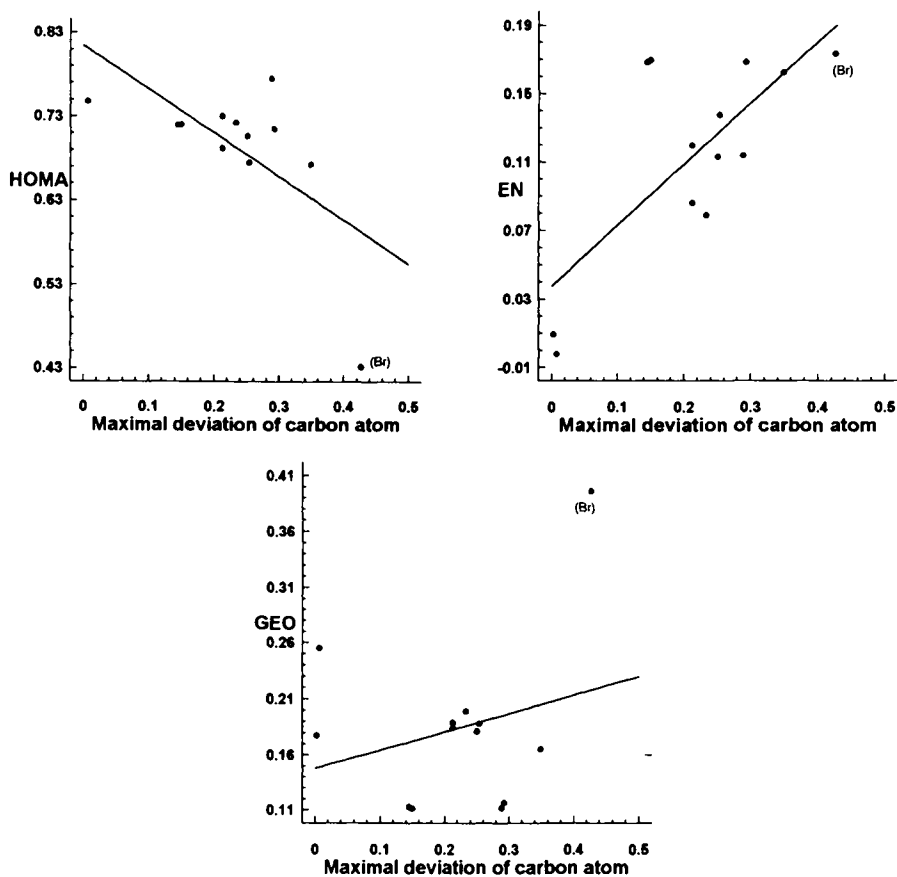


Fig. 26. Scatter plots of HOMA, EN and GEO values for 12 per-substituted naphthalene derivatives plotted against maximal deviation of the carbon atom from the best plane of all carbons in naphthalene. The correlation coefficients are -0.62 for HOMA, 0.71 for EN and -0.37 for GEO. The first two values are statistically significant at $\alpha = 0.05$. The point for Br was not taken into correlation because its low precision but it shows how far this strongly planarity-deforming substitution changes the geometric contribution to aromaticity.

occurs [25]. Now we will show what role these two terms, EN and GEO, play in dearomatization of the π -electron system in per-substituted naphthalenes. Figure 26 presents scatter graphs for the dependence of HOMA, EN and GEO terms on maximal deviation of the carbon atom of the naphthalene moiety from the best plane calculated by use of the least-squares method for its all carbons. The scatter plot is based on molecular geometries of 12 per-substituted derivatives of naphthalenes retrieved from the CSD base [2] (November 1996 release).

As it is clearly seen, the loss of aromaticity of the naphthalene moiety due to its deformation from planarity is roughly linear for the energetic term, and completely irregular for the geometric term. Moreover, the large value for perbromo-derivative (GEO = 0.367) appears as a deviation from the rather less differentiated values for other derivatives which are in the range between 0.105 and 0.263.

5. CONCLUSIONS

Molecular geometries are readily accessible from data bases (Cambridge Structural Database [2] and Inorganic Structural Database [3]) and routine measurements by X-ray diffraction techniques. Hence the models in which these data may be employed to give more descriptive chemical (biochemical, physical) information are of great value.

In spite of the enormous development of computer facilities and methods of quantum chemistry, empirical models may often be applied in the fields where the more precise methods cannot be used directly. The empirical treatments may often serve as initiations for more precise theoretical studies. Empirical rules of periodicity given by D. I. Mendeleev had been 50 years ahead of their theoretical explanation offered by quantum mechanics.

More detailed conclusions are as follows:

- (i) Ring energy contents (REC) of the phenylic rings may be calculated from CC bond lengths[14]. In general, for any molecular fragment built up of CC bonds its energy content may be calculated.
- (ii) These values are strongly differentiated and the topological environment of the ring in question plays a much greater role than the chemical changes in the environment [14, 22, 92].
- (iii) The nitro group in nitrobenzene does not interact (or interacts only very weakly) with the ring mesomerically [64]. The same tendency is observed for p-nitroaniline where the mesomeric effect of the nitro group is much smaller than that of the amino group. Nonlinear electric substituent effect [95] of the nitro group is responsible for its strange behavior [65].
- (iv) Detection of the mesomeric effect or generally of the intra- or inter-molecular charge transfer may be observed as a deviation from the model line based on the Bent-Walsh rule applied to systems without mesomeric effects.

(v) Aromaticity may be quantitatively described by models based solely on the bond lengths [25, 86]. Even the energetic and geometric components of the aromatic character may be estimated in this way [85, 93].

ACKNOWLEDGMENT

The BST/24/96 grant provided financial support for this study.

REFERENCES

1. R. Hoffmann, Foreword to the monograph by L.V. Vilkov, V.S. Mastryukov, N.I. Sadova, "Determination of the Geometrical Structure of Free Molecules, Mir Publishers, Moscow, 1993
2. F.H. Allen, J.E. Davies, J.J. Galloy, O. Johnson, O. Kennard, E.M. Mitchell, McRae, G.F. Mitchell, J.M. Smith, D.G. Watson, *J. Chem. Inf. Comput. Sci.*, 31 (1991) 187.
3. G. Bergerhoff, R. Hundt, R. Sievers, I.D. Brown, *J. Chem. Inf. Comput. Sci.*, 23 (1983) 66; G. Bergerhoff, R. Sievers, *Nachr. Dokum.* 40 (1989) 27.
4. An excellent review of modern techniques of determining the molecular geometry cf.: A. Domenicano, I. Hargittai (eds.), *Accurate Molecular Structures. Their Determination and Importance*, Oxford University Press, Oxford, 1992.
5. W. J. Hehre, L. Radom, P. v. Rague Schleyer and J. A. Pople, *Ab initio Molecular Orbital Theory*, J. Wiley, London, 1986.
6. G. Häfelinger, C. Regelmann, T.M. Krygowski and K. Woźniak *J. Comp. Chem.* 10 (1989) 329.
7. H.-B. Bürgi, *Inorg. Chem.*, 12 (1973) 2321.
8. H.-B. Bürgi, *Angew. Chem., Int. Ed. Engl.* 14 (1975) 460.
9. P. Murray-Rust, H.-B. Bürgi and J.D. Dunitz, *J. Am. Chem. Soc.* 97, (1975) 921
10. H.-B. Bürgi, J.D. Dunitz and E. Shefter, *J. Am. Chem. Soc.*, 95 (1973) 5065.
11. For reviews, cf.: J.D. Dunitz, *X-ray Analysis and the Structure of Organic Molecules*, Cornell Univ. Press, Ithaca, NY 1979.; H.B. Bürgi and J.D. Dunitz, *Acc. Chem. Res.*, 16 (1983) 153.
12. For an excellent collection of recent reviews, see: H.-B. Bürgi and J.D. Dunitz (eds.), *Structure Correlation*, VCH Weinheim, 1994.
13. H.S. Longuet-Higgins and L. Salem, *Proc. Roy. Soc.*, 251A (1959) 172.
14. T.M. Krygowski, A. Ciesielski, C.W. Bird, A. Kotschy, *J. Chem. Inf. Comput. Sci.*, 35 (1995) 203.
15. L. Pauling, *J. Am. Chem. Soc.*, 69 (1947) 542.
16. For a recent review, cf. article by A.S. Cieplak, p. 205 ff., in: *Structure Correlation*, H.-B. Bürgi and J.D. Dunitz (eds.) vol. 1 and 2, VCH Weinheim 1994.

17. H.S. Johnston and Ch. Parr, *J. Am. Chem. Soc.*, 84 (1963) 2544.
18. K. Hedberg, V. Schomacher, *J. Am. Chem. Soc.*, 73 (1951) 1482.
19. L.S. Bartell, E. Roth, C.D. Hollowell, K. Kuchitsu, J.E. Young, *J. Chem. Phys.*, 42 (1965) 2683.
20. P. George, *Chem. Rev.*, 75 (1975) 85.
21. J.B. Pedley, R.D. Naylor, S.P. Kirby, *Thermodynamical Data of Organic Compounds*, Chapman and Hall, London, 1986.
22. T.M. Krygowski, M. Cyrański, A. Ciesielski, B. Świrska, P. Leszczyński, *J. Chem. Inf. Comput. Sci.* 36 (1996) 1135.
23. I. Oonishi, I. Ohshima, S. Fujisawa, J. Aoki, Y. Ohashi, T.M. Krygowski, *J. Mol. Struct.*, 265 (1992) 283.
24. J. Kruszewski and T.M. Krygowski, *Tetrahedron Lett.* (1972) 3839.
25. T. M. Krygowski, *J. Chem. Inf. Comput. Sci.*, 33 (1993) 70.
26. T.M. Krygowski and M. Cyrański, *Tetrahedron*, 52, (1996) 1713.
27. For a review of the problem, cf. I. Gutman and S.J. Cyvin, *Introduction to the Theory of Benzenoid Hydrocarbons*, Springer-Verlag, Berlin, 1989.
28. K. Woźniak, T.M. Krygowski, *J. Mol. Struct.*, 191 (1989) 81.
29. H.J. Talberg, *Acta Chem. Scand.*, A29 (1975) 70.
30. H.J. Talberg, *Acta Chem. Scand.*, A31 (1977) 84.
31. T.M. Krygowski, R. Anulewicz, B. Pniewska, P. Milart, C.W. Bock, M. Sawada, Y. Takai, T. Hanafusa, *J. Mol. Struct.*, 324 (1994) 251.
32. T.M. Krygowski, M.K. Kalinowski, I. Turowska-Tyrk, C. Hiberty, P. Milart, A. Silvestro, R.D. Topsom, S. Dähne, *Struct. Chem.*, 2 (1991) 71.
33. T. Dziembowska, B. Szczodrowska, T.M. Krygowski, S.J. Grabowski, *J. Phys. Org. Chem.*, 7 (1994) 142.
34. T.M. Krygowski, A. Ciesielski and M. Cyrański, *Chem. Papers (Bratislava)* 49 (1995) 128.
35. T. M. Krygowski, A. Ciesielski and M. Cyrański, *J. Mol. Struct.*, 374 (1996) 277.
36. G.C. Pimentel and A.L. McClellan, *The Hydrogen Bond*, W.H. Freeman, San Francisco, CA, 1960.
37. G.A. Jeffrey and W. Sängner, *Hydrogen Bonding in Biological Structures*, Springer, Berlin, 1991.
38. E.R. Lippincott and R. Schröder, *J. Chem. Phys.*, 23 (1955) 1099.
39. J.D. Derissen and P.H. Smit, *Acta Cryst.* 34B (1978) 842.
40. G.W. Wheland, *Resonance in Organic Chemistry*, J. Wiley, New York, 1955
41. A. Pross, *Physical and Theoretical Principles of Organic Reactivity*, J. Wiley, New York, NY, 1995.
42. W.C. Herndon, *Tetrahedron*, 28 (1972) 3675; 29 (1973) 3; *J. Am. Chem. Soc.*, 95 (1973) 2404.
43. J. B. Hendrickson, D.J. Cram and G.S. Hammond, *Organic Chemistry*, 3rd Ed., McGraw-Hill, New York, 1970.
44. T. M. Krygowski, R. Anulewicz and J. Kruszewski, *Acta Cryst.*, B39 (1983) 732.

45. T. M. Krygowski, R. Anulewicz, M. Wisiorowski, *Pol. J. Chem.*, 69 (1995) 1579.
46. T.M. Krygowski, Chapt. 6, in: *Structure and Reactivity*, J.F. Liebman and A. Greenberg (eds.), VCH, Weinheim, 1988, p.231.
47. T.M. Krygowski, *Progr. Phys. Org. Chem.*, 17 (1991) 239.
48. J. Karolak-Wojciechowska, *Acta Cryst.*, B43 (1987) 574.
49. Chapman, N.B. and Shorter J. (eds.), *Advances in Linear Free Energy Relationships*, Plenum Press, New York, NY, 1972.
50. Chapman, N.B. and Shorter J., *Correlation Analysis in Chemistry. Recent Advances*. Plenum Press, New York, NY, 1978.
51. Zalewski, R.I., Krygowski T.M. and Shorter J., *Similarity Models in Organic Chemistry, Biochemistry and Related Fields*, Elsevier, Amsterdam 1991.
52. Shorter, J., *Correlation Analysis in Organic Chemistry: an Introduction to Free Energy Relationships*, Oxford Univ. Press, Oxford, 1973.
53. C.D. Johnson, *The Hammett Equation*, Cambridge University Press, Cambridge, 1973.
54. O. Exner, *Correlation Analysis of Chemical Data*, Plenum Press, New York, NY, and SNTL, Prague, 1988.
55. T.M. Krygowski and I. Turowska- Tyrk, *Collect. Czech. Chem. Commun.*, 55 (1990) 165.
56. H.C. Brown and Y. Okamoto, *J. Am. Chem. Soc.*, 80, (1958) 4979.
57. M. Cyrański and T.M. Krygowski, *Pol. J. Chem.*, 69 (1995) 1088.
58. T. M. Krygowski and J. Maurin, *J. Chem. Soc. Perkin 2* (1989) 695.
59. J. Maurin and T. M. Krygowski, *J. Mol. Struct.*, 158 (1987) 359.
60. J. Maurin and T. M. Krygowski, *J. Mol. Struct.*, 172 (1988) 413.
61. P.C. Hiberty and G. Ohanessian, *J. Am. Chem. Soc.*, 106 (1984) 6963.
62. J.B. Hendrickson, D.J. Cram and G.S. Hammond, *Organic chemistry*, McGraw-Hill, New York, NY, 1970.
63. O. Exner and T.M. Krygowski, *Chem. Soc. Rev.* 1996, 71.
64. R. Boese, D. Blaser, M. Nussbaumer and T.M. Krygowski, *Struct. Chem.* 3 (1992) 363.
65. S. Irle, T.M. Krygowski, J.E. Niu and W.H.E. Schwarz, *J. Org. Chem.*, 60 (1995) 6744.
66. T.M. Krygowski, R. Anulewicz, A. Jarmuła, T. Bąk, D. Rasała and S. Howard, *Tetrahedron*, 50 (1994) 13155.
67. S.T. Howard, T.M. Krygowski and M.L. Główska, *Tetrahedron*, 52 (1996) 11379
68. T.M. Krygowski, S.T. Howard, D. Martynowski and M.L. Główska, *J. Phys. Org. Chem.* in press
69. T.M. Krygowski, R. Anulewicz and Ph. C. Hiberty, *J. Org. Chem.* in press
70. A. Stanger, *J. Am. Chem. Soc.*, 113 (1992) 8277.
71. T.M. Krygowski, M. Cyrański, M. Wisiorowski, *Pol. J. Chem.* 70 (1996) 1351.
72. T.M. Krygowski, M. Cyrański, in *Advances in Molecular Research*, vol. 3, M. and I. Hargittai (eds). JAI Publishers, in press.

73. L.P. Hammett, *Physical Organic Chemistry*, McGraw-Hill, New York, NY, 1970.
74. A. Domenicano in *Accurate Molecular Structures. Their Determination and Importance*, A. Domenicano, I. Hargittai (eds.), Oxford University Press, Oxford, 1992.
75. A. Domenicano, A. Vaciago and C.A. Coulson, *Acta Cryst.* B31 (1975) 221.
76. A. Domenicano, A. Vaciago and C.A. Coulson, *Acta Cryst.*, B31 (1975) 1630.
77. J.E. Huheey *J. Phys. Chem.*, 69 (1965) 3284.
78. J.E. Huheey *J. Phys. Chem.*, 70 (1966) 2086.
79. A. Domenicano and A. Vaciago, *Tetrahedron Lett.* (1976) 1029.
80. A.D. Walsh, *Discussions Faraday Soc.*, 2 (1947) 18.
81. H. A. Bent, *Chem. Rev.* 61 (1961) 275.
82. T. M. Krygowski, I. Turowska-Tyrk, *Pol. J. Chem.*, 64 (1990) 289.
83. T. M. Krygowski, *J. Chem. Res. (s)* 1984, 238.
84. A. Julg and Ph. Francois, *Theor. Chim. Acta*, 7 (1967) 249.
85. T. M. Krygowski, M. Cyrański, *Tetrahedron*, 52 (1996) 10255.
86. C.W. Bird, *Tetrahedron*, 41 (1985) 1409; *Tetrahedron*, 42 (1986) 89; *Tetrahedron*, 43 (1987) 4725; *Tetrahedron*, 46 (1990) 5697; *Tetrahedron*, 48 (1992) 335; *Tetrahedron*, 48 (1992) 1992; *Tetrahedron*, 48 (1992) 7857; *Tetrahedron*, 49 (1993) 8441.
87. W. Gordy, *J. Chem. Phys.*, 15 (1947) 305.
88. A. R. Katritzky, P. Barczyński, G. Musumurra, D. Pisano, M. Szafran, *J. Am. Chem. Soc.*, 111 (1989) 7; A. R. Katritzky, V. Feygelman, G. Musumurra, P. Barczyński, M. Szafran, *J. Prakt. Chem.*, 332 (1990) 853; A.R. Katritzky, V. Feygelman, G. Musumurra, P. Barczyński and M. Szafran, *J. Prakt. Chem.* 332 (1990) 870; A. R. Katritzky, P. Barczyński, *J. Prakt. Chem.*, 332 (1990) 885.
89. K. Jug, A. Köster, *J. Phys. Org. Chem.*, 4 (1991) 163.
90. K. Uberla, *Faktorenanalyse*, Springer-Verlag, Berlin, 1977.
91. E. R. Malinowski, D. G. Howery, *Factor Analysis in Chemistry*, J. Wiley-Interscience, New York, NY, 1980.
92. M. Cyrański, T. M. Krygowski, *J. Chem. Inf. Comput. Sci.* 36 (1996) 1142.
93. M. Cyrański, T. M. Krygowski, *Tetrahedron*, 52 (1996) 13795.
94. L. W. Jenneskens, J. C. Klammer, H. J. B. de Boer, W. H. de Wolf, F. Bickelhaupt, C. H. Stam, *Angew. Chem.* 96 (1984) 236.
95. S. Marriott and R.D. Topsom, *J. Mol. Struct. (Theochem)*, 110 (1984) 337.

This Page Intentionally Left Blank

Average Local Ionization Energies: Significance and Applications

Jane S. Murray and Peter Politzer

University of New Orleans, Department of Chemistry, New Orleans,
Louisiana 70148

1. INTRODUCTION

The ionization energy I of an atomic or molecular system is equal to the energy that is required to remove one electron. However the chemical significance of the ionization energy is not limited to the formation of ions. It is also linked to the electronegativity χ ; if χ is written as [1-3],

$$\chi = -\left(\frac{\partial E}{\partial N}\right)_v \quad (1)$$

where E is the total energy of an N -electron system with nuclear potential v , then assuming that it is valid to integrate over N [2,4,5],

$$I = E(N-1) - E(N) = -\int_N^{N-1} \chi dN \quad (2)$$

Furthermore, χ is commonly approximated by [1,3,6,7],

$$\chi = 0.5(I + A) \quad (3)$$

in which A is the electron affinity. Finally, it has been found, at least for atoms, that I correlates with polarizability α [8,9]. Thus the ionization energy bears upon the reactive behavior of atoms and molecules in several different ways.

Chemical reactivity is typically site-specific. For example, certain sites may be more susceptible to electrophilic attack, others to nucleophilic. It could be argued, therefore, that there is a need for being able to determine the ionization energy as a function of position \mathbf{r} in the space of the molecule. We have introduced such a function [10], which is rigorously defined within the framework of the Hartree-Fock molecular orbital model by eq. (4):

$$\bar{I}(\mathbf{r}) = \sum_i \frac{\rho_i(\mathbf{r})|\epsilon_i|}{\rho(\mathbf{r})} \quad (4)$$

$\rho_i(\mathbf{r})$ is the electronic density of the i^{th} molecular orbital at the point \mathbf{r} , ϵ_i is the orbital energy, and $\rho(\mathbf{r})$ is the total electronic density function. (An analogous expression is of course applicable to atoms.)

By Koopmans' theorem, the magnitudes of Hartree-Fock orbital energies are good approximations to the ionization energies of the respective electrons [11]. Accordingly we interpret $\bar{I}(\mathbf{r})$ as the average energy required to remove an electron from any point \mathbf{r} in the space of an atom or molecule; it is the "average local ionization energy" [10]. $\bar{I}(\mathbf{r})$ focuses upon a particular point in space, not upon a particular molecular orbital.

In light of what was said earlier, the concept of local ionization energy suggests, in turn, local electronegativity and local polarizability. All three properties – I , χ and α – are rigorously defined in a global sense, in terms of the entire system; however they can also be conceptualized on a local basis, in which the emphasis is upon the point \mathbf{r} and not upon the highest-energy electron, as is normally the case for I and χ . We have already earlier suggested that $\bar{I}(\mathbf{r})$ might be a measure of local polarizability [9].

2. AVERAGE LOCAL IONIZATION ENERGIES OF ATOMS

Since the binding energies of the electrons in an atom decrease with radial distance from the nucleus, it is to be anticipated that $\bar{I}(\mathbf{r})$ will behave in a qualitatively similar fashion. This is indeed the case [9]; however $\bar{I}(\mathbf{r})$ shows a particularly interesting pattern. It decreases in a roughly stepwise manner, regions of gradual variation being separated by steps of more rapid change. This can be viewed as reflecting the shell structures of the atoms; $\bar{I}(\mathbf{r})$ changes slowly within a given shell, but then decreases sharply in going to the next shell.

This interpretation was confirmed for the atoms Li - Kr by integrating their Hartree-Fock electronic densities over the regions corresponding to the steps [9]; the points of inflection were taken to define the boundaries. For each atom, the number of these regions was found to be the same as the expected number of shells, and the integrated electronic populations were generally very similar to the formal shell occupancies, although with the introduction of $3d$ electrons, this was affected by the increasing degree of interpenetration that occurs between subshells [12]. In a recent study of the elements Li–Xe, Sen *et al* found that $\bar{I}(\mathbf{r})$ does not reveal the fifth electronic shell in the six elements Mo, Ru, Rh, Pd, Ag and Cd [13]. This may reflect the level of the Hartree-Fock wave functions.

Fricke has pointed out that there is a reasonably good inverse correlation between ionization energy and polarizability for the atoms H–Ra [8]. This led us to investigate whether there is a relationship between α and $\bar{I}(\mathbf{r})$ [9]. For this purpose we evaluated $\bar{I}(\mathbf{r})$ on the “surface” of each of the atoms He–Kr, the surface being defined either as (a) the spherical shell enclosing 98% of the electronic charge [14], or (b) the 0.001 au contour of the electronic density [15]. By either approach, α and $\bar{I}(\mathbf{r})$ were found to obey an equation of the form [9],

$$\alpha \bar{I}^a = b \quad (5)$$

in which a and b are constants that depend upon how the atomic surface is defined. Both correlation coefficients were 0.97. It was the existence of such relationships that prompted us to suggest that $\bar{I}(\mathbf{r})$ may be a measure of local variations in polarizability [9].

Finally, Nagy, Parr and Liu have demonstrated similarities, for a group of atoms, between the radial variation of $\bar{I}(\mathbf{r})$ and that of the “local temperature,” $T(\mathbf{r})$ [16]. The latter is defined in terms of the kinetic energy per electron at the point \mathbf{r} [17]. In particular, $T(\mathbf{r})$ also decreases in an approximately stepwise fashion, and the steps occur at about the same points as for $\bar{I}(\mathbf{r})$.

3. AVERAGE LOCAL IONIZATION ENERGIES OF MOLECULES

3.1 Applications to Reactivity

In applying $\bar{I}(\mathbf{r})$ to interpreting and predicting molecular reactive behavior, we felt it to be most reasonable to look at the magnitudes of $\bar{I}(\mathbf{r})$ on the molecular surface. We have defined the surface, following the suggestion of Bader *et al* [15], as either the 0.001 au or 0.002 au contour of the molecular electronic density. We have verified that our results are essentially independent of the choice of contour; $\bar{I}(\mathbf{r})$ decreases by less than 1% in going from the 0.002 au to the 0.001 au [10,18,19]. Defining the surface in terms of the molecular electronic density has the advantage that it reflects features, such as lone pairs, that are specific to the molecule.

The local and absolute minima of $\bar{I}(\mathbf{r})$ on a molecular surface, which we label $\bar{I}_{S,\min}$, correspond to points at which are found, on the average, the least tightly bound electrons. The values of these $\bar{I}_{S,\min}$ are generally greater, typically by 1-3 eV, than either the experimental ionization energy or the magnitude of the highest occupied orbital energy [10,19,20]; this indicates that there is a significant probability of finding inner, more tightly-bound electrons even on the molecular surface.

Since the positions of the $\bar{I}_{S,\min}$ are the locations, on the average, of the most easily removed electrons, these should also be the sites that are most reactive toward electrophilic attack. This has been fully confirmed for a group of monosubstituted benzene derivatives [10,21]. The $\bar{I}_{S,\min}$ correctly predict the *ortho/para*- or *meta*- directing tendencies of the substituents, even the rather unusual NH_3^+ , which is a *meta/para* director [22]. Furthermore, the magnitudes of the $\bar{I}_{S,\min}$ relative to that of unsubstituted benzene correctly indicate whether each substituent activates or deactivates the aromatic ring toward electrophiles. These analyses have been extended to other aromatic systems, including azines and azine N-oxides [18,23,24].

These encouraging results suggested that the $\bar{I}_{S,\min}$ at the *meta* and *para* positions of the benzene derivatives may correlate with the substituents' Hammett constants σ_m and σ_p [25,26]. These are measures of the electron-donating and electron-withdrawing tendencies of substituents on aromatic rings. Excellent linear relationships between $\bar{I}_{S,\min}$ and the Hammett constants have been found at three Hartree-Fock computational levels: STO-5G [23], 6-31G* [10] and 6-31+G** [21]. The correlation coefficients are 0.99. The 3-21G basis set is less successful in this respect, giving a correlation coefficient of 0.94 [10]. The magnitudes of the \bar{I}_S obtained with the various basis sets differ quite significantly, over a range of nearly 3 eV; however it is the trends rather than the actual magnitudes that are important for present purposes. We have accordingly been able to estimate σ_m and/or σ_p for three substituents for which they were not previously available: NHF [10], NF_2 [10] and $\text{N}(\text{NO}_2)_2$ [24].

We have also found reasonably good linear correlations between the Taft inductive constant σ_I [25,26] for a substituent X and (a) the nitrogen $\bar{I}_{S,\min}$ for $\text{H}_2\text{N}-\text{X}$ molecules [27], and (b) the oxygen $\bar{I}_{S,\min}$ in $\text{X}-\text{CH}_2-\text{COO}^-$ anions [23]. The points for $\text{X}=\text{F}$ are outliers [23,27], σ_I being underestimated. We have suggested that our relatively low predicted σ_I for fluorine is consistent with the considerable evidence that fluorine has a limited capacity for accepting additional electronic charge, despite a strong initial attraction for it [28-34]. The higher literature value of σ_I for fluorine, which was obtained from the measured pK_a of fluoroacetic acid [25,26], may be due to anomalous solution effects. For example, $\text{CH}_2\text{F}-\text{COOH}$ is less acidic than $\text{CH}_2\text{Cl}-\text{COOH}$ in the gaseous phase, while the reverse is true in aqueous solution [35].

The demonstrated effectiveness of $\bar{I}_{S,\min}$ in relation to interactions with electrophiles encouraged us to investigate whether it may correlate with pK_a , since H^+ is certainly an electrophile. A possible complication is that pK_a refers to aqueous solution, and our calculations are for single molecules, taking no explicit account of solvent effects. Our approach was to seek

relationships between the measured pK_a 's of acids and the computed (HF/6-31G*) $\bar{I}_{S,\min}$'s of their conjugate bases, since it is the latter that interact with H^+ .


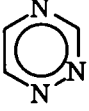







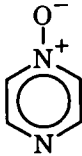

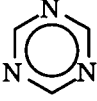
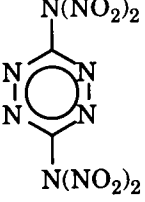
We found that pK_a does indeed correlate with $\bar{I}_{S,\min}$ [19,23,27,35-37], and it is gratifying that a wide variety of acids can be treated together: carboxylic acids, oxoacids, nitrogen acids, substituted methanes and hydrocarbons. (The correlations are of course better when the different families of acids are treated separately.) In all of these, the conjugate base is an anion, and the location of $\bar{I}_{S,\min}$ is usually near the atom from which H^+ has been removed. There is a separate relationship for unsaturated sp^2 -nitrogen-containing heterocycles [23,37], for which the conjugate base is the neutral molecule and the acid is a cation. It is particularly interesting that good correlation coefficients, from 0.97 to 0.99, are obtained despite the fact that the experimental data pertain to aqueous solution and the computed $\bar{I}_{S,\min}$ are for single molecules. These relationships have allowed us to predict pK_a for a number of systems for which it has not been determined experimentally (Table 1). Finally, since $\bar{I}_{S,\min}$ correlates with pK_a , it is not surprising that it does also with gas phase protonation enthalpies [35,36].

A property that has long been used for interpreting and predicting molecular reactivity [38-40] is the electrostatic potential $V(\mathbf{r})$ that is created in the space around a molecule by its nuclei and electrons. $V(\mathbf{r})$ is given rigorously by eq. (6),

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}') d\mathbf{r}'}{|\mathbf{r}' - \mathbf{r}|} \quad (6)$$

in which Z_A is the charge on nucleus A, located at \mathbf{R}_A , and $\rho(\mathbf{r})$ is the electronic density. Unlike $\bar{I}(\mathbf{r})$, which is by definition positive, $V(\mathbf{r})$ can be either positive or negative, depending upon whether the contribution of the nuclei or the electrons is dominant at the particular point. The electrostatic potential has been used extensively for analyzing electrophilic attack, since in principle, the possible sites can be identified and ranked by the locations and magnitudes of the most negative values of $V(\mathbf{r})$. In practice, however, the $\bar{I}_{S,\min}$ are much more effective and specific for this purpose. In aromatic systems, for example, the electrostatic potential above and below the ring does reveal whether substituents and/or heteroatoms have an activating or

Table 1.
Predicted pK_a values.

Molecule	Predicted pK_a	Reference	Molecule	Predicted pK_a	Reference
	43	[36]		-1.77	[37]
	43	[36]		-5.9	[24]
	26	[36]		-10.9	[24]
	32	[36]		-1.0	[18]
	36	[36]		3.1	[18]
$\text{HN}(\text{NO}_2)_2$	-5.6	[35]		-9.3	[24]
	-2.31	[37]		-15.0	[24]

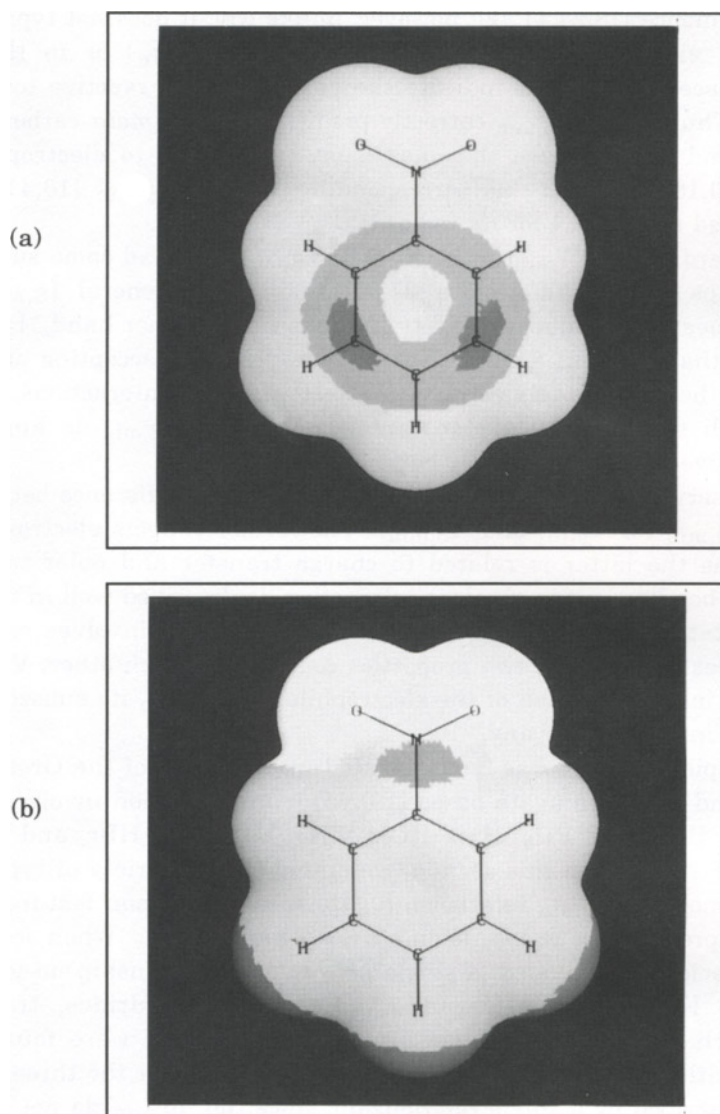


Figure 1. Calculated HF/6-31G* molecular surface properties of nitrobenzene: (a) average local ionization energy $\bar{I}(\mathbf{r})$, in eV; dark gray < 13.01 eV, 13.4 < light gray < 13.88; white > 13.88; and (b) electrostatic potential $V(\mathbf{r})$, in kcal/mole; dark gray > 14.0; 14.0 > light gray > 0.0; white < 0.0.

deactivating influence [18,24,41,42]; however, unlike $\bar{I}(\mathbf{r})$, it does not typically show minima, either on the molecular surface ($V_{S,\min}$) or in three-dimensional space (V_{\min}), that indicate the positions most reactive toward electrophiles. Thus, while $\bar{I}_{S,\min}$ correctly predicts that the *meta* carbons in nitrobenzene and pyridine are the ones most susceptible to electrophilic substitution [10,18], there are no corresponding V_{\min} or $V_{S,\min}$ [18,41,42]. This is illustrated in Figure 1 for nitrobenzene.

With regard to pK_a , V_{\min} (more so than $V_{S,\min}$) has had some success for certain groups of related molecules [27,36,43-45], but in general $\bar{I}_{S,\min}$ is superior as a measure of aqueous acidity [27,36]. On the other hand, $\bar{I}_{S,\min}$ is much worse than V_{\min} as an indicator of hydrogen bond accepting ability [27], reflecting the key role of electrostatic effects in these interactions. For correlating with the substituent constant σ_I , V_{\min} and $\bar{I}_{S,\min}$ do approximately equally well, and significantly better than $V_{S,\min}$ [27].

These observations emphasize what we feel is a key difference between the roles of $V_{S,\min}$ (or V_{\min}) and $\bar{I}_{S,\min}$. The former reflects electrostatic factors, whereas the latter is related to charge transfer and polarization. Thus hydrogen bonding, a non-covalent interaction, is described well in terms of the electrostatic potential [19,27,46], while pK_a , which involves charge transfer, requires $\bar{I}_{S,\min}$. The two properties complement each other: $V_{S,\min}$ determines the initial approach of the electrophile and $\bar{I}_{S,\min}$ its subsequent charge-sharing interaction, if any.

This complementarity was demonstrated in our study of the Group V-VII hydrides and their conjugate bases [19]. Our investigation involved the hydrides NH_3 , H_2O , HF , PH_3 , H_2S , HCl , AsH_3 , H_2Se and HBr , and their conjugate bases. Earlier in this section, we stated that a variety of types of acids fit the same pK_a - $\bar{I}_{S,\min}$ relationship. However a common feature was that the acidic proton in each case is on a first-row atom [35]. When second- and third-row acids are included, a single pK_a - $\bar{I}_{S,\min}$ relationship no longer suffices [19]. For the first-, second- and third-row hydrides, treated separately, both pK_a and gas phase protonation enthalpy were found to correlate well with the $\bar{I}_{S,\min}$ of the conjugate bases. However the three rows could not be described by a single relationship, since the $\bar{I}_{S,\min}$ do not show the correct trends within the vertical groups (V, VI and VII). On the other hand, both pK_a and protonation enthalpy can be represented as linear combinations of $V_{S,\min}$ and $\bar{I}_{S,\min}$:

$$pK_a = \alpha_1 V_{S,\min} + \beta_1 \bar{I}_{S,\min} + \gamma_1 \quad (7)$$

$$\Delta H_{pr} = \alpha_2 V_{S,\min} + \beta_2 \bar{I}_{S,\min} + \gamma_2 \quad (8)$$

In eqs. (7) and (8), the coefficients β_1 and β_2 are considerably larger in magnitude than are α_1 and α_2 ; nevertheless $V_{S,\min}$ is definitely needed to give the correct trends within the vertical groups. In Group VII, for example, $\bar{I}_{S,\min}$ decreases in going from F^- to Cl^- to Br^- , implying increasing reactivity toward H^+ and decreasing acidity. However $V_{S,\min}$ becomes less negative in the same direction, with the opposite implications. The changes in $V_{S,\min}$ are sufficiently greater than those in $\bar{I}_{S,\min}$ that eqs. (7) and (8) do describe the experimentally-observed variations in pK_a and ΔH_{pr} . The reason that $\bar{I}_{S,\min}$ is able to reproduce the trends in pK_a and ΔH_{pr} within each row even without the inclusion of $V_{S,\min}$ is that the latter varies relatively little in going from Group V to Group VII within the same row, whereas $\bar{I}_{S,\min}$ increases considerably. This explains why we were able, earlier, to represent pK_a in terms of $\bar{I}_{S,\min}$ alone for a variety of first-row acids [35].

Our present view is that the interaction with H^+ , and presumably other electrophiles as well, may involve two phases: the initial attraction that brings the H^+ into the vicinity of the molecule or anion, and possibly (but not necessarily) some subsequent degree of charge transfer or polarization. The first step is determined by $V_{S,\min}$ or V_{\min} , the second (to whatever extent it occurs) by $\bar{I}_{S,\min}$. In general, therefore, $V_{S,\min}$ (or V_{\min}) and $\bar{I}_{S,\min}$ play complementary roles; in some cases, however, one or the other may be a greatly dominating factor and may suffice for a quantitative representation of the interaction.

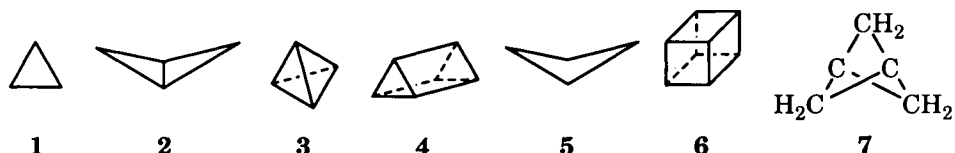
All of the studies discussed in this section have involved Hartree-Fock calculations. We have also carried out an initial investigation of the significance of $\bar{I}(\mathbf{r})$, as defined by eq. (4), when computed by density functional procedures. These have the important advantage of being able to treat much larger systems than can *ab initio* techniques of comparable accuracy, i.e. that include electronic correlation [3,47-51]. However while the widely-used Kohn-Sham density functional approach does produce one-particle orbitals and associated eigenvalues [3,50-52], there is no analogue of Koopmans' theorem [11] to provide assurance that the latter can be viewed as approximations to the electrons' ionization energies. Only for the eigenvalue corresponding to the highest occupied orbital has such an interpretation been made, that it equals (for the exact exchange-correlation potential) the negative of the first ionization energy of the system [2,3,5]. It is accordingly necessary to determine whether $\bar{I}(\mathbf{r})$ has the same sort of significance and relevance to molecular reactivity in Kohn-Sham density functional theory as in Hartree-Fock.

As a first step toward this objective, we computed $\bar{I}(\mathbf{r})$ at the DF/B3P86/6-31+G** level on the molecular surfaces of monosubstituted benzene

derivatives [21]. While the resulting $\bar{I}_{S,\min}$ are smaller in magnitude than the Hartree-Fock values for the same molecules [10,21,23], they do correctly predict the directing tendencies and ring-activating or -deactivating effects of the substituents, and correlate very well with the Hammett constants. Thus the initial indications are encouraging, with respect to computing $\bar{I}(\mathbf{r})$ by density functional methods.

3.2 Characterization of Bonds

In a Hartree-Fock study of the strained hydrocarbons 1 - 7 [20], we found the interesting feature that there are invariably $\bar{I}_{S,\min}$ near the midpoints of the C-C bonds in the three-membered rings of 1 - 4, but not the C-C bonds of the four-membered rings of 4 - 7. In triprismane, 4, which has both types of rings, the bonds in the three-sided ends have $\bar{I}_{S,\min}$; the bonds connecting these ends do not.

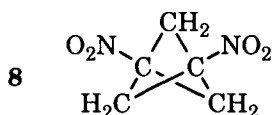


We have pointed out earlier that the C-C bonds in strained hydrocarbons have negative electrostatic potentials near their midpoints [39,53]. (This is not observed for more typical C-C bonds, such as those in ethane and propane.) In general, these potentials are more negative, and the calculated strain in the individual bonds is greater, in three-membered rings than in four-membered rings [53-58]. (We measure bond strain by the extent to which the actual bond path differs from a reference path [53,54,57].)

These factors – $\bar{I}_{S,\min}$, electrostatic potential, and bond strain – are all consistent with a relatively high degree of reactivity toward electrophiles on the part of C-C bonds in three-membered rings. The ethylene molecule also has both a negative potential and an $\bar{I}_{S,\min}$ near the midpoint of its double bond [20,59], and indeed it has long been recognized that cyclopropane and bicyclobutane have some olefin-like properties [60-64].

The $\bar{I}_{S,\min}$, more so than either the electrostatic potential or bond strain, represent a significant and seemingly characteristic difference between the C-C bonds in these saturated three- and four-membered hydrocarbon rings, and a similarity between the former and the C=C bonds in olefins. The fact that no bond $\bar{I}_{S,\min}$ were found for propellane, 7, is consistent with the absence of a bond between the central carbons; if there were one, then there would be three three-membered rings, and presumably $\bar{I}_{S,\min}$ near the midpoints of the C-C bonds. (The existence or non-existence

of a bond between the central carbons in **7** has been a matter of some controversy [58,65-68].) The only $\bar{I}_{S,\min}$ found in **7** are located to the outsides of the central carbons. This is consistent with our earlier conclusions that this molecule has biradical character [58], which is also supported by experimental observations that **7** can (a) undergo polymerization at the central carbons [69-71] and (b) react with N_2O_4 to form **8** [72].



4. SUMMARY

The full significance and range of applications of $\bar{I}(\mathbf{r})$ continue to be explored. When computed on molecular surfaces, it is clearly an effective tool for interpreting and predicting reactivity toward electrophiles, in solution as well as in the gas phase. There are indications that this effectiveness can be increased by combining $\bar{I}_{S,\min}$ with $V_{S,\min}$, the most negative values of the electrostatic potential on the molecular surface. $\bar{I}_{S,\min}$ can also be used to characterize, and hence identify, certain types of chemical bonds. Finally, $\bar{I}(\mathbf{r})$ computed within the internal regions of atoms has been found to reflect the detailed distributions of the electrons. This suggests that analogous analyses of molecules might be fruitful; for example, it might be interesting to examine the variation of $\bar{I}(\mathbf{r})$ along internuclear axes. Such studies are in progress.

ACKNOWLEDGEMENT

We greatly appreciate the support provided by the Eastman Kodak Company.

REFERENCES

1. R. G. Parr, R. A. Donnelly, M. Levy and W. E. Palke, *J. Chem. Phys.* **68** (1978) 3801.
2. J. P. Perdew, R. G. Parr, M. Levy and J. L. Balduz, Jr., *Phys. Rev. Lett.* **49** (1982) 1691.
3. R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, New York, 1989.
4. R. G. Parr and L. J. Bartolotti, *J. Phys. Chem.* **87** (1983) 2810.

5. J. P. Perdew, in *Density Functional Methods in Physics*, R. M. Dreizler and J. da Providencia, eds., Plenum Press, New York, 1985, p. 265.
6. R. S. Mulliken, *J. Chem. Phys.* **2** (1934) 782.
7. R. P. Iczkowski and J. L. Margrave, *J. Am. Chem. Soc.* **83** (1961) 3547.
8. B. Fricke, *J. Chem. Phys.* **84** (1986) 862.
9. P. Politzer, J. S. Murray, M. E. Grice, T. Brinck and S. Ranganathan, *J. Chem. Phys.* **95** (1991) 6699.
10. P. Sjoberg, J. S. Murray, T. Brinck and P. Politzer, *Can. J. Chem.* **68** (1990) 1440.
11. T. A. Koopmans, *Physica* **1** (1933) 104.
12. P. Politzer and K. C. Daiker, *Chem. Phys. Lett.* **20** (1973) 309.
13. K. D. Sen, M. Slamet and V. Sahni, *Chem. Phys. Lett.* **205** (1993) 313.
14. C. W. Kammeyer and D. R. Whitman, *J. Chem. Phys.* **56** (1972) 4419.
15. R. F. W. Bader, M. T. Carroll, J. R. Cheeseman and C. Chang, *J. Am. Chem. Soc.* **109** (1987) 7968.
16. A. Nagy, R. G. Parr and S. Liu, *Phys. Rev. A* **53** (1996) 3117.
17. S. K. Ghosh, M. Berkowitz and R. G. Parr, *Proc. Natl. Acad. Sci.* **81** (1984) 8028.
18. P. Lane, J. S. Murray and P. Politzer, *J. Mol. Struct. (Theochem)* **236** (1991) 283.
19. T. Brinck, J. S. Murray and P. Politzer, *Int. J. Quant. Chem.* **48** (1993) 73.
20. J. S. Murray, J. M. Seminario, P. Politzer and P. Sjoberg, *Int. J. Quant. Chem., Quant. Chem. Symp.* **24** (1990) 645.
21. P. Politzer, F. Abu-Awwad and J. S. Murray, *Int. J. Quant. Chem.*, in press.
22. R. T. Morrison and R. N. Boyd, *Organic Chemistry*, 3rd ed., Allyn and Bacon, Boston, 1973, ch. 11.
23. J. S. Murray, T. Brinck and P. Politzer, *J. Mol. Struct. (Theochem)* **255** (1992) 271.
24. P. Politzer, J. S. Murray, J. M. Seminario and R. S. Miller, *J. Mol. Struct. (Theochem)* **262** (1992) 155.
25. O. Exner, *Correlation Analysis of Chemical Data*, Plenum Press, New York, 1988.
26. C. Hansch, A. Leo and R. W. Taft, *Chem. Rev.* **91** (1991) 165.
27. J. S. Murray, T. Brinck, M. E. Grice and P. Politzer, *J. Mol. Struct. (Theochem)* **256** (1992) 29.
28. J. Hine and N. W. Burske, *J. Am. Chem. Soc.* **78** (1956) 3337.
29. J. E. Huheey, *J. Phys. Chem.* **69** (1965) 3284.
30. K. R. Brower, B. Gay and T. L. Konkol, *J. Am. Chem. Soc.* **88** (1966) 1681.
31. P. Politzer, *J. Am. Chem. Soc.* **91** (1969) 6235.
32. P. Politzer and J. W. Timberlake, *J. Org. Chem.* **37** (1972) 3557.
33. R. S. Evans and J. E. Huheey, *Chem. Phys. Lett.* **19** (1973) 114.
34. P. Politzer, J. E. Huheey, J. S. Murray and M. Grodzicki, *J. Mol. Struct. (Theochem)* **259** (1992) 99.

35. T. Brinck, J. S. Murray and P. Politzer, *J. Org. Chem.* **56** (1991) 5012.
36. J. S. Murray, T. Brinck and P. Politzer, *Int. J. Quant. Chem., Quant. Biol. Symp.* **18** (1991) 91.
37. T. Brinck, J. S. Murray, P. Politzer and R. E. Carter, *J. Org. Chem.* **56** (1991) 2934.
38. P. Politzer and K. C. Daiker, in *The Force Concept in Chemistry*, B. M. Deb, ed., Van Nostrand-Reinhold, New York, 1981, ch. 6.
39. P. Politzer and J. S. Murray, in *Reviews in Computational Chemistry*, Vol. 2, K. B. Lipkowitz and D. B. Boyd, eds., VCH Publishers, New York, 1991, ch. 7.
40. G. Náray-Szabó and G. G. Ferenczy, *Chem. Rev.* **95** (1995) 829.
41. P. Politzer, L. Abrahmsen and P. Sjoberg, *J. Am. Chem. Soc.* **106** (1984) 855.
42. J. S. Murray, K. Paulsen and P. Politzer, *Proc. Ind. Acad. Sci. (Chem. Sci.)* **106** (1994) 267.
43. J. S. Murray, J. M. Seminario and P. Politzer, *J. Mol. Struct. (Theochem)* **187** (1989) 95.
44. P. Nagy, K. Novak and G. Szasz, *J. Mol. Struct. (Theochem)* **201** (1989) 257.
45. P. Politzer and J. S. Murray, *Trans. Amer. Cryst. Assoc.* **26** (1990) 23.
46. J. S. Murray and P. Politzer, in *Quantitative Treatments of Solute/Solvent Interactions*, J. S. Murray and P. Politzer, eds., Elsevier, Amsterdam, 1994, ch. 8.
47. R. M. Dreizler and E. K. U. Gross, *Density Functional Theory: An Approach to the Quantum Many-Body Problem*, Springer-Verlag, Berlin, 1990.
48. E. S. Kryachko and E. V. Ludeña, *Density Functional Theory of Many Electron Systems*, Kluwer, Dordrecht, The Netherlands, 1990.
49. J. M. Seminario and P. Politzer, eds., *Modern Density Functional Theory*, Elsevier, Amsterdam, 1995.
50. R. G. Parr and W. Yang, *Ann. Rev. Phys. Chem.* **46** (1995) 701.
51. J. M. Seminario, ed. *Recent Developments and Applications of Modern Density Functional Theory*, Elsevier, Amsterdam, 1996.
52. W. Kohn and L. J. Sham, *Phys. Rev. A* **140** (1965) 1133.
53. P. Politzer and J. S. Murray, in *Structure and Reactivity*, J. F. Liebman and A. Greenberg, eds., VCH Publishers, New York, 1988, ch. 1.
54. P. Politzer, L. Abrahmsen, P. Sjoberg and P. R. Laurence, *Chem. Phys. Lett.* **102** (1983) 74.
55. P. Politzer, L. N. Domelsmith, P. Sjoberg and J. Alster, *Chem. Phys. Lett.* **92** (1982) 366.
56. P. Politzer, L. N. Domelsmith and L. Abrahmsen, *J. Phys. Chem.* **88** (1984) 1752.
57. P. Politzer, K. Jayasuriya and B. A. Zilles, *J. Am. Chem. Soc.* **107** (1985) 121.
58. P. Politzer and K. Jayasuriya, *J. Mol. Struct. (Theochem)* **135** (1986) 245.
59. J. Almlof and A. Stogard, *Chem. Phys. Lett.* **29** (1974) 418.

60. A. D. Walsh, *Trans. Faraday Soc.* **45** (1949) 179.
61. N. H. Cromwell and M. A. Graff, *J. Org. Chem.* **17** (1952) 414.
62. K. B. Wiberg, *Rec. Chem. Prog.* **26** (1965) 143.
63. K. B. Wiberg, G. M. Lampuran, R. P. Ciula, D. S. Connor, P. Schertler and J. Lavanish, *Tetrahedron* **21** (1965) 2749.
64. M. Charton, in *Chemistry of the Alkenes*, J. Zabicky, ed., Vol. 3, Wiley-Interscience, New York, 1970, ch. 10.
65. M. D. Newton and J. M. Schulman, *J. Am. Chem. Soc.* **94** (1972) 773.
66. K. B. Wiberg and F. H. Walker, *J. Am. Chem. Soc.* **104** (1982) 5239.
67. K. B. Wiberg, *Acc. Chem. Res.* **17** (1984) 379.
68. J. E. Jackson and L. C. Allen, *J. Am. Chem. Soc.* **106** (1984) 591.
69. A.-D. Schluter, *Macromolecules* **21** (1988) 1208.
70. P. Kaszynski and J. Michl, *J. Am. Chem. Soc.* **110** (1988) 5225.
71. J. Belzner, U. Bunz, K. Semmler, G. Szeimies, K. Opitz and A.-D. Schluter, *Chem. Ber.* **122** (1989) 397.
72. K. B. Wiberg and B. Ross, in *Proc. Eighth Annual Working Group Institute on Synthesis of High Energy Density Materials*, U.S. Army Research, Development and Engineering Center, Dover, NJ, 1990, p. 214.

Intrinsic Proton Affinity of Substituted Aromatics

Zvonimir B. Maksić ^a and Mirjana Eckert-Maksić ^b

^aQuantum Chemistry Group, Department of Chemistry, Rudjer Bošković Institute, P.O.B 1016, 10001 Zagreb, Croatia and Department of Physical Chemistry, Faculty of Science and Mathematics, University of Zagreb, Marulićev trg 19, 10000 Zagreb, Croatia

^bPhysical Organic Chemistry Laboratory, Department of Chemistry, Rudjer Bošković Institute, P.O.B 1016, 10001 Zagreb, Croatia

It is shown that the $MP2(fc)/6-31G^*//HF/6-31G^* + ZPE(HF/6-31G^*)$ model reproduces very well the experimental proton affinities in a large number of substituted benzenes and naphthalenes. Extensive applications of this model revealed that the proton affinity of polysubstituted aromatics followed a simple additivity rule, which have been rationalized by the ISA (independent substituent approximation) model. Performance of this model is surprisingly good. Applications of proton affinities, obtained by the transparent and intuitively appealing ISA model, in interpreting directional ability of substituents in the electrophilic substitution reactions of aromatics are briefly discussed.

1. Introduction

Proton transfer reactions play very important role in chemistry and biochemistry [1-3]. Considerable attention has been focused on the gas phase reactions in the last decades, since they are free of the solvent "pollution" thus being related to the intrinsic reactivity [4-6]. In particular, investigations of gas-phase acidities and basicities were some of the major undertakings in the field [7,8]. The proton affinity (*PA*), on the other hand is an interesting thermodynamic property by itself. It gives useful information on the electronic structure of base in question and serves as an indicator of the electrophilic substitution susceptibility of aromatic compounds [9]. It is the aim of this article to describe some recent advances in theoretical calculations of the proton affinities of substituted aromatics. We shall particularly dwell in more detail on the additivity rules, which enable simple and quick estimates of *PAs* in heavily substituted benzenes and naphthalenes. Some prospects for future developments will be briefly discussed too.

2. Absolute Proton Affinities

2.1. Experimental Basicity Scales

The gas phase basicity and proton affinity are intimately related entities being defined by the same hypothetical reaction:



where B and BH^+ stand for a base and its conjugated acid, respectively. The gas phase basicity is then defined as the negative of the free energy change, whereas PA is given by the negative of corresponding enthalpy change. The most popular way of determining PA s is based on measurements of the equilibrium constants of gas phase proton transfer reactions:



If the entropy change of reaction (2) can be reliably estimated, then relative PA of two bases B_1 and B_2 follows straightforwardly. Unfortunately, eqn. (2) cannot provide absolute values of the proton affinity as a rule. For this purpose standard base B_1 has to be chosen, possessing important property that its conjugated acid B_1H^+ can be generated in a mass spectrometer or by other experimental techniques like ion cyclotron resonance or flowing afterglow technique enabling measurements of its enthalpy of formation. Concomitantly, different selections of the anchor base B_1 lead to different ladders of the proton affinity [7,10,11]. A recently recommended absolute gas-phase proton affinity scale is based on the absolute value of the PA of CO (141.9 kcal/mol) [11]. It is noteworthy that this PA ladder is in good accordance with the theoretical estimates obtained by the very sophisticated G2 *ab initio* scheme [12]. There is, however, another drawback of the experimental approach in determining PA s in polyfunctional molecules. The measured data are related to the thermodynamically most stable place of protonation leaving information about alternative sites of attack to be desired. Modern computational methods of quantum chemistry [13] provide a very useful complementary approach, particularly since they treat all protonated forms on an equal footing.

2.2. Theoretical Models for Calculating Absolute PA s

It was mentioned already that very precise PA values within the so called chemical accuracy (1-2 kcal/mol) are obtained by the highly involved G2 procedure [12]. Since this intricate theoretical approach is not practical for large(r) molecular systems, a lot of efforts have been devoted to select a more feasible computational scheme capable of reproducing PA s in substituted aromatics [14-17]. It turns out that simpler procedures like G2(MP2) perform very well, but they are still not economical enough to be applicable in large polysubstituted aromatics [14]. The density functional methods (DFT) are more efficient, but they offer results which are not as accurate as one might wish at the present stage of development [14,16]. In fact the proton affinity could provide a useful hint in finding the optimal combination of the exchange and correlation functionals in DFT methods. Some recent attempts are encouraging in this respect [17]. It is gratifying, however, that a relatively simple *ab initio* MO model yields proton affinities of substituted aromatics in very good agreement with available experimental data [18-20]. This model deserves attention and it is described below.

The systematic procedure for calculating theoretical gas-phase basicity (GB) and the proton affinity (PA) is described by Chung-Phillips *et.al.* [21,22] and does not need to be repeated here in detail. In a nutshell, GB and PA for protonation process (1) are given by relations:

$$GB = -\Delta G \quad PA = GB - T\Delta S \quad (3)$$

where ΔG and ΔS are changes in the Gibbs free energy and entropy, respectively. Further, ΔG is related to the change in enthalpy ΔH :

$$\Delta G = \Delta H - T\Delta S \quad (4)$$

where

$$\Delta H = E_{el}(BH^+) - E_{el}(B) + \Delta_{ZPE} + \Delta(E - E_0) - 1.48 \text{ kcal/mol} \quad (5)$$

and

$$-T\Delta S = -T[S(BH^+) - S(B)] + 7.76 \text{ kcal/mol} \quad (6)$$

Here, the total electronic energy is given by E_{el} , Δ_{ZPE} denotes a change in the zero point energy (ZPE) between the conjugated acid and base, whereas $\Delta(E - E_0)$ represents the change in the internal energy between these two molecular systems in going from 0 K to 298 K. The energy E is given by the sum of four contributions $E = E_0 + E_{trans} + E_{rot} + E_{vib}$, where E_0 stands for the energy at 0 K thus being equal to $E_0 = E_{el} + ZPE$. Concomitantly, by using eqn.(3) the proton affinity PA is determined by expression:

$$PA = [E_{el}(B) - E_{el}(BH^+)] + [ZPE(B) - ZPE(BH^+)] + c(T) \quad (7)$$

where the temperature-dependent correction term $c(T)$ is given by $c(T) = T[S(BH^+) - S(B)] - \Delta E_{vib}(T) - 6.28 \text{ kcal/mol}$, since $\Delta E_{trans}(T) = \Delta E_{rot}(T) = 0$. It appears that $c(T)$ for $T=298 \text{ K}$ is fairly constant for protonated atoms of the same element within a family of related molecules [22], being relatively small at the same time [11]. We shall neglect $c(T)$ term keeping in mind that it just shifts the scale of the PA values being completely irrelevant if the trend of changes in the proton affinity is desired. To put it in another way, by gauging the theoretical model, which refers to $T = 0 \text{ K}$ and trying to reproduce the experimental PA s by using eqn.(7) with $c(T) = 0$, the temperature dependent correction term will be automatically included in the model of the choice. After some numerical experiments the latter appeared to be $MP2(fc)/6 - 31G^{**}/HF/6 - 31G^* + ZPE(HF/6 - 31G^*)$ model which implies that geometries and ZPE s were determined at the simple $HF/6 - 31G^*$ level. Vibrational frequencies were multiplied by the standard empirical weighting factor of 0.89 [23]. Explicit inclusion of the ZPE s is important for quantitative description of the absolute values of PA s since the protonated forms have one more atom and an additional chemical bond. Equally important for the absolute PA is an estimate of the correlation energy in aromatic moieties. This is achieved within the adopted model by the single point $MP2(fc)/6 - 31G^{**}/HF/6 - 31G^*$ calculation, where (fc) denotes frozen $(1s)^2$ core electrons in the course of the Møller-Plesset second order perturbation calculations. It should be mentioned that the use of larger $6 - 31G^{**}$ basis set is plausible in the final single point calculations, since a good description of H atoms in the protonation process is mandatory for reasonable performance of the model. It is noteworthy that the ZPE difference between BH^+ and B varies very little along the series of similar molecules implying that it can be neglected when relative PA s are considered. This is concomitant with findings of Schulman *et al.* [24] that ZPE is a quite robust entity and that it does not depend critically on the finer details of the electronic structure of molecules. In fact, ZPE

is a linear function of the number of atoms of each element constituting molecules under scrutiny, which holds to good accuracy. It follows as a corollary that the $HF/6-31G^*$ model is quite sufficient for estimating $ZPEs$ in "well behaving" molecules, being a useful by-product of the true minima search and checking procedures on the potential energy surfaces. An illustrative example which shows that the $HF/6-31G^*$ $ZPEs$ are reliable is provided by benzene and its σ -protonated Wheland complex (benzenium ion). We found that the difference in $ZPEs$ between benzene and benzenium ion was 7.1 and 6.9 (in kcal/mol) by the $MP2(fc)/6-31G^*$ and $HF/6-31G^*$ models, respectively [18]. These two values are practically indistinguishable thus supporting a use of the simpler and more economical Hartree-Fock model.

Finally, it should be mentioned that the $MP2(fc)/6-31G^*//HF/6-31G^* + ZPE$ ($HF/6-31G^*$) model of choice was selected by employing benzene and phenol as gauge molecules (*vide infra*). Its application to substituted aromatics was justified *a posteriori* by good agreement with available experimental data. It turned out that the model abbreviated heretofore as MP2(I) satisfied both criteria for a suitable vehicle in exploring PAs : it was practical enough to be feasible in large aromatic systems being at the same time reasonably accurate for a large variety of substituents. Naturally, each model has its limitations. Refinements of the MP2(I) model will be discussed at the later stage. All calculations have been performed by using GAUSSIAN 92 and GAUSSIAN 94 programs [25,26].

2.3. Proton Affinities in Monosubstituted Benzenes

Theoretical absolute proton affinities obtained for monosubstituted benzenes C_6H_5X are presented in Table 1. The examined substituents encompass $X = CH_3, OH, OCH_3,$

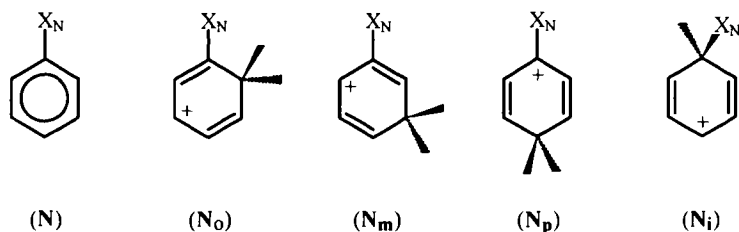


Figure 1. Schematic representation of benzene, its substituted derivatives and their ortho, meta, para and ipso protonated forms denoted by (No), (Nm), (Np), and (Ni) respectively. Here X_N represents the following atoms or groups - (X_N, N): (H, **1**), (OH, **2**), (OCH_3 , **3**), (CH_3 , **4**), (F, **5**), (CF_3 , **6**), (CHO, **7**), ($COOCH_3$, **8**), (CN, **9**), (NO_2 , **10**), (NO, **11**) and (Cl, **12**)

$F, CF_3, CHO, COOCH_3, NO_2$ and NO [18,19,27,28] (Fig.1). We commence discussion with benzene (**1**) and phenol (**2**). The MP2(I) model yields PA for benzene 179.9

Table 1

Proton affinities of monosubstituted benzenes (1) - (12) as obtained by the MP2(I) \equiv MP2(fc)/6-31G**//HF/6-31G* + ZPE(HF/6-31G*) model (in kcal/mol)^a.

System	PA[MP2(I)]	System	PA[MP2(I)]	System	PA[MP2(I)]
(1)	179.9[180.0] ^b	(5i)	156.7	(9i)	156.8
(2o)	193.0	(6o)	170.0	(9:N)	195.7 [195.7] ^d
(2m)	179.9	(6m)	169.5	(10o)	163.5
(2p)	195.5[195.0] ^c	(6p)	169.1	(10m)	162.6
(2i)	162.2	(6i)	168.0	(10p)	163.7
(2:O)	182.7	(7o)	172.7	(10i)	157.3
(3o)	197.9	(7m)	171.8	(10:O)	194.2[193.4] ^d
(3m)	183.0	(7p)	171.6	(11o)	168.2
(3p)	200.2[200.3] ^d	(7i)	167.9	(11m)	167.2
(3i)	165.4	(7:O)	199.7[200.2] ^d	(11p)	167.2
(3:O)	192.0	(8o)	176.0	(11i)	163.8
(4o)	186.2	(8m)	174.9	(11:N)	206.1[204.8] ^d
(4m)	182.9	(8p)	174.8	(11:O)	200.4
(4p)	187.3[189.1] ^d	(8i)	172.8	(12o)	178.6
(4i)	179.9	(8:O)	207.3[205.3] ^d	(12m)	172.2
(5o)	179.4	(9o)	166.8	(12p)	180.3[181.7] ^d
(5m)	172.5	(9m)	164.0	(12i)	164.0
(5p)	181.6[182.9] ^c	(9p)	166.7		

^a Theoretical results are taken from refs. [18,19,27,28]. Experimental values are given within the square parentheses. Protonation at the heteroatom is denoted by (N:X_N).

^b Ref. [11].

^c Ref. [29].

^d Ref. [7].

kcal/mol which is practically identical to the measured value [11]. This is important since benzene will serve as a reference level in estimating substituent effect on the proton affinity. Phenol (2) is a substituted benzene *par excellance* possessing OH group which exerts a strong lone pair π -back bonding effect by donating some electron density to the benzene π -orbital manifold. This type of substituents activates ortho and para positions of the ring making them more susceptible to the proton attack. This can be easily rationalized at the qualitative level by the corresponding resonance structures [19]. Actual *ab initio* calculations of the π -densities confirm this intuitive conjecture. The question arises, however, whether phenol is an oxygen or carbon base. The experimental evidence indicates that phenol protonates on the ring at the position para to the OH substituent. However, protonation at oxygen atom is also possible in solutions [30]. Moreover, DeFrees *et al.* [31] found, by using a particularly designed labeling experiments, that the oxygen proton affinity of phenol is some 13-20 kcal/mol smaller than the PA of the ring. Theory is capable to provide useful complementary information nowadays. Firstly, it appears that para position is the most favorable site for protonation, the second being the ortho

position (Table 1). Further, protonation at oxygen is energetically less prolific by 12.8 kcal/mol than the ring attack indicating that the lower experimental limit of DeFrees *et al.* [31] is the right value. On the other hand, the oxygen protonation is more propitious than either meta or ipso protonation. Putting together all these findings one can say that oxygen might become a competitive basic site in polar solvents. A very useful by-product of theoretical calculations are geometries of the protonated forms, which are very difficult to trap experimentally unless appropriate salts are prepared. Concomitantly, the most reliable information on the structure of conjugated acids is offered by theory (*vide infra*). As far as phenol is concerned, it is interesting to note that the most stable oxygen protonated form has C_s scissors conformation, the symmetry plane being coincident with the heavy atom framework. The transition structure (TS) between the two equivalent conformations is of the C_{2v} symmetry being as low as 0.9 kcal/mol [18].

Another very important facet of theoretical considerations is interpretation of the phenomena under study. This is accomplished by quantum mechanical models, which yield simplified but quintessential picture of the molecular behavior [32]. In order to rationalize ortho-para directional ability of OH group let us make an inspection of the valence bond (VB) resonance structure of the benzenium ion (Fig.2). It appears that a depletion

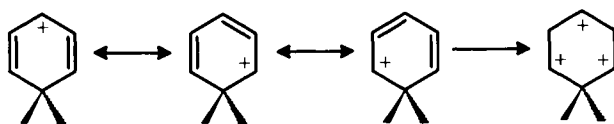


Figure 2. Distribution of the positive charge in the benzenium cation as a consequence of resonance. The dominant VB resonance structure is given on the far left. The resulting overall arrangement of the positive charge is described on the far right.

of the electron density takes place at ortho and para positions. This pattern of charge distribution, caused by formation of sp^3 carbon center and annihilation of one π -electron of the aromatic sextet, has a decisive effect on the preferential sites of the proton attack in e.g. benzenes fused to small ring(s). Analogously, OH group donates π -electron density to the ortho position as illustrated by the VB resonance structures (Fig.3) It follows that both ortho and para protonated forms yield a favorable distribution of alternating atomic charges in a nice and concerted way (Fig.4). One concludes that the advantageous ortho/para protonation occurs because of at least two reasons:(a) increase of the electron density at these positions due to the resonance effect caused by a substituent, which leads to favorable Coulomb interactions with the proton (the ground state effect) and (b) synergistic interaction of the negative charges induced by conjugation with positive charges produced by protonation and a concomitant formation of the sp^3 center in the conjugated acid (combined ground and the final state effect). It is easily retrieved that the latter is a consequence of the significant compatibility between the partial π -bond fixation caused

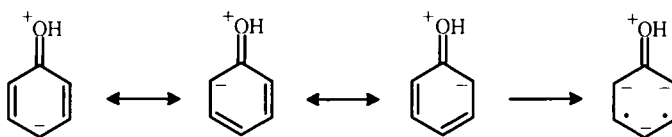


Figure 3. The π -electron back bonding effect in phenol.

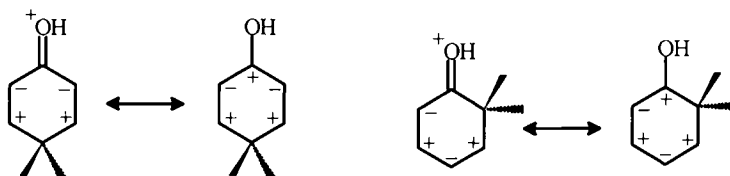


Figure 4. Cooperative overall distribution of the π -electron charges caused by the oxygen π -back bonding resonance effect and by the ortho/para protonation.

by the OH substituent and the sp^3 protonated center. In contrast, meta protonation leads to antagonistic charge demands between substituent and the protonated center as illustrated by Fig.5

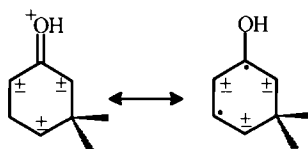


Figure 5. Disconcerted and noncooperative charge demand for the meta protonation.

Obviously, combined effect of the OH substituent and the proton meta attack is counteractive resulting in a less stable conjugated acid. An instructive case is provided by anisole (**3**) since it sheds light on the role of the alkyl group within alkoxy substituent. Perusal of data presented in Table 1 reveals that protonation at the ring positions in (**3**) is energetically more favorable by 3 – 5 kcal/mol relative to phenol. This feature is

easily understood in terms of the hyperconjugation of the CH_3 group leading to additional π -electron density transfer to the aromatic moiety [28] thus increasing the ground state electrostatic effect. The higher PA value for oxygen protonation in anisole (**3**) by 9 kcal/mol relative to phenol (**2**) illustrates the third important contribution to the proton affinity, which is a typical final state effect. This is a charge reorganization upon protonation. The electron density is relaxed in the conjugated acid in order to screen the created positive charge. Analysis of the $HF/6-31G^*$ wavefunctions shows that the oxygen electron density is actually higher in phenol than in anisole implying that the ground state Coulomb stabilization in the former compound is larger provided the density distribution of the rest of molecule is kept frozen. However, redistribution of the electron density in the final state toward the "positive hole" in anisole is obviously larger in view of the presence of the bulky CH_3 group leading eventually to higher PA value. This effect was apparent in a study of the proton affinity of some alcohols and ethers [18]. The PA s increased upon substitution by the alkyl group(s): the bulkier the substituent the higher the proton affinity. The alkyl groups served as "reservoirs" of the electron density, which in turn was drifted toward oxygen atom suffering the electron depletion due to the proton attack. In fact, the reorganization effect is so appreciable that the oxygen atom has larger electron density in the conjugated acid than in the initial base as a rule [18]. To put it in another way - bulkier substituent(s) is(are) more capable to accommodate the positive charge caused by protonation in a way which decreases the Coulomb repulsion between the positively charged atoms. Needless to say, multiple alkyl substitution increases a number of the relaxation "charge flow channels" thus increasing the reorganization energy.

A similar pattern of the ring protonation behavior like OH and OCH_3 exhibit also fluorine, chlorine and methyl substituents. A completely different picture is found in benzenes substituted by the π -electron accepting groups as exemplified by CHO and $COO(CH_3)$ atomic groupings as well as by the strong σ - and π -withdrawing substituents like CN , NO_2 , and NO groups. Let us consider the cyano group as a typical case. The resonance structures describing the π -electron depletion of the aromatic moiety are given in Fig.6: Our computations show that cyanobenzene protonates at nitrogen atom since

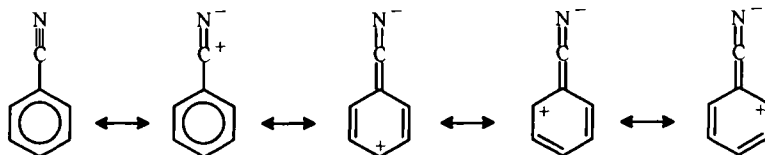


Figure 6. Valence bond resonance structures describing distribution of the positive charge around the aromatic perimeter and a polarization of the CN bond in cyanobenzene.

the estimated PA value by the MP2(I) model is in excellent agreement with experiment (Table 1). This is plausible since a strong electron withdrawing group like CN should

appreciably deactivate the ring positions as evidenced by the lower (**9o**), (**9m**), (**9p**) and (**9i**) *PA* values, which are all substantially below the benzene value. On the other hand, it is interesting to observe that the protonation at the heteroatom (195.7 kcal/mol) is much more favorable than e.g. in *HCN* as evidenced by its experimental value of 171.4 kcal/mol [33]. This is presumably a consequence of the increased resonance effect between the *CN* group and benzene fragment in (**9:N**) as easily derived from the VB structures and *ab initio* π -bond orders [27]. It is also reasonable to assume that the relaxation energy in (**9:N**) is significantly larger than in *HCN*. It should be finally stressed that protonation at heteroatom preserves a good deal of aromaticity of the benzene ring. On the contrary, the ring proton attack leads to the aromaticity defect because of formation of the *sp*³ center. Similarly the ring positions in nitrobenzene (**10**) and nitrosobenzene (**11**) are deactivated thus making them relatively strong oxygen and nitrogen bases, respectively. Analysis of the bonding descriptors (π -bond orders and densities) reveals that the heteroatom protonation increases the resonance interaction between the substituent group and benzene moiety in both cases.

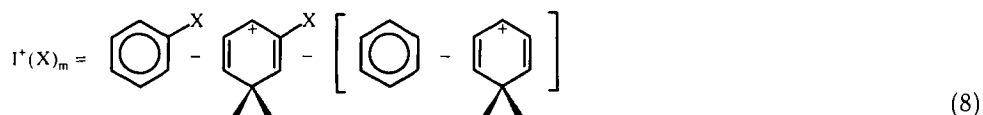
It is noteworthy that all our efforts to find a minimum on the potential energy surface of nitrobenzene protonated at nitrogen were futile since the proton always found its way to oxygen, where the absolute minimum was located. This is apparently a consequence of a substantial positive charge placed on nitrogen atom. In contrast, both heteroatoms are competitive in nitrosobenzene, the proton affinity of oxygen being lower by 4.5 kcal/mol. We mention as a final comment that the ipso protonation in (**9**), (**10**) and (**11**) is the least favorable site of the proton attack due to depleted electron density and an out of plane shift of the substituent group, which in turn leads to significant puckering of the aromatic ring. Benzaldehyde (**7**) and the benzoic acid derivative (**8**) protonate both at carbonyl oxygen atom. Theoretical estimates are in good accordance with the experimental data (Table 1). It is interesting to stress that *CHO* and *COOCH*₃ substituent groups deactivate the ring positions, the effect of the former substituent being more pronounced. Similarly, *CF*₃ group decreases *PA* values of the benzene sites quite uniformly by roughly 10 kcal/mol. This will be discussed in some more detail in the following section. We conclude with a general observation that the electron density withdrawing substituents *CF*₃, *CHO*, *COOCH*₃, *CN*, *NO*₂ and *NO*, which act both *via* inductive (-I) and resonance mechanism (-M) [34], exhibit practically no directional ring effect regarding the (electrophilic) proton attack since the calculated *PA* values are very close, ipso protonation being an exception for reasons mentioned above. This is in contradiction with the experimental data indicating a strong meta directional ability for a number of these substituents in the electrophilic reactions. This discrepancy deserves attention and should be better understood by additional investigations.

2.4. Proton Affinities in Polysubstituted Benzenes—The Additivity Rule

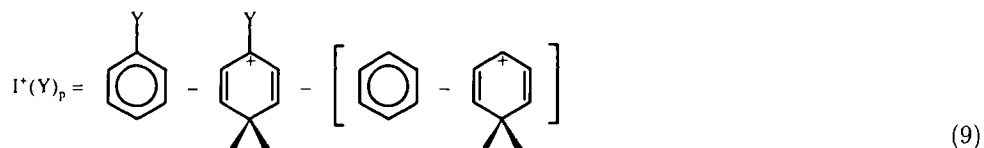
2.4.1. Increments

Before proceeding further we introduce increments which describe a change in the aromatic ring proton affinities in monosubstituted benzenes. This is most conveniently done by using homodesmotic reactions [35]. In this type of hypothetical reaction the number of atoms of the same element, specific types of covalent bonds and approximate hybridization states are preserved being equal in reactants and products. This is very important since it

diminishes the role of the electron correlation to a minimum in contrast to its paramount significance in chemical reactions in general. Furthermore, shortcomings introduced by imperfect theoretical models (e.g. by employing modest basis sets etc.) are minimized since errors tends to cancel out to a large extent. Increments $I^+(X)_m$ and $I^+(Y)_p$ caused by meta and para positioned substituents X and Y respectively are given by:



and



Increments of substituents considered in this article are given in Table 2. Perusal of the

Table 2

Increments describing the change in the *PA* of benzene caused by a presence of the substituent at a particular position of the ring as obtained by the MP2(I) model (in kcal/mol).

Substituent(X)	$I^+(X)_o$	$I^+(X)_m$	$I^+(X)_p$	$I^+(X)_i$	$I^+(X)_{av}^a$
<i>OH</i>	13.1	0	15.6	-17.7	8.4
<i>OCH</i> ₃	18.0	3.1	20.3	-14.5	12.5
<i>CH</i> ₃	6.3	3.0	7.4	0	5.2
<i>F</i>	-0.5	-7.4	1.7	-23.2	-2.8
<i>Cl</i>	-1.3	-7.7	0.4	-15.9	-3.5
<i>CF</i> ₃	-9.9	-10.4	-10.8	-11.9	-10.3
<i>CHO</i>	-7.2	-8.1	-8.3	-12.0	-7.8
<i>COOCH</i> ₃	-3.3	-5.0	-5.1	-7.1	-4.3
<i>CN</i>	-13.1	-15.9	-13.2	-23.1	-15.3
<i>NO</i> ₂	-16.4	-17.3	-16.2	-22.6	-16.7
<i>NO</i>	-11.7	-12.7	-12.7	-16.1	-12.3

^a The average increments are defined by eqn.(10).

presented data offer several interesting conclusions. The ipso positions is dramatically deactivated except in the case of a methyl group. In the latter case the proton affinity of benzene remains unchanged, which is fortuitous. In what follows we shall focus on increments of the ortho, meta and para sites. The largest activation of the benzene ring is produced by the *OCH*₃ substitution implying that methoxy group substantially increases basicity of the benzene ring. The same holds for the hydroxy group but to a lesser extent.

Curiously enough, its meta substitution does not affect the basicity of benzene at all. The methyl group belongs to this category of activating substituents, but its influence seems to be considerably smaller. Substituents OCH_3 , OH and CH_3 possess a strong ortho - para directing property. Halogen F and Cl atoms activate the para position and deactivate the remaining benzene sites for the electrophilic attack. Thus, they represent the borderline cases since the rest of the atomic groupings dramatically decrease basicity of the benzene framework leading to heteroatomic bases as a rule. The largest deactivation is exerted by the NO_2 group. It would be useful to compare the calculated increments with some experimental data. This is possible only for the energetically most favorable modes of protonation. Thus, our $I(OCH_3)_p$, $I(OH)_p$, $I(CH_3)_p$, $I(F)_p$ and $I(Cl)_p$ increments are in good agreement with the measured $-\Delta H_f^0$ values relative to benzene obtained by the equilibrium proton transfer reactions at 600 K, which assume values 17.8, 13.4, 8.4, 1.3 and 1.1 (in kcal/mol), respectively, the measured values being systematically somewhat lower [29]. Further, measurements by Mason *et al.* [36] show that para protonation in fluorobenzene is more profitable than ortho protonation by 1.1 kcal/mol. This is in accordance with a difference $I^+(F)_p - I^+(F)_o = 2.2$ kcal/mol. On the other hand, empirical evidence based on the electrophilic substitution reactivity indicates that both ortho and para centres are deactivated in halobenzenes relative to benzene [9]. This contention is not corroborated by our results for F and Cl substituents (Table 2). In fact, the energy profile curve for the electrophilic reaction in fluoro- and chlorobenzene should be split for ortho and para substitutions, the former lying somewhat above and the latter below the calibration free benzene curve. Utilizing the proton as a probe of the electronic structure and reactivity of the phenyl ring, one can quantify the average activation/deactivation effect of various substituents by the following $I^+(X)_{av}$ index:

$$I^+(X)_{av} = \frac{1}{5} \sum_{\alpha \neq i} I^+(X)_\alpha \quad (10)$$

where the summation runs over all positions of the phenyl fragment excluding the ipso position, and the total sum is divided by the number of terms in the sum. The ipso position is omitted because protonation at this site considerably perturbs the ring by its puckering and by changing the substituent itself in view of the rehybridization within the $C - X$ bond. Obviously, this site has to be treated separately. The calculated average increments $I^+(X)_{av}$ presented in Table 2 support conjectures discussed above reached by inspection of the VB structures and by survey of the separate increments. In particular, dramatic decrease in stability of the benzene ring in general and its diminished ability to protonate caused by NO_2 and CN substituents is reflected in the corresponding $I^+(NO_2)_{av}$ and $I^+(CN)_{av}$ values.

Finally, we address the question of reliability of the increments estimated by the MP2(I) model. Let us consider phenol as a typical example. Increments obtained by various models are presented in Table 3. It appears that the increments are quite insensitive on the basis set used provided a model of the MP2 level of sophistication is employed. Their robustness is illustrated by very similar $I^+(OH)$ values, if the electron correlation is taken explicitly into account. In contrast, the $HF/6 - 31G^*$ increments are sometimes at variance with the MP2 results.

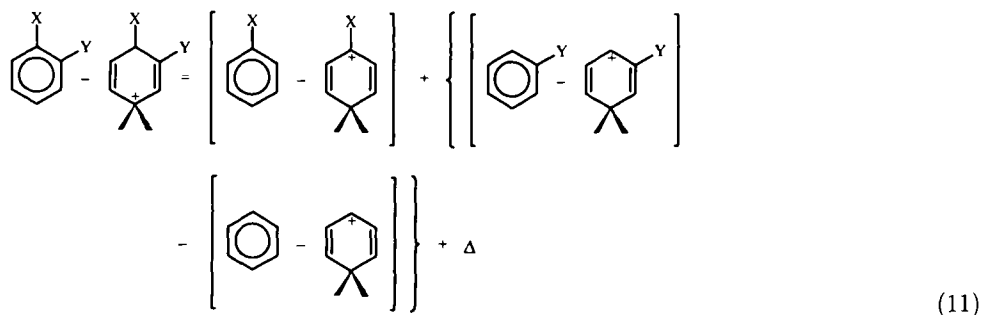
Table 3

Dependence of the *PA* increments in phenol on the theoretical model used (in kcal/mol)

Model	$I^+(OH)_o$	$I^+(OH)_m$	$I^+(OH)_p$	$I^+(OH)_i$	$I^+(OH)_{av}$
HF/6 - 31G*	13.0	-3.7	16.1	-15.8	6.9
MP2(fc)/6 - 31G**//	12.8	-0.4	15.13	-18.2	8.0
HF/6 - 31G*					
MP2(fc)/6 - 31G**//	13.1	0.0	15.6	-17.7	8.4
HF/6 - 31G*					
MP2(fc)/6 - 311G**//	13.7	0.4	16.4	-18.0	8.9
HF/6 - 31G*					

2.4.2. Disubstituted Benzenes - the Independent Substituent Approximation

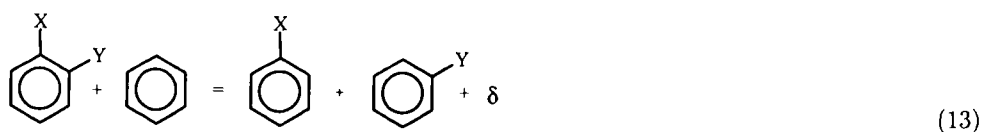
Additivity of PAs in polysubstituted benzenes is easily obtained by means of homodesmotic chemical reactions, if it is assumed that the interaction between substituents is either reasonably small or comparable in the initial base and the final state conjugated acid. Consider for example the *PA* of 1,2-disubstituted benzene, which can be resolved by employing the corresponding homodesmotic reaction in a following way:



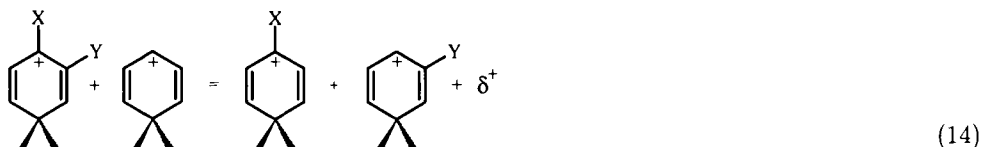
Adding and subtracting the proton affinity of benzene $PA(\text{benzene})$ on the right hand side of eqn.(11) one obtains:

$$PA(C_6H_4X_pY_m) = PA(\text{benzene}) + I^+(X)_p + I^+(Y)_m + \Delta \quad (12)$$

where $I^+(X)_p$ and $I^+(Y)_m$ are increments due to substituents X and Y attached at para and meta positions, respectively, defined by eqns. (8 and 9). It is easy to see that Δ is given by a difference $\Delta = \delta - \delta^+$, where δ and δ^+ measure the interaction energy between the two substituents in the base 1,2- C_6H_4XY and determine the interference energy in the conjugated acid obtained by the protonation at a site para to X and meta to Y , respectively. They are given by following equations:



and



If $\Delta = \delta - \delta^+$ is small, then the proton affinity of disubstituted benzenes is obtained to a good accuracy by the *PA* of benzene corrected by a sum of increments appearing due to substituents *X* and *Y*. Consequently, if Δ can be abandoned, each substituent influences the proton affinity as if the other was absent thus leading to the independent substituent approach (ISA). To put it in another way, deviation from additivity, Δ reflects the influence of the collective effect on the proton affinity in disubstituted benzene. Some typical results obtained by utilizing formula (12) and the MP2(I) model are presented in Table 4. Survey of the given data shows that the simple additivity formula works very well. It is important to point out that both the MP2(I) model and the additivity rule of thumb are accurate enough to distinguish between different experimental *PA* ladders. In particular, it appears that the more recent experimental data of the proton affinity in 1,2- and 1,4- difluorobenzenes of Szulejko and McMahon [11] are more accurate than earlier experimental estimates. Finally, it is noteworthy that the average absolute deviation $\Delta(av)_{abs}$ is 1.0 kcal/mol.

2.4.3. Polysubstituted Benzenes

The generalisation of eqn.(12) is straightforward. The *PA* of a multiply substituted benzene is given by:

$$PA(\text{subst. benzene}) = PA(\text{benzene}) + \sum_N I^+(N)_{\alpha(N)} \quad (15)$$

where the summation runs over all substituents *N*, whilst $\alpha(N)$ denotes the position of the substituent relative to the protonation site ($\alpha = \text{o, m, p, i}$). Some representative results are presented in Table 5. Performance of the additivity rule is very good since a difference between *PA*(add.) values and MP2(I) results or the measured data rarely exceeds 1-2 kcal/mol implying that they are within the experimental errors. Our calculations carried out so far show that the average absolute deviation of the additivity formula from the MP2(I) estimates $\Delta(av)_{abs}$ is as low as 1.2 kcal/mol. It should be strongly pointed out that the additivity rule of thumb may be very useful in checking the experimental and/or theoretical data. In this connection it is interesting to mention that the additivity *PA* values of heavily fluorinated benzenes are in very good accordance with recent experimental findings thus providing additional piece of evidence that parts of earlier proton affinity ladders should be revised.

Table 4

Proton affinities of some disubstituted benzenes obtained by the MP2(I) model ($PA[MP2(I)]$) and by the additivity rule ($PA(add.)$). Deviations from the additivity are given by Δ (in $kcal/mol$)

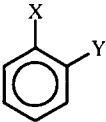
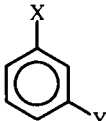
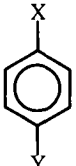
X	Y	$PA[MP2(I)]^a$	$PA(add)$	Δ	Exptl.
o-disubstituted benzenes					
$(CH_3)_m$	$(CH_3)_o$	189.0	189.2	-0.2	
$(CH_3)_p$	$(CH_3)_m$	189.7	193.6	-3.9	193.3 ^b
$(OH)_m$	$(OH)_o$	193.5	193.0	0.5	
$(OH)_p$	$(OH)_m$	197.5	195.5	2.0	
$(F)_m$	$(F)_o$	172.2	172.0	0.2	
$(F)_p$	$(F)_m$	175.0	174.1	0.9	181.8 ^b ; 175.7 ^c
$(OH)_m$	$(CH_3)_o$	186.5	186.2	0.3	
$(OH)_p$	$(CH_3)_m$	198.2	198.5	-0.3	
$(OH)_m$	$(CH_3)_p$	187.4	187.3	0.1	
$(OH)_o$	$(CH_3)_m$	194.4	196.0	-1.6	
m-disubstituted benzenes					
$(CH_3)_o$	$(CH_3)_o$	192.1	192.5	-0.4	
$(CH_3)_p$	$(CH_3)_o$	193.3	193.6	-0.3	195.9 ^b ; 188.1 ^d
$(CH_3)_m$	$(CH_3)_m$	186.0	185.9	0.1	
$(F)_o$	$(F)_o$	178.3	178.9	-0.6	
$(F)_p$	$(F)_o$	181.0	181.1	-0.1	181.5 ^b ; 181.6 ^d
$(F)_m$	$(F)_m$	165.6	165.0	0.6	
$(OH)_o$	$(CH_3)_o$	198.3	199.3	-1.0	
$(OH)_p$	$(CH_3)_o$	201.2	201.8	0.6	
$(OH)_m$	$(CH_3)_m$	183.3	182.9	0.4	
$(OH)_o$	$(CH_3)_p$	199.7	200.4	-0.7	

Table 4

Proton affinities of some disubstituted benzenes obtained by the MP2(I) model ($PA[MP2(I)]$) and by the additivity rule ($PA(\text{add.})$). Deviations from the additivity are given by Δ (in $kcal/mol$)

X	Y	$PA[MP2(I)]^a$	$PA(\text{add.})$	Δ	Exptl.
					
p-disubstituted benzenes					
$(CH_3)_m$	$(CH_3)_o$	189.1	189.2	-0.1	192.0 ^b ; 191.8 ^c
$(OH)_o$	$(OH)_m$	193.1	193.0	0.1	
$(F)_m$	$(F)_o$	172.4	172.0	0.4	181.2 ^b ; 171.5 ^c
$(OH)_o$	$(CH_3)_m$	185.4	186.2	-0.8	
$(OH)_m$	$(CH_3)_o$	195.6	196.0	-0.4	

^a Theoretical results are taken from [19].

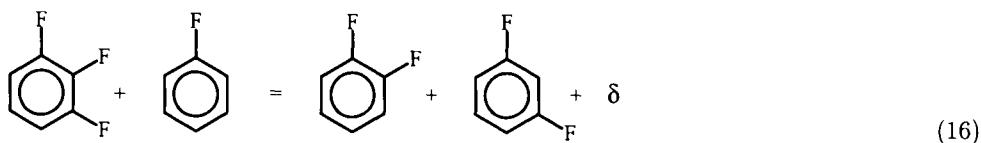
^b Ref. [33].

^c Ref. [11].

^d Ref. [37].

2.4.4. The Ipso protonation

The ipso protonation deserves a special scrutiny as discussed earlier. Here we show that the additivity rule is operative for the ipso protonation too, if proper reference level is found. Let us consider multiply substituted fluorobenzenes. They are schematically depicted in Fig.7. It is obvious that the proton affinity of benzene cannot serve as a gauge value for the ipso protonation. Instead, we shall employ once again homodesmotic reactions and proceed as follows. Protonation at position 1 of 1,2,3-trifluorobenzene will provide an illuminating example in this respect. The corresponding homodesmotic reactions read:



and

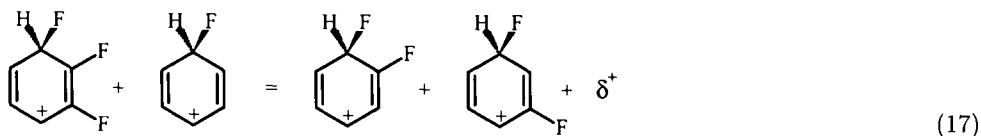


Table 5

Proton affinities of some polysubstituted benzenes obtained by the MP(I) model ($PA[MP2(I)]$) and by the additivity rule ($PA(\text{add})$). Deviations from the additivity are given by Δ in kcal/mol

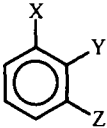
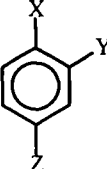
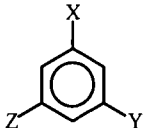
X	Y	Z	$PA[MP2(I)]^a$	$PA(\text{add})$	Δ	Exptl.
1,2,3-tri-substituted benzenes						
$(CH_3)_p$	$(CH_3)_m$	$(CH_3)_o$	195.1	196.6	-1.5	
$(CH_3)_m$	$(CH_3)_p$	$(CH_3)_m$	192.1	192.1	-1.2	
$(F)_p$	$(F)_m$	$(F)_o$	174.0	173.7	0.3	173.0 ^b
$(F)_m$	$(F)_p$	$(F)_m$	186.6	166.8	1.8	
$(CN)_p$	$(CH_3)_m$	$(F)_o$	170.5	169.2	1.3	
$(CN)_m$	$(CH_3)_p$	$(F)_m$	165.9	164.0	1.5	
$(CN)_o$	$(CH_3)_m$	$(F)_p$	173.0	171.5	1.5	
$(OH)_p$	$(CH_3)_m$	$(CH_3)_o$	203.0	204.8	-1.8	
$(OH)_m$	$(CH_3)_p$	$(CH_3)_o$	190.0	190.3	-0.3	
$(OH)_o$	$(CH_3)_m$	$(CH_3)_p$	200.1	202.3	-2.2	
1,2,4-tri-substituted benzenes						
$(F)_m$	$(F)_o$	$(F)_o$	171.1	171.5	-0.4	
$(F)_m$	$(F)_p$	$(F)_o$	175.0	173.7	1.3	174.5 ^b ; 183.4 ^c
$(F)_o$	$(F)_m$	$(F)_m$	165.5	164.6	0.9	
$(F)_m$	$(CH_3)_o$	$(CH_3)_o$	185.2	185.1	0.1	
$(F)_m$	$(CH_3)_p$	$(CH_3)_o$	186.9	186.2	0.7	
$(F)_o$	$(CH_3)_m$	$(CH_3)_m$	186.1	185.4	0.7	
$(F)_m$	$(OH)_o$	$(CH_3)_o$	191.2	191.9	-0.7	
$(F)_m$	$(OH)_p$	$(CH_3)_o$	195.6	194.4	1.2	
$(F)_o$	$(OH)_m$	$(CH_3)_m$	181.3	182.4	-1.1	
$(CH_3)_m$	$(F)_o$	$(CN)_o$	170.5	169.3	1.2	
$(CH_3)_m$	$(F)_p$	$(CN)_o$	172.4	171.5	0.9	
$(CH_3)_o$	$(F)_m$	$(CN)_m$	164.8	162.9	1.9	

Table 5

Proton affinities of some polysubstituted benzenes obtained by the MP(I) model ($PA[MP2(I)]$) and by the additivity rule ($PA(add)$). Deviations from the additivity are given by Δ in $kcal/mol$

X	Y	Z	$PA[MP2(I)]^a$	$PA(add)$	Δ	Exptl.
1,3,5-tri-substituted benzenes						
$(F)_p$	$(F)_o$	$(F)_o$	180.0	180.6	-0.6	178.4 ^b ;181.0 ^c
$(F)_p$	$(CH_3)_o$	$(CH_3)_o$	164.2	194.2	0.1	
1,2,3,4-tetrafluorobenzene (5-protonated)			168.2	166.2	1.9	
1,2,4,5-tetrafluorobenzene (3-protonated)			164.2	164.1	0.1	
1,2,3,4,5-pentafluorobenzene (6-protonated)			166.8	165.8	1.0	164.3 ^b ;179.9 ^c

^a Theoretical results are taken from [19].

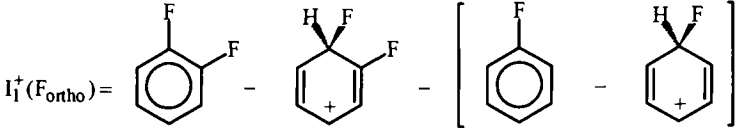
^b Ref. [11].

^c Ref. [33].

Taking a difference (16)-(17) and rearranging the resulting terms one obtains:

$$PA[1,2,3\text{-trifluorobenzene}]_1 = PA[1\text{-fluorobenzene}]_1 + I_1^+(F_{ortho}) + I_1^+(F_{meta}) + (\delta - \delta^+) \quad (18)$$

where the increments are defined by :

$$I_1^+(F_{ortho}) = \text{Structure 1} - \text{Structure 2} - \left[\text{Structure 3} - \text{Structure 4} \right] \quad (19)$$


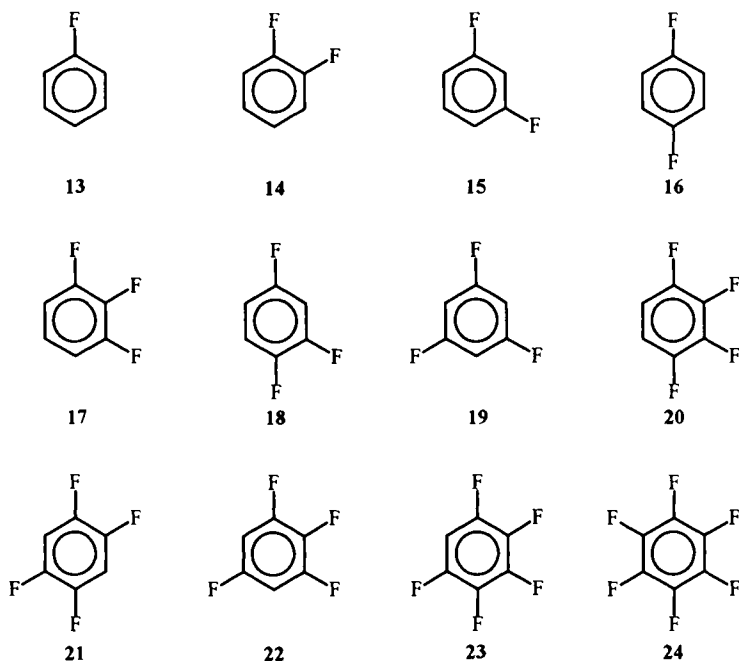


Figure 7. Schematic representation of fluorobenzenes used in studying the additivity property of the proton affinity for the ipso attack.

and

$$I_1^+(F_{meta}) = \text{Structure 1} - \text{Structure 2} - \left[\text{Structure 3} - \text{Structure 4} \right] \quad (20)$$

Here, the ipso protonation at position 1 of 1-fluorobenzene is denoted by the subscript 1, whereas ortho and meta signify positions of other substituent(s) relative to the protonation site. The difference $\Delta = (\delta - \delta^+)$ determines deviations from strict additivity. It will appear that δ and δ^+ cancel to a large extent, although their particular values are sometimes not negligible. Relation (18) is easily generalized to encompass polyfluorinated benzenes:

$$PA[pfb]_1 = (PA)_1 + n_o I_1^+(F_{ortho}) + n_m I_1^+(F_{meta}) + n_p I_1^+(F_{para}) \quad (21)$$

where *pfb* stands for a polyfluorinated benzene and $(PA)_1$ is a short hand notation for $PA(1\text{-fluorobenzene})$. Integers n_o , n_m and n_p assume values: $n_o, n_m = 0, 1, 2$ and $n_p = 0, 1$

depending on the number of substituents. Employing our adopted MP2(I) model one obtains the following increments: $I_1^+(F_{ortho}) = 2.0$, $I_1^+(F_{meta}) = -7.6$ and $I_1^+(F_{para}) = 3.0$ (in kcal/mol) [38]. It appears that ortho and para fluorine substitutions stabilize the ipso protonated form of 1-fluorobenzene, whereas meta fluorination leads to its destabilization. Ipso proton affinities are given in Table 6. The average absolute deviation $\Delta(av)_{abs} = |PA(MP2) - PA(add)|$ is as small as 0.8 kcal/mol, which reflects excellent performance of the simple additivity rule.

Table 6

Ipso proton affinities of fluorinated benzenes (17)-(24) as estimated by the MP2(I) model and by the additivity formula (21) in kcal/mol. ^a

Molecule	$PA[MP2(I)]^b$	$PA(add)$	Δ	Exptl.
17 (i,o,o)	164.0	164.4	-0.4	-
17 (i,o,m)	155.0	154.8	0.2	-
18 (i,o,p)	165.6	165.4	0.2	-
18 (i,o,m)	155.3	154.8	0.5	-
18 (i,m,p)	156.6	155.8	0.8	-
19 i	145.7	145.2	0.5	-
20 (i,o,m,p)	158.2	157.8	0.4	-
20 (i,o,o,m)	156.6	156.8	-0.2	-
21 i	159.5	157.8	1.7	-
22 (i,o,m,m)	148.1	147.2	0.9	-
22 (i,o,o,p)	167.1	167.4	-0.3	-
22 (i,m,m,p)	149.9	148.2	1.7	-
23 (i,o,m,m,p)	152.5	150.2	2.3	-
23 (i,o,o,m,p)	160.4	159.8	0.6	-
23 (i,o,o,m,m)	149.6	149.2	0.4	-
24 i	153.8	152.2	1.6	153.8 ^c

^a The first index within parentheses denotes ipso protonation, whereas remaining indices determine positions of the remaining F atoms relative to the protonated C atom.

^b Theoretical results are taken from Ref. [38].

^c Ref. [11].

As a rule, deviations Δ are very small with very few exceptions (Table 6). However, some interference energies δ and δ^+ might be quite appreciable. For instance, interactions between F atoms in system **14**, **17** and **18** are 4.7, 4.6 and 4.6 (in kcal/mol) respectively, being quite comparable to those in the protonated forms **14i**, **17(i,o,m)** and **18(i,m,p)** which given in the same order read 2.6, 4.4 and 3.8 (in kcal/mol). Consequently, they practically cancel out in the process of calculating the ipso proton affinity by the additivity rule of thumb yielding an "error" of ~ 0.8 kcal/mol. We note in passing that the interference energy is substantial whenever F atoms assume vicinal positions implying that the destabilization interaction depends on proximity of fluorines as intuitively expected.

It is clear that reasonable estimates of interference energies δ and δ^+ is a prerequisite for establishing the additivity formulas. For instance, they should not be sensitive on the basis set within the adopted theoretical framework. It is plausible to assume that δ and δ^+ are quite insensitive on the finer details of the employed molecular wavefunctions in view of the homodesmicity of the twins eqns. (16) and (17). In order to illustrate this intuitive conjecture, we have calculated δ and δ^+ by employing a more flexible (but also a more expensive) basis set within the $MP2(fc)/6-311G^{**}/HF/6-31G^*$ model. The corresponding δ and δ^+ values for **17** and **17(i;m,p)** are 4.5 and 4.2 kcal/mol, respectively, thus being practically equal to results obtained by the MP2(I) model given above. It should also be mentioned that δ and δ^+ are not influenced by ZPEs.

Finally, a word on the significance of the additivity formula is in place here. It enables a quick estimates of the ipso PAs in heavily substituted aromatics, avoiding time consuming or even non-feasible full calculations. We tacitly assume here that a very practical additivity rule of thumb will work successfully for other substituents too. Some preliminary calculations show that this is indeed the case. A distinct advantage of the additivity formula is that it can provide information on the ipso proton affinity for a number of situations which are not amenable to experimental investigation. It is gratifying that our estimated PA value of perfluorobenzene is in harmony with the experimental finding [11]. It is also intellectually pleasing that the additivity offers at the same time a transparent rationalization of the variation in PA in multiply substituted aromatics.

2.4.5. Limitations of the MP2(I) Model – The Aniline Story

Our extensive studies of the proton affinity of substituted benzenes involved aniline, too. It turned out that the MP2(I) model failed in reproducing the experimental value of aniline of 209.3 kcal/mol [29]. This has led to a careful examination of aminoalkanes in order to select a proper theoretical model, which would be still practical enough to be used in larger aromatics. It appeared that such a model was the $MP2(fc)/6-311+G^{**}/HF/6-31G^*+ZPE(HF/6-31G^{**})$ approach [39]. This model will be referred to as MP2(II). A use of a more flexible basis set involving diffuse function in the MP2 single point calculation is plausible in view of the fact that the lone pair electrons in sp^3 nitrogen are relatively loosely bound. Concomitantly, they are spatially more diffuse as compared to the lone pairs in oxygen or fluorine. Some typical proton affinities of aminoalkanes obtained by the MP2(II) model are compared with the experimental data in Table 7. Agreement with the measured data is very good, NF_3 being an exception which calls for reexamination of this experimental value since $MP2(II)$ model cannot fail as much as 16.5 kcal/mol. This conclusion is supported by additional calculations performed at G2 and G2(MP2) levels of theory. These methods give for the proton affinity of NF_3 values 130.2 and 130.8 kcal/mol, respectively [39], which are very in good accordance with the estimate of the considerably simpler MP2(II) model (131.5 kcal/mol). As an important outcome of the MP(II) study of the PA of aminoalkanes is a strong inverse linear dependence of the proton affinity on the s-character of the nitrogen lone pair. It appears also that the relaxation energy plays a significant role as evidenced by PAs in the series NR_nH_{3-n} with $R = CH_3$ and C_2H_5 . In the latter case PAs were higher by 3–7 kcal/mol presumably because C_2H_5 is a larger electron density reservoir, which is capable to screen the proton positive charge in a more efficient way [39].

Table 7

Proton affinities of ammonia and some of its alkyl and fluoro derivatives (in kcal/mol)

Molecule	(<i>PA</i> [MP2(II)])	Exptl.	Molecule	(<i>PA</i> [MP2(II)])	Exptl.
NH_3	204.4	205.0 ^b	$(CF_3)(CH_2)NH_2$	201.5	203.3 ^b
CH_3NH_2	214.9	214.1 ^b	$(CF_3)(CH_3)_2N$	200.0	195.0 ^b
$(CH_3)_2NH$	220.8	220.5 ^b	FNH_2	181.6	181±5 ^c
$(CH_3)_3N$	226.4	224.3 ^b	F_2NH	157.6	163±5 ^c
$C_2H_5NH_2$	217.5	217.1 ^b	NF_3	131.5	148 ^d
$(C_2H_5)_2NH$	226.6	225.1 ^b	$(CH_3)_2FN$	203.5	
$(C_2H_5)_3N$	223.1	231.2 ^b	$(CH_3)F_2N$	172.5	

^a Theoretical results are taken from [39].^b Ref. [40]^c Ref. [41]^d Ref. [41]

Having established that the MP2(II) model described *PAs* in aminoalkanes rather satisfactory, we applied it to aniline and pyridine. It turned out that the calculated *PAs* of these two important molecules were 209.5 and 219.9 kcal/mol, respectively, thus being in very good agreement with experimental values of 209.5 and 220.8 kcal/mol as given by Aue and Bowers [40]. It was also conclusively shown that aniline was a nitrogen base thus resolving a long standing controversy which gave rise to a dispute for almost 20 years [39].

2.4.6. Proton Affinities of Larger Aromatics–Naphthalenes

Investigations performed on polysubstituted benzenes have shown that the independent substituent approximation (ISA) provided an extremely simple and useful rule of thumb enabling quick estimates of proton affinity values in heavily substituted benzenes with surprisingly good accuracy. It is desirable to check the generality of the additivity rule in larger aromatics. One can assume on intuitive grounds that the additivity should work in large aromatics even better since perturbations exerted by the substituents on more sizeable structural framework are smaller. We shall dwell here on naphthalene substituted by *F* and *CN* substituents, because they are free of conformational problems thus being perfectly suited to test the additivity of the *PA*. Some of the studied systems are schematically depicted in Fig.8. The additivity formula is *mutatis mutandis* the same as eqn.(22):

$$PA(\text{subst.naphthalene}) = PA(\text{naphthalene}) + \sum_X I^+(X)_{H(n)} \quad (22)$$

where the summation is extended over all substituents *X* and *n* denotes positions of the proton attack. The increment is given by $I^+(X)_{H(n)}$. It measures the change in the proton affinity of naphthalene due to substituent *X* attached to particular positions around the molecular perimeter [43]. The applied MP2(I) model gives the *PA* value of naphthalene of 194.9 kcal/mol, which is in excellent agreement with the experimental results [43]. It is related to protonation at position 1. In order to determine the increments, it is important to know that the *PA* value for protonation at C(2) carbon atom is 190.5

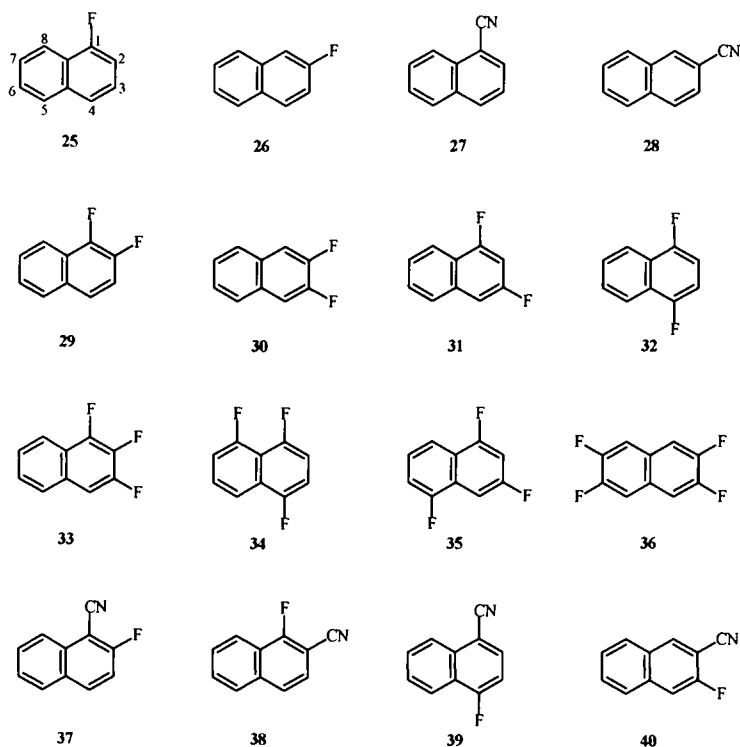


Figure 8. Schematic representation and numbering of atoms of substituted naphthalenes

kcal/mol. Increments caused by *F* and *CN* substituents are presented in Table 8, where position of the proton attack are given in the first row. It appears that fluorine activates positions 2 and 4 in **25** and 1 in **26**, which is compatible with the dominant VB resonance structures [43]. The computed proton affinities obtained by the MP2(I) model and the additivity rule are given in Table 9 together with the deviations Δ and the interference energies δ and δ^+ .

It is interesting to point out that the lowest proton affinity in polyfluorinated naphthalenes is found for ipso protonation (viz. systems **33**, **35** and **36**). It is a consequence of the out-of-plane shift of fluorine and the accompanying ring puckering. However, this is at the same time a manifestation of the π -electron fluoro effect put forward by Liebman *et al.* [45]. It is very well known that multiply fluorinated compounds possess considerably stabilized σ -MOs if the systems are planar, the π -manifold being almost unaffected [46]. However, in nonplanar systems all MOs at the carbon skeleton are significantly stabilized [45,46] which is exactly the case for the ipso protonation. Now, it can be easily shown

Table 8

Proton affinities of molecules **25-28** and the corresponding increments as obtained by the MP2(I) model (in kcal/mol)^a

Mol.	Entity	1	2	3	4	5	6	7	8
25	<i>PA</i>	180.7	192.5	185.8	197.2	191.4	186.2	187.4	191.7
	$I^+(F_1)_{H(n)}$	-14.1	2.0	-4.7	2.4	-3.4	-4.3	-3.1	-3.1
26	<i>PA</i>	195.9	173.7	187.9	187.9	190.7	189.6	186.1	193.2
	$I^+(F_2)_{H(n)}$	1.1	-16.8	-2.6	-6.9	-4.1	-0.9	-4.4	-1.6
27	<i>PA</i>	-	178.5	178.9	182.6	184.8	178.8	180.2	184.8
	$I^+(CN_1)_{H(n)}$	-	12.0	-11.6	-12.2	-10.0	-11.7	-10.3	-10.0
28	<i>PA</i>	183.3	-	180.1	180.8	183.5	180.0	178.5	184.7
	$I^+(CN_2)_{H(n)}$	-11.5	-	-10.4	-14.0	-11.3	-10.5	-12.0	-10.1

^a Theoretical data are taken from Refs.[43] and [44].

that the following relationship holds:

$$PA(M) = D_e(M - H) + IP(H) - IP(M)_1 \quad (23)$$

where $D_e(M - H)$ is the bond dissociation energy of $M - H$ bond. M stands here for a molecule to be protonated, its first ionization potential is denoted by $IP(M)_1$, whereas the ionization potential of the hydrogen atom is given by $IP(H) = 13.6 \text{ eV}$. Since $IP(M)_1$ of the nonplanar fluorinated molecular system increases it implies that the corresponding ipso $PA(M)$ value decreases.

Perusal of the presented data in Table 9 shows that the additivity formula based on the independent substituent approach performs very well in describing PA s of substituted naphthalenes. If the ipso protonation is excluded, then the average absolute deviations $|\Delta|_{av} = |\delta - \delta^+|_{av}$ is 0.5 kcal/mol, which is remarkable indeed. The origin of this very high accuracy of the extremely simple and intuitively appealing ISA model lies in a fact that the interference energies δ and δ^+ in the initial base and the final state conjugated acid, respectively, are very similar. They could be sometimes as large as 9 kcal/mol like in compounds **33** and **36**, but their difference is very small as a rule (Table 9). It is also worth mentioning that δ and δ^+ are usually rather small in most cases. The present results are encouraging and probably will open up a new avenue of research of properties of heavily substituted large aromatic compounds in general and their proton affinities in particular.

3. Miscellaneous Applications of the Additivity Rule

The additivity rule governing the proton affinity of various multiply substituted aromatics may be useful in rationalizing many closely related properties of this family of compounds. For instance, Brown and Brady [47] determined the relative basicities of a number of aromatic compounds including methylbenzenes. Unfortunately, experimental PA values related to methylbenzenes are scarce. Hence, proton affinities obtained by the additivity rule may prove very useful here in predicting the missing experimental data

Table 9

Proton affinities of compounds **25-40** as obtained by the MP2(I) model and the additivity formula (22) (in kcal/mol)

Molecule	Entity	1	2	3	4	5	6	7	8
29	<i>PA</i>	184.4	179.1	182.7	191.3	187.5	185.5	182.9	190.4
	<i>PA_{ad}</i>	181.8	175.7	183.2	190.3	187.3	185.3	183.0	190.1
	Δ	2.6	3.4	-0.5	1.0	0.2	0.2	-0.1	0.3
	δ	4.8	4.8	4.8	4.8	4.8	4.8	4.8	4.8
	$-\delta^+$	-2.2	-1.4	-5.3	-3.8	-4.6	-4.6	-4.9	-4.5
30	<i>PA</i>	188.6	173.0	-	-	189.1	185.4	-	-
	<i>PA_{ad}</i>	189.0	171.1	-	-	189.1	185.2	-	-
	Δ	-0.4	1.9	-	-	0.0	0.2	-	-
	δ	4.3	4.3	-	-	4.3	4.3	-	-
	$-\delta^+$	-4.7	-2.4	-	-	-4.3	-4.1	-	-
31	<i>PA</i>	173.8	189.3	169.0	197.9	189.8	181.9	186.7	187.5
	<i>PA_{ad}</i>	173.8	189.9	169.0	198.3	189.8	181.8	186.5	187.6
	Δ	0.0	-0.6	0.0	-0.4	0.0	0.1	0.2	-0.1
	δ	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
	$-\delta^+$	-0.6	-1.2	-0.6	-1.0	-0.6	-0.5	-0.4	-0.7
32	<i>PA</i>	184.0	188.0	-	-	188.1	182.9	-	-
	<i>PA_{ad}</i>	183.1	187.8	-	-	188.3	183.1	-	-
	Δ	0.8	0.2	-	-	-0.2	-0.2	-	-
	δ	0.8	0.8	-	-	0.8	0.8	-	-
	$-\delta^+$	0.0	-0.6	-	-	-1.0	-1.0	-	-
33	<i>PA</i>	177.1	178.1	167.8	191.3	185.8	181.4	182.2	186.4
	<i>PA_{ad}</i>	172.7	173.1	166.4	191.4	185.7	180.9	182.1	186.0
	Δ	2.2	5.0	1.4	-0.1	0.1	0.5	0.1	0.4
	δ	9.7	9.7	9.7	9.7	9.7	9.7	9.7	9.7
	$-\delta^+$	-7.5	-4.7	-8.3	-9.8	-9.6	-9.2	-9.6	-9.3
34	<i>PA</i>	181.4	185.8	183.6	181.7	191.3	178.0	185.8	174.7
	<i>PA_{ad}</i>	180.0	184.7	184.4	179.7	190.7	178.4	185.1	174.9
	Δ	1.3	1.0	-0.1	2.0	0.6	-0.4	0.7	0.5
	δ	4.3	4.3	4.3	4.3	4.3	4.3	4.3	4.3
	$-\delta^+$	-3.0	-3.3	-4.4	-1.3	-3.8	-4.9	-3.7	-4.2
35	<i>PA</i>	170.4	185.0	165.8	194.8	176.3	184.0	181.8	189.8
	<i>PA_{ad}</i>	170.4	185.6	165.9	195.2	175.7	183.8	181.8	190.0

Table 9
Proton affinities of compounds **25-40** as obtained by the MP2(I) model and the additivity formula (22) (in kcal/mol)

Molecule	Entity	1	2	3	4	5	6	7	8
	Δ	0.0	-0.6	-0.1	-0.4	0.6	0.2	0.0	-0.1
	δ	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	$-\delta^+$	-0.8	-1.5	-0.8	-1.1	-0.1	-0.5	-0.7	-0.8
36	<i>PA</i>	183.1	168.5	-	-	-	-	-	-
	<i>PA_{ad}</i>	183.1	165.8	-	-	-	-	-	-
	Δ	-0.2	2.7	-	-	-	-	-	-
	δ	9.3	9.3	-	-	-	-	-	-
	$-\delta^+$	-9.5	-6.6	-	-	-	-	-	-
37	<i>PA</i>	-	-	177.0	176.6	181.3	178.4	176.5	183.4
	<i>PA_{ad}</i>	-	-	176.3	175.7	180.7	177.9	175.8	183.2
	Δ	-	-	0.7	0.9	0.6	0.5	0.7	0.2
	δ	-	-	1.0	1.0	1.0	1.0	1.0	1.0
	$-\delta^+$	-	-	-0.3	-0.1	-0.4	-0.5	-0.3	-0.8
38	<i>PA</i>	-	-	176.2	183.5	180.3	176.3	175.7	181.9
	<i>PA_{ad}</i>	-	-	175.4	183.2	180.1	175.4	175.4	181.6
	Δ	-	-	0.8	0.3	0.2	0.6	0.3	0.3
	δ	-	-	1.0	1.0	1.0	1.0	1.0	1.0
	$-\delta^+$	-	-	-0.2	-0.7	-0.8	-0.4	-0.7	-0.7
39	<i>PA</i>	-	173.8	181.1	-	181.7	175.7	176.0	181.5
	<i>PA_{ad}</i>	-	173.8	180.9	-	181.7	175.7	175.9	181.4
	Δ	-	0.0	0.2	-	0.0	0.0	0.1	0.1
	δ	-	-0.3	-0.3	-	-0.3	-0.3	-0.3	-0.3
	$-\delta^+$	-	0.3	0.5	-	0.3	0.4	0.4	0.4
40	<i>PA</i>	177.9	-	-	182.7	182.3	176.5	178.2	181.3
	<i>PA_{ad}</i>	176.4	-	-	181.9	181.9	175.6	177.6	180.6
	Δ	1.5	-	-	0.8	0.4	0.9	0.5	0.7
	δ	1.5	-	-	1.5	1.5	1.5	1.5	1.5
	$-\delta^+$	0.0	-	-	-0.7	-1.1	-0.6	-1.0	-0.8

and in correlating the relative basicities measured by Brown and Brady. Indeed, there is a very good linear correlation between these relative basicities RB and the PA values offered by the additivity rule [23]. The least square fit method gives the following relation:

$$RB = 0.050 \cdot PA - 8.4 \text{ kcal/mol} \quad (24)$$

with the correlation coefficient $R = 0.985$. It follows as a corollary that the additivity works very well for polymethylbenzenes and that solvent effects do not change the order of their proton affinities or basicities. Additionally, the additivity results supply theoretical relative basicities for durene and pentamethylbenzene, where experimental data are not available. It is interesting to mention that for hexamethylbenzene $PA(\text{add}) = 207.3 \text{ kcal/mol}$ agrees well with the measured value of 206.2 kcal/mol [7]. Furthermore, the predicted relative basicities are the same for penta- and hexamethylbenzene, since $I^+(\text{CH}_3)_i = 0$. An experimental confirmation of this prediction would be desirable.

Another useful application of the proton affinity is related to rationalization of the electrophilic reactivity of aromatics. Several illustrative examples will be presented here. The well-known empirical para directive ability of halogens [48] is easily rationalized by the PA increments. For instance, the relative yields of 4- and 2-protonated 1-fluoro-3,5-dimethylbenzene are 84 % and 16 %, respectively [48]. The corresponding sums of the PA increments are 14.3 and 13.2 kcal/mol, respectively. The difference can be ascribed to the activating (para) and deactivating (ortho) action of the F atom, since the methyl groups themselves favor protonation in position 2. This example shows rather nicely how the additivity concept can describe the origin of the directional properties of the electrophilic reactivity of the substituted benzene ring at least at the qualitative level. Further, the relative rates of protodetritiation in trifluoroacetic acid for various positions in toluene and o-, m- and p-xylenes [49] are in qualitative agreement with the increase in PA for the corresponding ring carbon atoms given by the additivity rule. It is also well known that o-cresol is protonated mainly ortho and para to the OH group and not to the CH_3 group. This is perfectly clear since substitutions ortho and para to the methyl group would imply that the OH group does not enter into play, because $I^+(\text{OH})_m = 0.0$, in contrast to other positions, where the OH fragment contributes either 13.1 ($I^+(\text{OH})_o$) or 15.6 kcal/mol ($I^+(\text{OH})_p$). The cooperative action of OH and CH_3 groups is reflected by the fact to o-cresol reacts with bromine about five times more rapidly than does phenol [50]. Finally, the electrophilic regioselectivity of benzenes fused to small rings is discussed elsewhere [9,51,52] and needs not to be repeated here.

4. Conclusion

We have conclusively shown that the ISA model was very useful in predicting and interpreting proton affinities in polysubstituted benzenes and naphthalenes. A compelling evidence is provided which documents that the PA values obtained by the additivity rule of thumb are in very good accordance with available experimental data and/or the accurate theoretical results offered by the *ab initio* models at the MP2 level of sophistication. Analogous formulas should work in larger aromatic systems too. If there are some

exceptions, which do not satisfy the additivity rule, they should be treated separately. A typical case is provided by the ipso protonation of polysubstituted benzenes. The additivity was restored by selecting a proper reference level. Generally, if the additivity fails, this might provide a clue for some interesting interactions between substituents thus yielding a deep insight into the intramolecular interactions in aromatics. Preferential sites of the proton attack are determined by an interplay of several contributions to the proton affinity. The increased electron density placed at a particular atom increases the favorable Coulomb attraction (the ground state effect). A synergistic interaction between negative atomic charges induced by resonance effect upon substitution and positive atomic charges produced by protonation and concomitant formation of the carbon sp^3 center within the aromatic moiety represents a combination of the ground (initial base) and final state (conjugated acid) effects. The aromaticity defect caused by the carbon sp^3 center might lead to a preferential heteroatom attack on the substituent together with deactivation of the aromatic ring due to its electron density depletion. A typical clear-cut final state effect is given by the reorganization of the electron density upon creation of the positive charge in the conjugated acid. These four modes of intramolecular interactions vary from one substituent to another. It is difficult to delineate them in a quantitative sense at present. Finally, the proton affinity estimated by the additivity rule proved useful in distinguishing the experimental values belonging to different *PA* ladders and in interpreting directional ability of substituents in determining the most susceptible site for an electrophilic attack.

Acknowledgement: This work has been supported by the Ministry of Science and Technology of Republic of Croatia through the program 009808. A part of this work has been performed during our visit to the Institute of Organic Chemistry of the Westfälische Wilhelms-Universität in Münster. M.E.M. and Z.B.M. would like to thank the Alexander von Humboldt-Stiftung and the Deutsche Forschungsanstalt für Luft und Raumfahrt e.v. in Bonn, respectively, for financial support.

REFERENCES

1. C.H. Bamford and C.F.H. Tipper (eds.), *Comprehensive Chemical Kinetics*, Vol.8, Proton Transfer, Elsevier, Amsterdam, 1977.
2. T.H. Lowry and K.S. Richardson, *Mechanics and Theory in Organic Chemistry*, Harper & Row, New York, NY, 1976.
3. R. Stewart, *The Proton: Applications to Organic Chemistry*, Academic Press, Orlando, FL, 1985.
4. E.M. Arnett, *Acc. Chem. Res.*, **7**(1973)404.
5. M. Meot-Ner (Mautner), *Acc. Chem. Res.*, **17**(1984)186.
6. F. Cacace, *Acc. Chem. Res.*, **21**(1988)215.
7. S.G. Lias, J.F. Liebman and R.D. Levin, *J. Phys. Chem. Ref. Data*, **13**(1984)695.
8. G.A. Harrison, *Chemical Ionization Mass Spectrometry*, Sixth Printing, CRC Press, Boca Raton, FL, 1989.
9. R. Taylor, *Electrophilic Aromatic Substitution*, J. Wiley, Chichester, 1990.

10. M. Meot-Ner (Mautner) and L.W. Sieck, *J. Am. Chem. Soc.*, **113**(1991)4448.
11. J.E. Szulejko and T.B. McMahon, *J. Am. Chem. Soc.*, **115**(1993)7839.
12. B.J. Smith and L. Radom, *J. Am. Chem. Soc.*, **115**(1993)4885.
13. W.J. Hehre, L. Radom, P.V.R. Schleyer and J.A. Pople, *Ab Initio Molecular Orbital Theory*, J. Wiley-Interscience, New York, NY, 1986.
14. B.J. Smith and L. Radom, *Chem. Phys. Lett.*, **231**(1994)345.
15. J.W. Ochterski, G.A. Peterson and K.B. Wiberg, *J. Am. Chem. Soc.*, **117** (1995) 11299
16. A. M. Schmiedekamp, I. A. Topol and C. J. Michejda, *Theor. Chim. Acta*, **92** (1995) 83.
17. A. K. Chandra and A. Goursot, *J. Phys. Chem.*, **100** (1996) 11596.
18. M. Eckert-Maksić, M. Klessinger and Z. B. Maksić, *Chem. Phys. Lett.*, **232** (1995) 472.
19. M. Eckert-Maksić, M. Klessinger and Z. B. Maksić, *J. Phys. Org. Chem.*, **8** (1995) 435; *Chem. Eur. J.*, **2** (1996) 155.
20. D. Kovaček, Z. B. Maksić and I. Novak, *J. Phys. Chem.*, in print.
21. K. Zhang, D. M. Zimmerman, A. Chung-Phillips and C. J. Cassady, *J. Am. Chem. Soc.*, **115** (1993) 10812.
22. K. Zhang, C. J. Cassady and A. Chung-Phillips, *J. Am. Chem. Soc.*, **116** (1994) 11512.
23. J. A. Pople, H. B. Schlegel, R. Krishnan, D. J. De Frees, J. S. Binkley, M. J. Frisch, R. W. Whitesides, R. F. Hout and W. J. Hehre, *Int J. Quant. Chem. Symp.* **15** (1981) 269.
24. J. M. Schulman and R. L. Disch, *Chem. Phys. Lett.*, **113** (1985) 291; M. R. Ibrahim and Z. A. Fataftah, *Chem. Phys. Lett.*, **125** (1986) 149.
25. GAUSSIAN 92, Revision B, M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzales, R. L. Martin, D. J. Fox, D. J. De Frees, J. Baker, J. J. P. Stewart and J. A. Pople, Gaussian, Pittsburg PA, 1992.
26. GAUSSIAN 94, Revision B.2, M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Peterson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkare, M. Challacombe, C. Y. Peng, P. A. Ayala, W. Chen, M. W. Wong, J. L. Anders, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. De Frees, J. J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzales and J. A. Pople, Gaussian, Pittsburg PA, 1995.
27. M. Eckert-Maksić, M. Hodošček, D. Kovaček, Z. B. Maksić and M. Primorac, *J. Mol. Struct. (Theochem)*, in press.
28. D. Kovaček, S. Kučina, Z. B. Maksić and M. Primorac, *El. J. Theoret. Chem.*, in press.
29. Y. K. Lau and P. Kebarle, *J. Am. Chem. Soc.*, **98** (1976) 7452.
30. D. M. Brouwer, E. L. Mackor and C. MacLean, in *Carbonium Ions*, Vol. 2, G. A. Olah and P. von R. Schleyer (eds.), J. Wiley, New York, NY, 1970.
31. D. J. De Frees, R. T. McIver Jr. and W. J. Hehre, *J. Am. Chem. Soc.*, **99** (1977)

- 3854.
32. Z. B. Maksić (ed.), *Theoretical Models of Chemical Bonding*, Vols. 1-4, Springer Verlag, Berlin-Heidelberg, 1990-91.
 33. S. G. Lias, J. E. Bartmess, J. F. Liebman, J. L. Holmes, R. D. Lewin and W. G. Mallard, *J. Phys. Chem. Ref. Data Suppl.*, **17** (1988).
 34. J. March, *Advanced Organic Chemistry*, 3rd. ed., J. Wiley, New York, NY, 1985, p. 238.
 35. P. George, M. Trachtman, C. W. Bock and A. M. Brett, *Tetrahedron* **32** (1976) 313; *J. Chem. Soc. Perkin Trans 2* (1976) 1222.
 36. R. S. Mason, M. T. Fernandez and K. R. Jennings, *J. Chem. Soc., Faraday Trans. 2*, **83** (1987) 89.
 37. R. Walder and J. L. Franklin, *Int. J. Mass. Spectr. Ion Phys.*, **36** (1980) 85.
 38. Z. B. Maksić, M. Eckert-Maksić and M. Klessinger, *Chem. Phys. Lett.*, **260** (1996) 572.
 39. C. Hillebrand, M. Klessinger, M. Eckert-Maksić and Z. B. Maksić, *J. Phys. Chem.*, **100** (1996) 9698.
 40. D.M. Aue and M.T. Bowers, in *Gas Phase Ion Chemistry*, M.T. Bowers (ed.), Vol.2, Academic Press, New York, NY, 1979, Chapter 9.
 41. R. Lascola, R. Withnall and L. Andrews, *J. Phys. Chem.*, **92**(1988)2145.
 42. C.E. Dorion and T.B. McMahon, *Inorg. Chem.*, **19**(1980)3037.
 43. D. Kovaček, Z.B. Maksić and I. Novak, *J. Phys. Chem.*, in print.
 44. M. Eckert-Maksić, M. Klessinger, I. Antol and Z.B. Maksić, *J. Phys. Org. Chem.*, in press.
 45. J.F. Liebman, P. Politzer and D.C. Rosen, in *Chemical Applications of Atomic and Molecular Electrostatic Potentials*, P. Politzer and D. Truhlar (eds.), Plenum Press, New York, 1981, p.295.
 46. C.R. Brundle, M.B. Robin, N.A. Kuebler and H. Basch, *J. Am. Chem. Soc.*, **94** (1972)1451.
 47. H.C. Brown and J.D. Brady, *J. Am. Chem. Soc.*, **74**(1952)3570.
 48. G.A. Olah and Y.K. Mo, in: *Carbonium Ions*, Vol.5, G.A. Olah and P.v.R. Schleyer (eds.), J. Wiley-Interscience, New York, 1976, p.2135.
 49. A. Streitwieser, Jr. and C.H. Heathcock, *Introduction to Organic Chemistry*, Ch.29, Mc Millan, New York, 1976.
 50. P.B.D. De la Mare, *Tetrahedron* **5**(1959)107.
 51. M. Eckert-Maksić, Z.B. Maksić and M. Klessinger, *Int. J. Quant. Chem.*, **49**(1994)383; *J. Chem. Soc., Perkin Trans. 2*(1994)285; M. Eckert-Maksić, W.M.F. Fabian, R. Janoschek and Z.B. Maksić, *J. Mol. Struct. (Theochem)*, **338**(1995)1243.
 52. M. Eckert-Maksić, M. Klessinger, D. Kovaček and Z.B. Maksić, *J. Phys. Org. Chem.*, **9**(1996)269.

This Page Intentionally Left Blank

Dipole Moments of Aromatic Heterocycles

Cyril Párkányi^a and Jean-Jacques Aaron^b

^aDepartment of Chemistry and Biochemistry, Florida Atlantic University, 777 Glades Road,
P.O. Box 3091, Boca Raton, FL 33431-0991, USA

^bInstitut de Topologie et de Dynamique des Systèmes, Université Paris 7 - Denis Diderot,
1, rue Guy de la Brosse, F-75005 Paris, France

1. INTRODUCTION

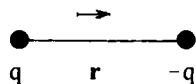
A dipole moment represents a direct measure of electron distribution in a molecule of known geometry. It is a physical constant which can be obtained experimentally and which can also be calculated.

In molecules containing atoms of different electronegativity, as is normally the case with heterocyclic compounds, the electrons are not shared equally by the respective atoms and this results in regions of high electron density and of low electron density. Because of this uneven distribution of electrons, a molecule, which as a whole is electroneutral, will possess a center of positive charge (positive end) and a center of negative charge (negative end). If these two centers do not coincide, the molecule has a permanent electric dipole moment.

Thus, the electric dipole moment μ is a vector and is defined as

$$\mu = q r \quad (1)$$

where q is the magnitude of charge and r is the vector distance between the centers of positive charge and negative charge.



The dimension of dipole moments is charge \times distance (esu-cm). The charge of the electron is 4.80×10^{-10} esu, i.e., it is of the order of 10^{-10} esu, and the distances are typically in Å ($1 \text{ \AA} = 10^{-8}$ cm), thus the order of magnitude is $(1 \times 10^{-10})(10^{-8})$ esu-cm. The unit used for dipole moments is the Debye unit, D, where $1 \text{ D} = 1 \times 10^{-18}$ esu-cm when the distances are expressed

in Å. This means that the dipole moment between a proton and an electron at a distance of 1 Å is 4.80 D. In this chapter, the positive direction of the dipole moment will be defined as the direction from the center of the positive charge towards the center of the negative charge.

The SI units for dipole moments, coulomb meters (C-m), are very rarely used (1 D = 3.336×10^{-30} C-m).

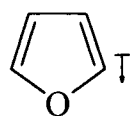
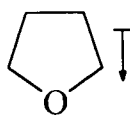
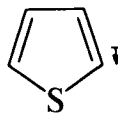
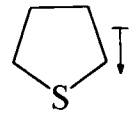
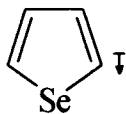
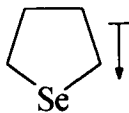
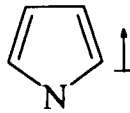
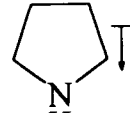
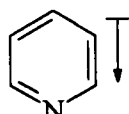
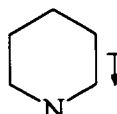
A number of excellent texts, monographs, chapters, and reviews devoted to dipole moments are available and contain a detailed discussion of the theory of dipole moments [1-23] which will not be presented here. However, most of the above references are not very recent. Some of them are specifically devoted to heterocyclic compounds [7,12,13] and several sets of tables of dipole moments are available [24-31], the most complete and valuable being the tables compiled by McClellan [29-31]. All the above references have served as an excellent source of information and a starting point for this chapter.

It seems worthwhile to point out here that, prior to the development and widespread applications of modern spectroscopic techniques (ir and nmr spectroscopy, mass spectrometry), dipole moments represented one of the most important sources of structural information about organic molecules. Although the importance of dipole moments in structural chemistry has diminished, they are still very useful and represent a way of obtaining different types of valuable data.

Examples of the various practical applications of dipole moments include, but are not limited to: differentiation between isomers (*cis* and *trans*, *o*, *m*, and *p*, tautomers, etc.), conformational analysis, studies of molecular geometry, supporting evidence for resonance hybrids, information about the polar character of molecules (important for solubility in different solvents and permeability through membranes), information about electrical effects of substituents (inductive, resonance), studies of hydrogen bonding, and studies of donor-acceptor interactions (e.g., charge transfer complexes). Practical cases describing the use of dipole moments for different types of structural studies mentioned above can be found in numerous publications mentioned in this chapter.

Most heterocyclic compounds possess an uneven distribution of charges resulting in a permanent dipole moment. Typical dipole moments for most organic molecules are in the range between 0 and 12 D but there are some compounds such as, e.g., polymethine dyes, which have dipole moments of 20 D or higher.

The role of heteroatoms in ground- and excited-state electronic distribution in saturated and aromatic heterocyclic compounds is easily demonstrated by a comparison of a number of heteroaromatic systems with their perhydro counterparts. In π -excessive heteroaromatic systems, because of their resonance structures, their dipole moments are less in the direction of the heteroatom than in the corresponding saturated heterocycles: furan (**1**, 0.71 D) vs. tetrahydrofuran (**2**, 1.68 D), thiophene (**3**, 0.52 D) vs. tetrahydrothiophene (**4**, 1.87 D), and selenophene (**5**, 0.40 D) vs. tetrahydroselenophene (**6**, 1.97 D). In the case of pyrrole (**7**, 1.80 D), the dipole moment is reversed and is actually higher than that of pyrrolidine (**8**, 1.57 D) due to the acidic nature of the pyrrole ring (the N-H bond). In contrast, the dipole moment of π -deficient pyridine (**9**, 2.22 D) is higher than that of piperidine (**10**, 1.17 D). In all these compounds, with the exception of pyrrole (**7**), the direction of the dipole moment is from the ring towards the heteroatom [32-34].

**1****2****3****4****5****6****7****8****9****10**

2. EXPERIMENTAL GROUND-STATE DIPOLE MOMENTS

Experimental dipole moments can be obtained in several different ways. The first and most widely used approach is based on the measurement of dielectric constants. The second group of methods utilizes microwave spectroscopy and molecular beams (the Stark effect method, the molecular beam method, the electric resonance method, Raman spectroscopy, etc.).

2. 1. Dielectric constant methods

Many dipole moments available in the literature were obtained from the measurements of dielectric constants in the vapor phase or in diluted solutions of polar compounds in nonpolar solvents [24-31]. All these methods are based on Debye's statistical theory [35].

The Debye equation relating the dipole moment, μ , to the total polarization of a substance, P , has the form

$$P = (4\pi N/3) [\alpha_0 + (\mu^2/3kT)] \quad (2)$$

where P is the molar polarization, N is Avogadro's number, α_0 is the polarizability of the molecule by distortion (atomic and electronic polarization), and $\mu^2/3kT$ is the polarization of the molecule by orientation. T is the absolute temperature and k is the Boltzmann constant.

Molar polarization, P , can be expressed as

$$P = [(\epsilon - 1)/(\epsilon + 2)] \cdot M/d \quad (3)$$

where ϵ is the dielectric constant, M is the molecular weight of the substance, and d its density.

Hence,

$$[(\epsilon - 1)/(\epsilon + 2)] \cdot M/d = (4\pi N/3) [\alpha_0 + (\mu^2/3kT)] \quad (4)$$

When the constants are expressed numerically and the total polarization (expression on the left-hand side) is plotted against T^{-1} , the dipole moment is obtained from the slope, b :

$$\mu = 0.01283 \sqrt{b} \quad (D) \quad (5)$$

A satisfactory range of temperatures (at least 100 K) is needed. In the above form, the Debye equation can be used for gas-phase determinations only and thus it is not applicable to most heterocyclic compounds.

However, a second approach developed by Debye [35] can be used for dilute solutions of a polar substance in nonpolar solvents. This approach assumes additivity of the properties of the components in solution, i.e., of the solute and of the solvent. The overall polarization has to be separated into contributions from the solute and the solvent, in the form of the respective specific or molar polarizations. The actual value of the dipole moment is again obtained from a linear plot. Throughout the years, a number of different techniques and equations for the determination of P were developed, the procedure introduced by Halverstadt and Kumler being one of the most successful ones [36].

The Halverstadt-Kumler method uses solute polarization at infinite dilution, $P_{2\infty}$, based on extrapolation of several P_2 values obtained for different concentrations of the solute. The equation has the form:

$$P_{2\infty} = 3\alpha M_2/[d_1(\epsilon_1 + 2)^2] + M_2[(\epsilon_1 - 1)/(\epsilon_1 + 2)][(1/d_1) + \beta] \quad (6)$$

where the indices 1 and 2 refer to the solvent and the solute, respectively, and α and β are the slopes for the two dependences. The $P_{2\infty}$ value is then used to obtain the dipole moment from the approximate equation below (with atomic polarization neglected and R_D used in place of electronic polarization):

$$\mu = 0.01283 \sqrt{P_{2\infty} - R_D T} \quad (7)$$

where R_D is the molar refraction.

Hedestrand [37] used extrapolation of the dielectric constant and the density of the

solution directly to zero concentration of the solute. This gives an equation

$$P_{2\infty} = [(\epsilon_1 - 1)/(\epsilon_1 + 2)] \cdot M_2/d_1 + (M_1/d_1) \cdot [3\alpha\epsilon_1 - \beta(\epsilon_1 - 1)(\epsilon_1 + 2)]/(\epsilon_1 + 2)^2 \quad (8)$$

Again, the α and β are the coefficients characterizing the slopes of the straight lines $\epsilon = \epsilon_1(1 + \alpha x_2)$ and $d = d_1(1 + \beta x_2)$ describing the dependence of ϵ and d on the molar fraction of the solute, x_2 .

Other approaches replace densities with refractivity indices (Guggenheim [38], Smith [39]), while Higasi uses only dielectric constant measurements [40]. Numerous other equations have been introduced. An excellent description, discussion, and critical evaluation of these methods can be found in the literature [11,18]. An improved method of dipole moment measurements was developed by Imbach *et al.* [41].

The equation we have found most useful in the determination of dipole moments of heterocyclic compounds is based on equations developed by Hedestrand [37], Guggenheim [38], and Smith [39] and has the form

$$\mu^2 = (27kT/4\pi N) \cdot [1/d_1(\epsilon_1 + 2)^2] \cdot (A_{\epsilon_2} - A_{n_2})M_2 \quad (9)$$

where k is the Boltzmann constant (1.381×10^{-16} erg-deg $^{-1}$), T is the absolute temperature, N is Avogadro's number, d_1 and ϵ_1 the density and the dielectric constant of the solvent, respectively, A_{ϵ_2} and A_{n_2} are the numerical values obtained from the solute dielectric constant and refractive index measurements, respectively, and M_2 is the molecular weight of the solute.

Exner has compared and statistically tested the various equations used to determine dipole moments in solution [42].

The dielectric constants needed for the determination of dipole moments are obtained by a heterodyne beat method or the resonance and bridge methods. The instrument specifically adapted for measurements of dielectric constants for dipole moments is a dipole meter model DM-01 manufactured by Wissenschaftlich-Technische Werkstätten, Weilheim, Germany, equipped with two different thermostatted cells for two ranges of dielectric constants (cell DFL-1, ϵ 1.0 to 3.4, cell DFL-2, ϵ 2.0 to 6.9). The other dipole meter which used to be available from Toshniwal Brothers, Ltd., Madras, India, is not sufficiently sensitive for measurements of dielectric constants of diluted solutions and is better suited for measurements of large differences, such as the dielectric constants and dipole moments of different solvents.

The measurements involve the determination of differences between the dielectric constant of a solution containing the solute in question, and the dielectric constant of the pure solvent. This means that solvents of low polarity (with a low dielectric constant) are the preferred choice as the results are more accurate. A sufficient concentration of the solute is needed to make the two dielectric constants sufficiently different. The remaining needed experimental data - the refractive indices and densities - are easily obtained.

The two nonpolar solvents most commonly used for the determination of dipole moments by the dielectric constant method are benzene and dioxane, however, numerous other solvents have been used for this purpose. The use of other solvents often becomes a necessity in heterocyclic chemistry where many heterocycles with polar groups are not soluble in solvents of low polarity. For example, we have successfully used ethyl acetate, acetic acid, and *n*-butyl ether. In some cases, the use of mixed solvents was necessary (dioxane-morpholine, dioxane-

dimethylformamide). Table 1 summarizes the dielectric constants and refractive indices of some most commonly used solvents.

It is not surprising that, for a number of various reasons (solute-solvent interactions, hydrogen bonding, complex formation, solvent polarity), dipole moments depend on the nature of the solvent. Thus, somewhat different values can be obtained in different solvents, and they are normally different from gas-phase measurements.

Table 1
Dielectric constants and refractive indices of solvents used in dipole moment measurements ^a

Solvent	ϵ^{25}	n_D^{25}	Solvent	ϵ^{25}	n_D^{25}
Pentane	1.836	1.3548	Toluene	2.379	1.4893 ^b
Hexane	1.889	2.3722	Carbon disulfide	2.641 ^c	1.6254 ^d
Heptane	1.917	1.3851	Diethyl ether	4.265	1.3526 ^c
Cyclohexane	2.015	1.4233	Chloroform	4.724	1.4459 ^c
Dioxane	2.209	1.4206	Chlorobenzene	5.621	1.5241 ^c
Carbon tetrachloride	2.228	1.4570	Ethyl acetate	6.020	1.3723 ^c
Benzene	2.274	1.4979	Acetic acid	6.150 ^c	1.3716 ^c

^aAt 25°C unless indicated otherwise. The data were taken from the literature, from several reference sources. The table lists solvents with low dielectric constants, suitable for use with the dipole meter DM-01. ^b24°C. ^c20°C. ^d23.5°C.

2.2. Microwave methods

The second most important general method of determination of dipole moments is microwave spectroscopy. Many reliable values were obtained from the frequencies of the lines in rotational spectra by calculating the three principal moments of inertia of a molecule with respect to the axes *x*, *y*, and *z* and using them to evaluate the geometric parameters of a particular structure [43,44]. All polar molecules give pure rotational spectra whereas molecules with no dipole moments give no such spectra. An external electric field is applied and its intensity is determined by calibration (commonly with carbonyl sulfide, COS) [45].

A simplified but much less accurate method determines only the basic features of a molecule and not its complete molecular structure [46]. Numerous dipole moments were determined by the dielectric absorption method in the microwave region [47-49].

Because microwave spectroscopy is less commonly used for the determination of dipole moments of aromatic heterocycles than the dielectric constant methods, only a brief summary of the various modifications and possibilities will be presented here.

One important point is worth mentioning. It is well known that dipole moment values

measured in solution (dielectric constant measurements) are subject to a solvent effect which depends on the interactions between the solute and the solvent. In the microwave method, the measurements are carried out in the gas phase, without any solvent interference. However, it turns out that, in most cases, there is a good agreement between the values obtained by the two methods [29].

2.3. The Stark effect method

The Stark effect is related to a change in energy levels of atoms and molecules in the presence of a strong external electric field and is observed as a shift and splitting of the spectral rotational lines. The applicability of the method is restricted to the gas phase and more complex compounds require the use of isotopically labeled molecules [43,44,50-52].

2.4. Molecular beam method

The molecular beam method employs deflection of molecular beams in a nonuniform electric field. The displacement of the beam is used to calculate the dipole moment. Typically, the accuracy of this method is relatively low [3,53,54].

2.5. Electric resonance method

The electric resonance method (molecular beam electric resonance method) uses three different electric fields. It is time-consuming, however, while it can be used only for simple molecules, it gives very accurate results [3,53].

2.6. Raman spectroscopy

Bakhshiev developed a method for the determination of ground- and excited-state dipole moments based on Raman spectra which has been successfully used for different compounds [55-57].

2.7. Sign and direction of the dipole moment

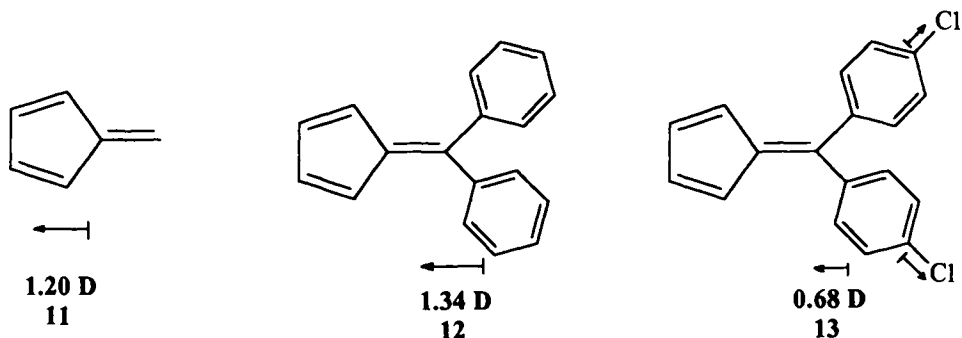
As mentioned in the introduction, in organic chemistry the positive direction of the dipole moment is normally defined as the direction from the center of the positive charge towards the center of the negative charge [19,58,59]. This convention also prevails in physical organic chemistry and in inorganic chemistry [14]. However, while the dipole still points from the positive charge towards the negative charge, in physical chemistry and in chemical physics the positive direction of the dipole moment is defined in the opposite way, i.e., from the negative charge to the positive charge [60-63].

Whereas calculated dipole moments are obtained as a vector, with a specified direction, experimentally obtained dipole moments typically do not specify direction and thus are obtained as a scalar. Thus, additional work is usually needed to obtain the direction. The answer can be provided in several different ways.

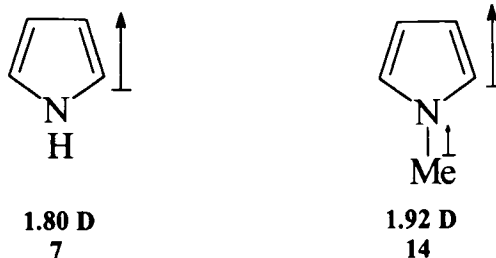
One possibility is to use hybrid orbitals. For example, any s hybridization in the HCl bond will give the correct direction of the dipole moment, $\text{H}\overset{+}{\text{C}}\text{Cl}$ [58]. In a similar fashion, consideration of the possible resonance structures in a molecule and determination of the most important one (based on the electronegativities of the respective atoms) will give the correct direction. For example, in HF , $\text{H}\overset{+}{\text{F}}\text{F}^-$ is much more important than $\text{H}^-\overset{+}{\text{F}}$, thus giving the direction of the dipole moment as $\overset{+}{\text{H}}\text{F}$.

Precise measurements of the effect of isotopic substitution on hyperfine Zeeman splitting effects [64] have provided the correct direction of the dipole moment for carbon monoxide, CO [65], deuterium iodide, DI [66], and carbonyl sulfide, OCS [67].

In organic chemistry, another approach utilizes the measurement of dipole moments of additional substituted derivatives of compounds in question. Thus, e.g., theoretical calculations indicate that the direction of the dipole moment in fulvene (11) is from the exocyclic methylene group toward the five-membered ring (exocyclic methylene group positive, the ring negative). The experimental dipole moment of fulvene is 1.20 D [68]. The two derivatives of fulvene used in the study, 6,6-diphenylfulvene (12) and 6,6-bis-(*p*-chlorophenyl)fulvene (13), have $\mu = 1.34$ D and 0.68 D, respectively. Because in chlorobenzene the bond moment of the C-Cl bond is directed towards the chlorine atom and the bis-*p*-chlorophenyl derivative has a lower dipole moment than the diphenyl derivative, it is clear that the dipole moment in fulvene is directed towards the five-membered ring [68].



One more example with a heterocyclic compound will be presented. As mentioned before, the dipole moment of pyrrole (7) is 1.80 D [69,70] and the positive direction (+ to -) is from the nitrogen heteroatom towards the remaining portion of the five-membered ring. To confirm this, one can use 1-methylpyrrole (14) as a reference compound. The dipole moment of 1-methylpyrrole is 1.92 D [69]. Because the bond moment of the N-Me bond is directed from the methyl group towards the nitrogen and 1-methylpyrrole has a higher dipole moment than pyrrole (1.92 D and 1.80 D, respectively), the dipole moment of pyrrole must be directed from the nitrogen atom towards the ring.



3. CALCULATED GROUND-STATE DIPOLE MOMENTS

In quantum-chemical calculations of dipole moments, each possible or expected structure is characterized by its inherent wave function, ψ , and its inherent electronic configuration [11]. The dipole moment is then defined by eq. (10) [11].

$$\mu = -e \int \sum_i r_i [\psi(1,2,3, \dots, n)]^2 dV_1 dV_2 dV_3 \dots dV_n + e \sum_j z_j r_j \quad (10)$$

where e is the electronic charge, r_i is the radius vector of the i -th electron, and z_j and r_j are the charge of the j -th atomic nucleus and its radius vector, respectively.

In principle, methods that can be used to calculate dipole moments are the various MO methods, which can be divided into three large groups: empirical, semiempirical, and *ab initio*. In contrast, VB methods are generally not suited for calculations of dipole moments because here the motion of electrons is completely synchronized. If, however, one uses additional ionic,

excited VB structures, it is possible to analyze bond moments, such as the $\overset{\delta-}{\text{C}}-\overset{\delta+}{\text{H}}$ or $\overset{\delta-}{\text{N}}-\overset{\delta+}{\text{H}}$, in their r dependence (r = interatomic distance) using the FORS-IACC (full optimized reaction space-intraatomic correlation correction) model [71].

In the various MO methods (HMO, SCF-MO, PPP, etc.), the dipole moment, which corresponds to a system of point electric charges (charge densities), is defined by the relation

$$\mu = \sum_i q_i r_i \quad (11)$$

where q_i are the charges and r_i are their vector distances from the origin. The magnitude of the dipole moment, μ , does not depend on the choice of the origin in the coordinate system, and the most convenient point is usually selected as the origin.

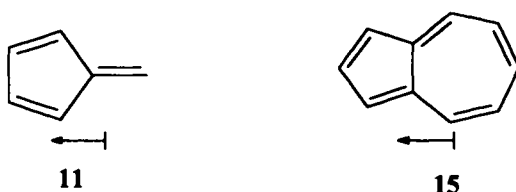
A significant advantage of calculated dipole moments is that they are vectors and thus give a direction of the dipole moment whereas experimental dipole moments are normally obtained without any defined direction.

3.1. Empirical methods

In the HMO method, only the π -component of the dipole moment is obtained, while with all-electron methods (extended Hückel, EHT), a total dipole moment is calculated. If only a π -moment is available by using a π -electron method, the σ -component can be obtained as a vector

sum of the individual σ -bond and group moments. The total dipole moment is then computed as a vector sum of the π -moment and σ -moment. It should be kept in mind that, in this procedure, the polarization interaction of the π - and σ -moments is not taken into consideration.

One of the deficiencies of the MO methods, especially the simple ones, is that they tend to exaggerate uneven distribution of electrons in a molecule and thus make it more polar (with a higher dipole moment) than it actually is. The result is that dipole moments which are based on charge densities obtained from eigenfunctions of the MO approximations are usually considerably higher than the actual experimental dipole moments. Two old, well-known examples of theoretical dipole moments obtained by the HMO method are fulvene (11) whose calculated dipole moment is 4.7 D [72-74] and the experimental value is 1.2 D [68], and azulene (15), with a calculated dipole moment of 6.9 D [72] and the experimental value of 1.0 D [68].

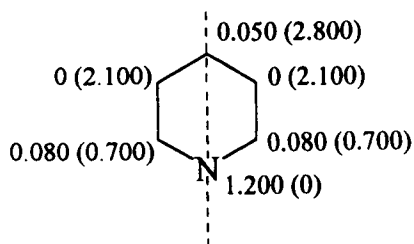


Numerous other similar examples are available. Various improvements of the simple MO theory usually give lower dipole moments which are closer to the actual experimental values (variation of β with bond distances, the ω -technique, etc.).

As a rule, a much better agreement can be obtained for compounds with heteroatoms because the additional parameters used in the calculations (the Coulomb integral of the heteroatoms, the resonance integral of the carbon-heteroatom bonds, etc.) can be adjusted to give a reasonable fit with experimental data.

Pyridine can be used as a practical example (the same compound was used by Streitwieser [58]; Minkin *et al.* used imidazole as a similar example [11]). In pyridine only the vectors along its symmetry axis (y -axis as shown below) need to be considered as all the moments along the x -axis cancel out. In our HMO calculation, the following parameters were used: $\alpha_N = \alpha_C + 0.5\beta_{CC}$; $\beta_{CN} = \beta_{CC}$; regular hexagon, all bond distances $d = 1.40 \text{ \AA}$) [75].

π -Moment:



In the diagram on the previous page, charge densities are shown at each atom; the respective y -component of each r_i is shown in parentheses. The nitrogen atom is the origin.

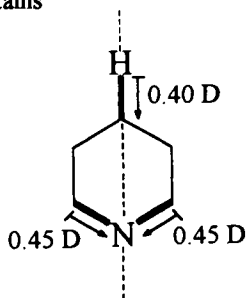
Thus,

$$\mu_{\pi} = -4.80[(2 \times 0.080 \times 0.700) + (2 \times 0 \times 2.100) + (0.050 \times 2.800)] = -1.21 \text{ D} \quad (12)$$

(directed towards nitrogen)

σ -Moment:

To calculate the σ -moment, one needs to consider only the two $\overset{+}{\text{C}}-\overset{-}{\text{N}}$ bonds (C-N, 0.45 D) and the C(4)-H bond ($\overset{+}{\text{H}}-\overset{-}{\text{C}}$, 0.40 D) (the y -components only as all other vectors cancel out). Using the σ -bond moments, one obtains



$$\mu_{\sigma} = -(2 \times 0.45 \times 0.500) - 0.40 = -0.85 \text{ D} \quad (13)$$

(directed towards nitrogen)

The minus sign indicates that the dipole moment is directed towards nitrogen. If we neglect the sign and are interested in the numerical value only, the total dipole moment will be:

$$\mu_{\text{t}} = \mu_{\pi} + \mu_{\sigma} = 1.21 + 0.85 = 2.06 \text{ D} \quad (14)$$

(directed towards nitrogen)

The experimental dipole moment of pyridine is 2.22 D [76,77], in a good agreement with the calculated value.

It should be pointed out that, in spite of the above success, in most cases dipole moments obtained on the basis of HMO π -components are much higher than experimental values. Because the HMO method represented an important approach in the past, a number of efforts were made to improve the agreement between the theoretical and experimental dipole moments [68,78-80]. Good values of dipole moments were obtained for several heterocyclic compounds by the use of a correction factor in MO calculations [81].

To obtain a total dipole moment, a combination of the simple HMO method and the Del

Re approach were used. In this way, dipole moments for several nitrogen heterocycles were obtained that were in an excellent agreement with the experimental values [82].

Similarly as with the HMO method, the use of the EHT method (extended Hückel) which considers both the π - and the σ -electrons, leads to exaggerated values of the dipole moments [83,84].

3.2. Semiempirical methods

Numerous semiempirical methods have been used to calculate dipole moments (e.g., PPP, CNDO/2, CNDO/S, other CNDO variations, INDO, INDO/S, MNDO, MINDO/3, AM1, HAM3, etc.). They can be divided into π -electron and all-valence-electron methods. In π -electron methods such as, e.g., the PPP (LCI-SCF-MO) method, only the π -component of the dipole moment is obtained and the σ -component has to be computed separately. As in the case of empirical methods, one possibility is to calculate the σ -component as a vector sum of the individual σ -bond and group moments. These values are readily available from several sources [4-6,11,18,19,85]. The resulting total (overall) dipole moment is then computed as a vector sum of the π -moment and the σ -moment.

The σ -components can also be obtained by calculation of all σ -moments according to the approach suggested by Mulliken and Coulson [86,87], with the inclusion of homopolar and atomic dipoles and using Slater atomic orbitals [88-90]. Other possibilities are to calculate the σ -electronic charges on the individual atoms [91-93] or the Del Re approach [94,95]. Some of the early papers devoted to dipole moments calculated by the PPP method were later subjected to a criticism by Exner [18]. In one of the older studies, satisfactory dipole moments were obtained for pyridine and pyrrole using the variable electronegativity SCF method [96].

The use of the various all-valence-electron methods (CNDO/2, CNDO/S, INDO, and the other above-mentioned methods) gives values of total dipole moments which are generally in good agreement with experimental values [97-99].

The CNDO/2 method [100,101] and the related methods have been successfully used to compute dipole moments [97,102] and there are numerous publications devoted to the use of the CNDO and INDO methods for calculations of dipole moments of heterocyclic compounds. However, in certain cases, some of these methods tend to lead to somewhat higher dipole moments than the actual experimental values.

A number of authors have compared the validity of various semiempirical methods and optimized the parameters [103-106]. It does not seem practical to give references to all calculations on this topic. Selected recent calculations of dipole moments of pyrimidine and purine bases and their derivatives and analogs will be mentioned here as practical examples [107-114]. Numerous references on semiempirical calculations of dipole moments of different types of heterocycles including pyrimidines and purines can be found in our publications [115-133].

In our work, we have successfully used the PPP method (with the σ -component obtained as a vector sum of the σ -bond and group moments) and the CNDO/2 method, with an excellent or at least a good agreement between calculated and experimental values. The compounds studied were 1,3-diaryl and 1-aryl-3-heteroaryltriazenes [115], indoles [116], purines [118,119,133], quinazolines [120], phenothiazines [121,124,127], pyrimidines [122,133], pteridines [123], coumarins [125,128], acridines [126], phenazines [126], benzo[*a*]phenothiazines [127,129,132], and the heterocyclic dye merocyanine 540 [130,131].

3.3. *Ab initio* methods

The number of papers with *ab initio* calculations of dipole moments of heterocycles has been continually increasing throughout the last few years. Today, *ab initio* dipole moments are available for most major parent heterocyclic systems and numerous derivatives, especially the nitrogen-containing heterocycles in general and pyrimidine and purine bases in particular. A number of different basis sets and approaches were used.

Some of the recent calculations on pyrimidine and purine bases can be listed here [134-144]. In the case of simple monocyclic compounds with one heteroatom and their derivatives, *ab initio* approach has been used for pyrrole [145,146], furan [146-148], thiophene [146,147], pyridine [149-151], and phosphabenzene [149].

Ab initio dipole moments are also available for numerous other heterocyclic compounds (heterocycles with fused rings, heterocycles with two or more heteroatoms, etc.).

3.4. Semiempirical and *ab initio* methods - a comparison

Rather than listing the actual numerical values of dipole moments for various types of heterocycles, it seems more practical to make a simple comparison of several methods based predominantly on our work. As an example, the compounds presented in Table 2 are selected pyrimidines and purines studied in our previous work for which a good agreement between the experimental and calculated dipole moments has been found [118,119,122,133]. The comparison includes the experimental dipole moments and the theoretical values obtained by the PPP + σ -bond moment calculations, by the CNDO/2 method, and by *ab initio* calculations. HMO calculations are not included because, as a rule, the values are too high. A comparison of the three sets of calculated values indicates that there is no clear-cut preference for any of the methods as far as the numerical values of the dipole moments are concerned, although only the *ab initio* calculations provide values which are most sound from the theoretical and physical point of view. All three methods yield results which, in general, are in a good agreement with the experimental values.

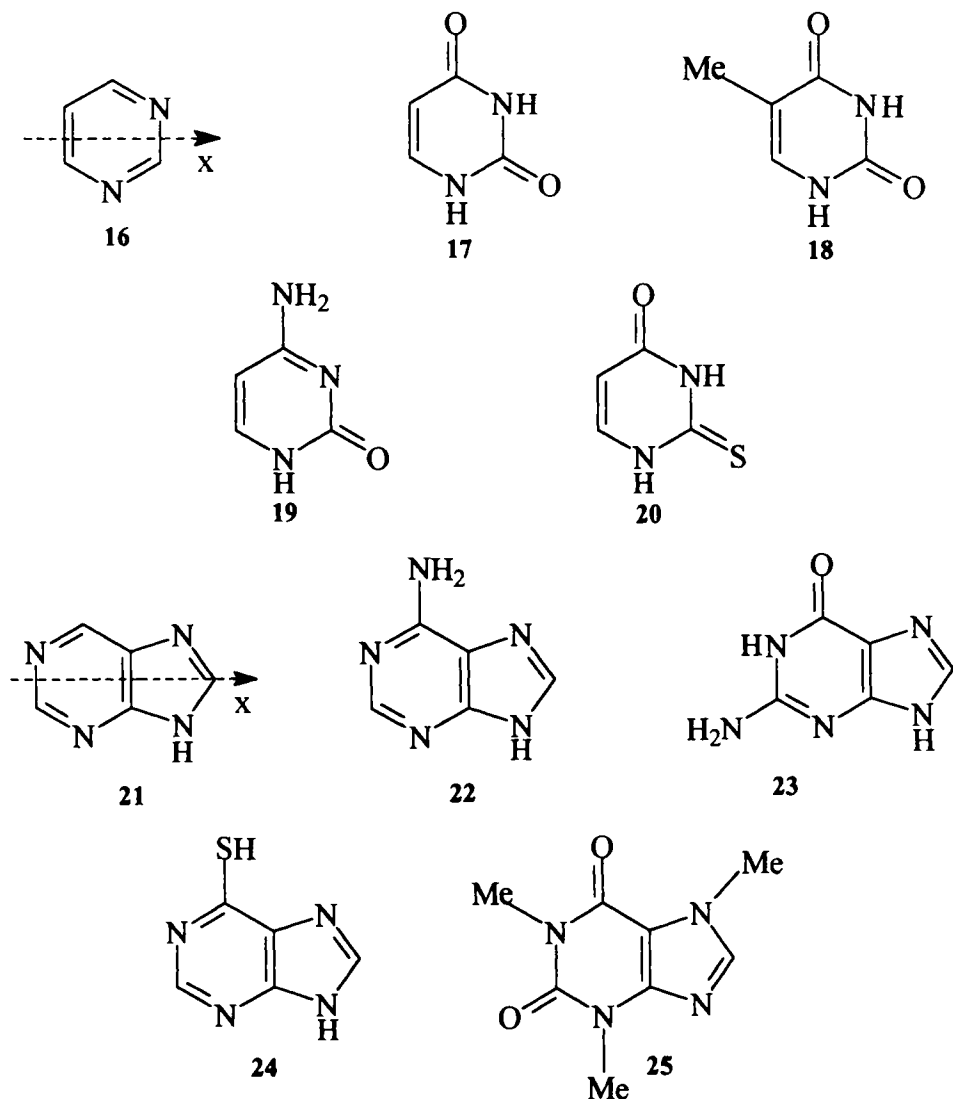
It is important to point out, however, that in many cases the agreement is not as good as for the compounds shown in Table 2.

The role of the solvent (solute-solvent interactions) has not been considered in these calculations. However, efforts to include the role of the solvent have been made in various other calculations.

4. EXPERIMENTAL EXCITED-STATE DIPOLE MOMENTS

Electronically excited states of organic molecules possess a different distribution of electrons within a molecule and, consequently, a different dipole moment. It is important to mention that ground-state dipole moments and the corresponding singlet or triplet excited-state dipole moments are not necessarily collinear although, for the sake of simplicity, it is usually assumed that they are. Excited-state dipole moments have been reviewed in previous publications [11,18,167].

In general, experimental methods for the determination of excited-state dipole moments are based on experimental ground-state dipole moments and a change of the position of a spectral band (in an electronic spectrum) caused by an electric field which can be external (electro-



chromism) or internal (solvatochromism).

Electrooptical methods (electrochromism) which are more accurate but which are experimentally more difficult include electric polarization of fluorescence or phosphorescence, electric dichroism, and absorption spectra in the vapor phase in an electric field (Stark effect).

The solvatochromic methods (solvent-shift methods) are experimentally much simpler as they do not use any external field. However, they are less reliable because of the numerous

Table 2
Experimental and calculated ground-state dipole moments of selected pyrimidines and purines

No.	Compound	Dipole moment, μ^a				Direction, θ^f
		Exp ^b	PPP+ σ^c	CNDO/2 ^d	<i>Ab initio</i> ^e	
16	Pyrimidine	2.0 ^g , 2.44 [154,155]	1.86	2.00 [156]	2.42	330°
17	Uracil	4.16, 5.18 [157,158]	5.06	4.61 [159]	4.72	53°
18	Thymine	3.95, 4.13 [157,158]	4.93	4.35 [159]	4.64	52°
19	Cytosine	~7.0 [160]	7.21	7.61 [159]	7.12	339°
20	2-Thiouracil	4.21 [161]	5.34	4.68 [122]	5.32	39°
21	Purine	4.32 [162]	4.35	4.19 [163]	3.66	225°
22	Adenine	3.00, 3.85 [17,119, 164]	3.66	2.86 [159]	2.47, 2.83 [165]	168°
23	Guanine	5.50, [119]	8.37	7.26 [159]	6.84	52°
24	6-Mercaptopurine	3.59 ^h [118]	3.79	-	4.38	240°
25	Caffeine	4.60 [166]	4.58	4.41 [109]	4.35	172°

^aIn Debye units. In the case of the purines, the calculated values are for the 9*H*-tautomers wherever appropriate (as shown in the formulas). The numbers in the brackets are the references. ^bIn dioxane unless a different solvent is indicated. ^cVector sum of the π - and σ -components; π -component: PPP method, σ -component: σ -bond (and σ -group) moments. References: pyrimidines [122], purines [118,119]. ^dFor another set of values on CNDO/2 dipole moments of pyrimidines, see [122] (our calculations). ^eSPARTAN, 6-31G** level. Our work, to be published [133]. ^f θ is the angle between the positive direction of the x-axis and the calculated dipole moment (PPP + σ) read counterclockwise, for the orientation of the structures as shown in the formulas [118,119,122]. ^gIn benzene. ^hIn acetic acid.

simplifications. In this approach, various solvatochromic equations are used which utilize ground-state dipole moments and shifts of the absorption and emission (fluorescence, phosphorescence) maxima in solvents of different polarity. The equations most commonly used for compounds which give fluorescence (and/or phosphorescence) are those developed by Kawski, Chamma and Viallet [168-170] and by Bakhshiev [171]. In the case of nonemitting compounds, only absorption spectra can be employed (McRae [172] and Suppan [173,174] equations). To conserve space, the actual equations will not be shown here. Numerous examples of their use and the respective solvatochromic equations can be found in our publications [120-132]. All these equations require the use of solvent functions and Onsager cavity radii which can be conveniently determined from solid-state densities of the compounds under study using the Suppan equation [175].

Because of the many assumptions and approximations used in the solvatochromic methods, the agreement between the experimental and calculated excited-state dipole moments is not always good. For obvious reasons, solvatochromic equations using both the absorption and emission spectral data (Kawski-Chamma-Viallet, Bakhshiev) give considerably better results than the equations based strictly on absorption spectra (McRae, Suppan). However, even the results obtained by using the Kawski-Chamma-Viallet and Bakhshiev equations are often considerably different. It is quite difficult to decide which of the equations should represent the preferred choice.

In addition to the collinearity assumption (the ground and excited-state dipole moments are taken as collinear), specific solute-solvent interactions are not considered and solvent effects on absorption and emission profiles are neglected. Incomplete relaxation prior to emission is always possible. Also, the use of some solvatochromic equations can lead to negative values and imaginary values for some compounds. Finally, it is important to remember that, even if the ground- and excited-state dipole moments are considered to be collinear or at least approximately collinear, parallel and antiparallel orientations of the ground- and excited-state dipole moments should be considered.

A statistical evaluation and comparison of solvatochromic methods used to determine excited-state dipole moments has been carried out by Koutek [167]. Solvent effects can be taken into consideration using the reaction field theory developed by Katritzky, Zerner, Szafran, and Karelson [176-179] and Siretskii, Kirillov, and Bakhshiev [180] have proposed an equation containing a $\cos \phi$ where ϕ is the angle between the direction of the ground-state dipole moment and the excited-state dipole moment. The equation worked well for certain aromatic dyes but its general applicability has not been tested.

Prabhumirashi and Kunte [181] have proposed a new procedure employing Bakhshiev's equation for solvatochromic frequency shifts for excited-state dipole moments and specific solute-solvent interaction energies based on absorption spectra only, without using emission spectra. Suppan [182] has expanded upon the solvatochromic shift method and discussed the effect of the medium on the energies of electronic states and Ghoneim and Suppan [183] discussed solvatochromic shifts of non-dipolar molecules in polar solvents.

Ayachit and Tonannavar [184] have critically examined Suppan's method and developed two improvements intended to resolve the uncertainty concerning the direction of excited-state dipole moments. Kurbako *et al.* [185] have developed a new solvent shift approach based on a theory of universal intermolecular interactions. Abe has proposed a method of estimating the angle between ground- and excited-state dipole moments based on improved solvatochromic

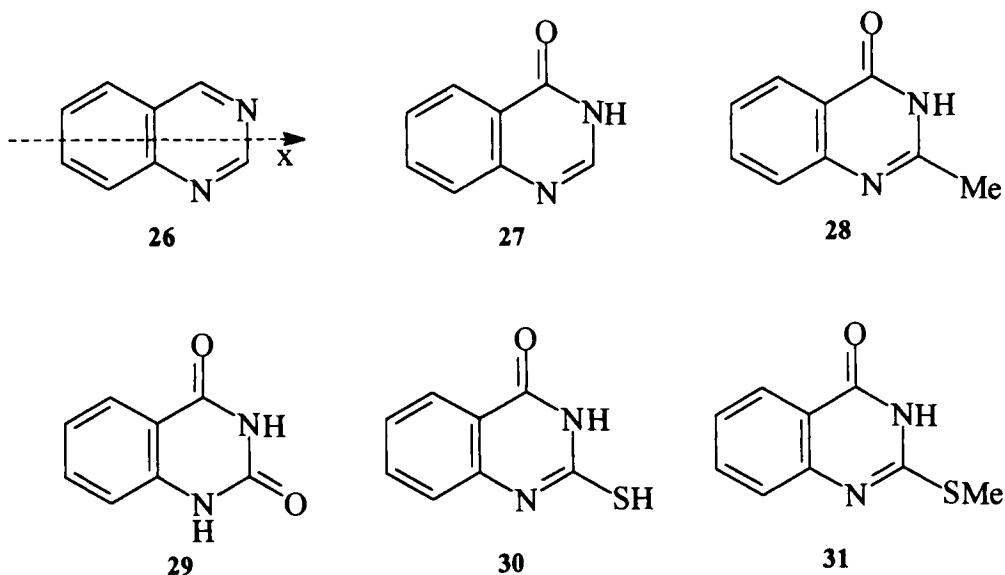
equations [186-189].

5. CALCULATED EXCITED-STATE DIPOLE MOMENTS

In our PPP + σ -moment calculations, we assume that the excited-state π -component (as obtained by the PPP calculation) is different from the ground-state dipole moment, whereas the σ -moment remains unchanged (i.e., the same as in the ground state). This approach has worked very well for a number of different series of organic heterocycles.

The number of calculated excited-state dipole moments is still relatively limited (see the references in our publications mentioned above) and, because of the numerous approximations and problems related concerning experimental excited-state dipole moments, the agreement between the experimental and calculated values is often poor.

As an example, the experimental and calculated first excited singlet-state dipole moments for selected fluorescent quinazolines are presented in Table 3, with the ground-state values included for comparison [120]. For most compounds, the agreement is good or at least acceptable.



Our studies of excited-state dipole moments of various types of heterocyclic compounds indicate that, while in many cases the polarity of the respective molecules increases in electronically excited states as compared to the ground states and this results in excited-state dipole moments that are higher than ground-state values, in some cases the opposite is true and excited-state dipole moments are lower than their ground-state counterparts. This can be

Table 3
Experimental and calculated ground and first excited singlet-state dipole moments of selected quinazolines^a

No.	Compound	Dipole moment, ground state			Dipole moment, first excited singlet state		
		Exp ^b	Calc ^c	θ^d	Exp ^e	Calc ^f	θ^g
26	Quinazoline	3.40	2.33	323°	1.10	3.62	330°
27	4(3 <i>H</i>)-Quinazolinone	3.12	3.36	148°	1.54	3.78	143°
28	2-Methyl-4(3 <i>H</i>)-Quinazolinone	3.83	3.65	150°	5.36	4.98	143°
29	2,4(1 <i>H</i> ,3 <i>H</i>)-Quinazolinone	3.85	4.20	45°	5.40	5.64	85°
30	2-Mercapto-4(3 <i>H</i>)-quinazolinone	3.21	3.77	159°	5.65	5.45	147°
31	2-Methylmercapto-4(3 <i>H</i>)-quinazolinone	5.06	4.75	167°	6.14	7.23	142°

^aIn Debye units. Ref. [120]. ^bMeasured in morpholine-dioxane (5 : 3, vol.). ^cComputed as the vector sum of the π -component (PPP calculation) and the σ -component (σ -bond and group moments; cf. Table 2). ^dAngle between the positive direction of the x-axis and the ground-state dipole moment read counterclockwise, for the orientation of the structures as shown in the formulas. ^eFrom Kawski-Chamma-Viallet correlations. ^fVector sum of the first excited singlet state π -contribution (PPP) and the σ -component which is the same as in the ground state. ^gAngle between the positive direction of the x-axis and the first excited singlet-state dipole moment read counterclockwise.

observed with experimental values as well as with calculated dipole moments.

In principle, experimental and calculated triplet excited-state dipole moments can be obtained in the same fashion as singlet excited-state dipole moments. However, the number of studies in this area is limited (references [17,162,190-193]) can serve as examples.

Additional work in the area of excited-state dipole moments is needed, to make the agreement between experimental and theoretical values better and the values more reliable than they are today. However, the results obtained so far seem to be promising.

6. CONCLUSION

The above material represents a concise overview devoted to experimental and theoretical dipole moments of aromatic heterocycles, their determination and computation. Only selected relevant references are included, with a particular attention to our own work. The coverage is not exhaustive.

The material presented in the chapter illustrates the difficulties connected with experimental determination and theoretical computation of dipole moments in general and heterocyclic compounds in particular. It is obvious that more work is needed on solute-solvent interactions, role of the polarity of the solvent, dipole moments in solution vs. dipole moments in the gas phase, dielectric constant measurements vs. microwave spectroscopic data, improvements in calculations, better solvatochromic equations, etc.

It is also clear that dipole moments still represent one of the important physical characteristics of compounds with an uneven distribution of charge in general and aromatic heterocycles in particular.

ACKNOWLEDGMENT

We appreciate the financial support provided by the North Atlantic Treaty Organization, Brussels, Belgium (collaborative grant 0352/87). Also, we wish to thank Professors Jacques E. Dubois and Pierre C. Lacaze (Paris, France), Jacques Barbe (Marseille, France), William C. Herndon (El Paso, TX), László von Szentpály and Ratna Ghosh (Kingston, Jamaica), Jane S. Murray and Peter Politzer (New Orleans, LA), Marwan Dakkouri (Ulm, Germany), and Otto Exner (Prague, Czech Republic) for their interest in our work and for valuable discussions.

REFERENCES

1. G.E.K. Branch and M. Calvin, *The Theory of Organic Chemistry*, Chap. 18, Prentice-Hall, Englewood Cliffs, NJ, 1941.
2. R.J.W. Le Fèvre, *Dipole Moments*, Methuen, London, 1953.
3. J.W. Smith, *Electric Dipole Moments*, Butterworths, London, 1955.
4. C.P. Smyth, *Dielectric Behaviour and Structure*, McGraw-Hill, New York, NY, 1955.
5. L.E. Sutton, in: E.A. Braude and F.C. Nachod (eds.), *Determination of Organic Structures by Physical Methods*, p. 373, Academic Press, New York, NY, 1955.

6. H. Fröhlich, *Theory of Dielectrics*, 2nd ed., Clarendon Press, Oxford, 1958.
7. S. Walker, in: A. R. Katritzky (ed.), *Physical Methods in Heterocyclic Chemistry*, Vol. I, p. 189, Academic Press, New York, NY, 1963.
8. B.L. Shaw, in: J.C.P. Schwarz (ed.), *Physical Methods in Organic Chemistry*, Chap. 20, p. 323, Oliver & Boyd, Edinburgh, 1964.
9. K. Higasi, H. Baba, and A. Rembaum, *Quantum Organic Chemistry*, Chap. 8, p. 154, J. Wiley-Interscience, New York, NY, 1965.
10. N.E. Hill, W.E. Vaughn, A.H. Price, and M. Davies, *Dielectric Properties and Molecular Behaviour*, Van Nostrand Reinhold, London, 1969.
11. V.I. Minkin, O.A. Osipov, and Yu. A. Zhdanov, *Dipole Moments in Organic Chemistry*, Plenum Press, New York, NY, 1970; *Dipol'nye Momenty*, Khimiya, Leningrad, 1968.
12. A.D. Garnovskii, Yu.Y. Kolodyazhnyi, O.A. Osipov, V.I. Minkin, S.A. Giller, I.B. Mazeika, and I.I. Grandberg, *Khim. Geterotsikl. Soedin.*, (1971) 867.
13. J. Kraft and S. Walker, in: A.R. Katritzky (ed.), *Physical Methods in Heterocyclic Chemistry*, Vol. IV, p. 237, Academic Press, New York, NY, 1971.
14. G.J. Moody and J.D.R. Thomas, *Dipole Moments in Inorganic Chemistry*, E. Arnold, London, 1971.
15. C.P. Smyth, in: A. Weissberger and B.W. Rossiter (eds.), *Physical Methods of Chemistry*, Part 4, p. 397, J. Wiley-Interscience, New York, NY, 1972.
16. K. Bergström, *Measurements of Dielectric Constants and Electric Dipole Moments in the Gas Phase*, Lund Institute of Technology, Lund, 1973.
17. W. Liptay, in: E.C. Lim (ed.), *Excited States*, Vol. I, p. 129, Academic Press, New York, NY, 1974.
18. O. Exner, *Dipole Moments in Organic Chemistry*, G. Thieme Verlag, Stuttgart, 1975.
19. L.N. Ferguson, *Organic Molecular Structure*, Willard Grant Press, Boston, MA, 1975.
20. O.A. Osipov, A.D. Garnovskii, and V.I. Minkin, *Dipole Moments in the Chemistry of Complex Compounds*, Izd. Rostov. Univ., Rostov-on-Don, 1976.
21. L. Sobczyk, H. Engelhardt, and K. Bunzl, *The Hydrogen Bond*, 3 (1976) 937.
22. *Digest of Literature on Dielectrics*, National Academy of Sciences, Washington, DC, published annually until 1977.
23. V.I. Minkin, in: H.B. Kagan (ed.), *Stereochemistry: Fundamentals and Methods*, Vol. 2: *Determination of Configurations by Dipole Moments, CD or ORD*, p. 1, G. Thieme Verlag, Stuttgart, 1977.
24. C.P. Smyth, *Dielectric Constant and Molecular Structure*, Chap. 1, Chemical Catalog Co., New York, NY, 1931.
25. L.G. Wesson, *Tables of Electric Dipole Moments*, The Technology Press (Massachusetts Institute of Technology), Cambridge, MA, 1948.
26. A.A. Maryott and F. Buckley, *Tables of Dielectric Constants and Electric Dipole Moments of Substances in Gaseous State*, Circular 573, National Bureau of Standards, Washington, DC, 1953.
27. R.D. Nelson, D.R. Lide, and A.A. Maryott, *Selected Values of Electric Dipole Moments for Molecules in the Gas Phase*, NSRDS-NBS (National Bureau of Standards) 10, Washington, DC, 1967.
28. O.A. Osipov, V.I. Minkin, and A. D. Garnovskii, *Spravochnik po Dipol'nym Momentam (Handbook of Dipole Moments)*, Vysshaya Shkola, Moscow, 1971.

29. A.L. McClellan, *Tables of Experimental Dipole Moments*, Vol. 1, W.H. Freeman, San Francisco, CA, 1963.
30. A.L. McClellan, *Tables of Experimental Dipole Moments*, Vol. 2, Rahara Enterprises, El Cerrito, CA, 1974.
31. A.L. McClellan, *Tables of Experimental Dipole Moments*, Vol. 3, Rahara Enterprises, El Cerrito, CA, 1989.
32. J.A. Joule and G.F. Smith, *Heterocyclic Chemistry*, p.12, Van Nostrand Reinhold, London, 1972 (repr. 1975).
33. M.V. Sargent and T.M. Cresp, in: P.G. Sammes (ed.), *Comprehensive Organic Chemistry*, Part 18.4., p. 624, Pergamon Press, New York, NY, 1979.
34. G.R. Newkome and W.W. Paudler, *Contemporary Heterocyclic Chemistry. Syntheses, Reactions, and Applications*, p. 13, J. Wiley-Interscience, New York, NY, 1982.
35. P. Debye, *Polar Molecules* (Eng. Trans.), Chemical Catalog Co., New York, NY, 1929.
36. I.F. Halverstadt and W.D. Kumler, *J. Am. Chem. Soc.*, 64 (1942) 2988.
37. G. Hedestrand, *Z. Physik. Chem.*, B2 (1929) 428.
38. E.A. Guggenheim, *Trans. Faraday Soc.*, 45 (1949) 714.
39. J.W. Smith, *Trans. Faraday Soc.*, 46 (1950) 394.
40. K. Higasi, *Bull. Inst. Phys. Chem. Res. (Tokyo)*, 22 (1943) 805.
41. J.L. Imbach, R.A.Y. Jones, A. R. Katritzky, and R.J. Wyatt, *J. Chem. Soc., B*, (1967) 499.
42. O. Exner, *Collect. Czech. Chem. Commun.*, 46 (1981) 1002.
43. J.E. Wollrab, *Rotational Spectra and Molecular Structure*, Academic Press, New York, NY, 1967.
44. E.B. Wilson, *Quart. Rev. Chem. Soc.*, 1 (1972) 293.
45. J. S. Muentner, *J. Chem. Phys.*, 48 (1968) 4544.
46. A.A. Shapkin, L.N. Gunderova, N.N. Magdesieva, and N.M. Pozdeev, *Zh. Strukt. Khim.*, 14 (1973) 1037.
47. W.F. Hassell, M.D. Magee, S.W. Tucker, and S. Walker, *Tetrahedron*, 20 (1964) 2137.
48. J. Crossley, A. Holt, and S. Walker, *Tetrahedron*, 21 (1965) 3141.
49. M.D. Magee and S. Walker, *Trans. Faraday Soc.*, 62 (1966) 3093.
50. R.M. Redheffer, in: C.G. Montgomery (ed.), *Technique of Microwave Measurements, MIT Radiation Laboratory Series*, Vol. 11, Chap. 10, McGraw-Hill, New York, NY, 1947.
51. D.H. Whiffen, *Quart. Rev. (London)*, 4 (1950) 131.
52. S. Walker and H. Straw, *Atomic, Microwave, and Radiofrequency Spectroscopy*, Chapman and Hall, London, 1961.
53. K.F. Smith, *Molecular Beams*, J. Wiley, New York, NY, 1955.
54. N.F. Ramsey, *Molecular Beams*, Clarendon Press, Oxford, 1956.
55. N.G. Bakhshiev, *Opt. Spektrosk.*, 10 (1961) 717; *Opt. Spectry.*, 10 (1961) 379.
56. N.G. Bakhshiev, *Opt. Spektrosk.*, 13 (1962) 192.
57. L.M. Kutsyna and L.U. Voevoda, *Opt. Spektrosk.*, 18 (1965) 520.
58. A. Streitwieser, Jr., *Molecular Orbital Theory for Organic Chemists*, Chap. 6, p. 139, J. Wiley, New York, NY, 1961.
59. E. Heilbronner and H. Bock, *Das HMO-Modell und seine Anwendung*, p. 262, Verlag Chemie, Weinheim/Bergstr., 1968.

60. R.S. Berry, S.A. Rice, and J. Ross, *Physical Chemistry*, p. 155, J. Wiley, New York, NY, 1980.
61. P.W. Atkins, *Quanta: A Handbook of Concepts*, 2nd ed., Oxford University Press, Oxford, 1991.
62. P.W. Atkins, *Physical Chemistry*, 5th ed., Oxford University Press, Oxford, 1994.
63. P.W. Atkins and R.S. Friedman, *Molecular Quantum Mechanics*, 3rd ed., Oxford University Press, Oxford, 1997.
64. C.H. Townes, G.C. Dousmanis, R.L. White, and R.F. Schwarz, *Discussions Faraday Soc.*, 19 (1955) 56.
65. B. Rosenblum, A.H. Nethercot, Jr., and C.H. Townes, *Phys. Rev.*, 109 (1958) 400.
66. C.A. Burns, *J. Chem. Phys.*, 30 (1959) 976.
67. W.H. Flygare, W. Hüttner, R.L. Shoemaker, and P.D. Foster, *J. Chem. Phys.*, 50 (1969) 1714.
68. G.W. Wheland and D.E. Mann, *J. Chem. Phys.*, 17 (1949) 264.
69. H. Kofod, L.E. Sutton, W.A. de Jong, P.E. Verkade, and B.M. Wepster, *Recl. Trav. Chim. Pays-Bas*, 71 (1952) 521.
70. A.D. Buckingham, B. Harris, and R.J.W. Le Fèvre, *J. Chem. Soc.*, (1953) 1626.
71. B. Lam, M.W. Schmidt, and K. Ruedenberg, *J. Phys. Chem.*, 89 (1985) 2221.
72. C.A. Coulson and H.C. Longuet-Higgins, *Rev. Sci. Instr.*, 85 (1947) 927.
73. A. Pullman, B. Pullman, and P. Rumpf, *Bull. Soc. Chim. Fr.*, (1948) 757.
74. C. Sándorfy, N.Q. Trinh, A. Laforgue, and R. Daudel, *J. Chim Phys.*, 46 (1949) 655.
75. R. Zahradnik and C. Párkányi, *Collect. Czech. Chem. Commun.*, 30 (1965) 355.
76. D.G. Leis and B.C. Curran, *J. Am. Chem. Soc.*, 67 (1945) 79.
77. A.R. Katritzky, E.W. Randall, and L.E. Sutton, *J. Chem. Soc.*, (1957) 1769.
78. A. Julg, *J. Chim. Phys.*, 52 (1955) 377.
79. P. François and A. Julg, *J. Chim. Phys.*, 57 (1960) 490.
80. T. Schaefer and W.G. Schneider, *Can. J. Chem.*, 41 (1963) 966.
81. L.E. Orgel, T.L. Cottrell, W. Dick, and L.E. Sutton, *Trans. Faraday Soc.*, 47 (1951) 113.
82. H. Berthod and A. Pullman, *J. Chim. Phys.*, 62 (1965) 942.
83. R. Hoffmann, *J. Chem. Phys.*, 39 (1963) 1397.
84. W. Adam and A. Grimison, *Theor. Chim. Acta*, 7 (1967) 342.
85. C.W.N. Cumper, *Tetrahedron*, 25 (1969) 3131.
86. R.S. Mulliken, *J. Chem. Phys.*, 3 (1935) 573.
87. C.A. Coulson, *Trans. Faraday Soc.*, 38 (1942) 433.
88. J.H. Gibbs, *J. Phys. Chem.*, 59 (1955) 644.
89. H.M. Hameka and A.M. Liquori, *Mol. Phys.*, 1 (1958) 9.
90. B.J. Lounsbury, *J. Phys. Chem.*, 67 (1963) 721.
91. G. Klopman, *Tetrahedron, Suppl. 2* (1963) 111.
92. K. Fukui, in: P.-O. Löwdin and B. Pullman (eds.), *Molecular Orbitals in Physics, Chemistry, and Biology; A Tribute to R.S. Mulliken*, p. 513, Academic Press, New York, NY, 1964.
93. N.D. Sokolov, *Usp. Khim.*, 36 (1967) 2195.
94. G. Del Re, *J. Chem. Soc.*, (1958) 4031.
95. G. Del Re and T. Yonezawa, *Biochim Biophys. Acta*, 75 (1963) 153.
96. R.D. Brown and M.L. Heffernan, *Aust. J. Chem.*, 12 (1959) 319, 330, 543, 554.

97. J.A. Pople and M. Gordon, *J. Am. Chem. Soc.*, 89 (1967) 4253.
98. R.D. Brown and F.R. Burden, *Theor. Chim. Acta*, 12 (1968) 95.
99. J.S. Yadav, P.C. Mishra, and D.K. Rai, *J. Mol. Struct.*, 13 (1972) 311.
100. J.A. Pople, D.P. Santry, and G.A. Segal, *J. Chem. Phys.*, 43 (1965) 129, 136.
101. J.A. Pople and G.A. Segal, *J. Chem. Phys.*, 44 (1966) 3289.
102. J.E. Bloor and D.L. Breen, *J. Phys. Chem.*, 72 (1968) 716.
103. A. Chung-Phillips, *J. Comput. Chem.*, 10 (1989) 17.
104. J.J.P. Stewart, *J. Comput. Chem.*, 10 (1989) 221.
105. P.R. Livotto and Y. Takahata, *An. Acad. Bras. Cienc.*, 61 (1989) 135; *Chem. Abstr.*, 112 (1990) 166360y.
106. A. Birenzvege, L.M. Stuedivan, G.R. Famini, and P.N. Krishnan, *J. Comput. Chem.*, 17 (1993) 33.
107. L.P. Bokacheva and S.G. Semenov, *Vestn. Leningr. Univ., Ser. 4: Fiz. Khim.*, (1988) 73; *Chem. Abstr.*, 110 (1989) 30606j.
108. A.R. Katritzky, M. Szafran, and J. Stevens, *J. Chem. Soc., Perkins Trans. 2*, (1989) 1507.
109. A.I. Raznoshinskii, S.N. Shcherbo, and V.I. Yuzhakov, *Zh. Fiz. Khim.*, 64 (1990) 1266.
110. K. Singh, D.K. Rai, and J.S. Yadav, *THEOCHEM*, 77 (1991) 103.
111. D.N. Govorun, V.D. Danchuk, Ya.R. Mishchuk, I.V. Kondratyuk, N.F. Radomskii, and N.V. Zheltovskii, *J. Mol. Struct.*, 267 (1992) 99.
112. J.G. Contreras and J. B. Alderete, *Mol. Eng.*, 2 (1992) 29.
113. J.G. Contreras and J.B. Alderete, *THEOCHEM*, 102 (1993) 283.
114. A.O. Alyoubi and R.H. Hilal, *Biophys. Chem.*, 55 (1995) 231.
115. G. Vermin, M. Meyer, L. Bouscasse, J. Metzger, and C. Párkányi, *J. Mol. Struct.*, 68 (1980) 209.
116. C. Párkányi, S.R. Oruganti, A.O. Abdelhamid, L. von Szentpály, B. Ngom, and J.J. Aaron, *J. Mol. Struct. (THEOCHEM)*, 135 (1986) 105.
117. C. Párkányi, A. Brehon, A. Couture, A. Lablache-Combier, and A. Pollet, *Heterocycles*, 24 (1986) 355.
118. J.J. Aaron, M.D. Gaye, C. Párkányi, N.S. Cho, and L. von Szentpály, *J. Mol. Struct.*, 156 (1987) 119.
119. C. Párkányi, C. Boniface, J.J. Aaron, and M. Buna, to be published.
120. J.J. Aaron, A. Tine, M.D. Gaye, C. Párkányi, C. Boniface, and T.W.N. Bieze, *Spectrochim. Acta*, 47A (1991) 419.
121. C. Párkányi, C. Boniface, J.J. Aaron, F. Meuguelati, J.S. Murray, P. Politzer, and K.S. Raghuveer, in: H. Keyzer, G.M. Eckert, I.S. Forrest, R.R. Gupta, F. Gutmann, and J. Molnár (eds.), *Thiazines and Structurally Related Compounds, Proc. Sixth Intl. Conf. Phenothiazines and Struct. Related Psychotropic Drugs*, Pasadena, CA, Sept. 11-14, 1990, p. 103, Krieger Publishing, Malabar, FL, 1992.
122. C. Párkányi, C. Boniface, J.J. Aaron, M.D. Gaye, K.S. Raghuveer, L. von Szentpály, and R. Ghosh, *Struct. Chem.*, 3 (1992) 277.
123. J.J. Aaron, M.D. Gaye, C. Párkányi, C. Boniface, T.W.N. Bieze, S.S. Atik, K.S. Raghuveer, L. von Szentpály, and R. Ghosh, *Pteridines*, 3 (1992) 153.
124. C. Párkányi, C. Boniface, J.J. Aaron, and M. Maafi, *Spectrochim. Acta*, 49A (1993) 1714.

125. C. Párkányi, M.S. Antonious, J.J. Aaron, M. Buna, A. Tine, and L. Cissé, *Spectrosc. Lett.*, 27 (1994) 439.
126. J.J. Aaron, M. Maafi, C. Párkányi, and C. Boniface, *Spectrochim. Acta*, 51A (1995) 603.
127. C. Párkányi, M.S. Antonious, J.J. Aaron, M. Maafi, O. Gil, C. Kersebet, and N. Motohashi, in: J. Barbe, H. Keyzer, and J.C. Soyfer (eds.), *Biological and Chemical Aspects of Thiazines and Analogs, Proc. Seventh Intl. Conf. Phenothiazines and Struct. Related Psychotropic Compds.*, Marseille, France, Aug. 29-Sept. 2, 1993, p. 177, Enlight Associates, San Gabriel, CA, 1995.
128. J.J. Aaron, M. Buna, C. Párkányi, M.S. Antonious, A. Tine, and L. Cissé, *J. Fluorescence*, 5 (1995) 337.
129. J.J. Aaron, M. Maafi, C. Kersebet, C. Párkányi, M.S. Antonious, and N. Motohashi, *Spectrosc. Lett.*, 28 (1995) 1111.
130. C. Párkányi, A. Adenier, and J.J. Aaron, in: E. Kohen and J.G. Hirschberg (eds.), *Analytical Use of Fluorescent Probes in Oncology*, p. 371, Plenum Press, New York, NY, 1996.
131. A. Adenier, J.J. Aaron, C. Párkányi, G. Deng, and M. Sallah, *Heterocycl. Commun.*, 2 (1996) 403.
132. J.J. Aaron, M. Maafi, C. Kersebet, C. Párkányi, M.S. Antonious, and N. Motohashi, *J. Photochem. Photobiol. A: Chemistry*, 101 (1996) 127.
133. C. Párkányi, M. Dakkouri, and J.J. Aaron, to be published.
134. J. Bandekar, *Spectrosc. Lett.*, 22 (1989) 173.
135. J.S. Kwiatkowski and J. Leszczyński, *THEOCHEM*, 67 (1990) 35.
136. J. Leszczyński, *Chem. Phys. Lett.*, 174 (1990) 347.
137. R. Czermiński, K. Szczepaniak, W.B. Person, and J.S. Kwiatkowski, *J. Mol. Struct.*, 237 (1990) 151.
138. J. Pranata, S.G. Wierschke, and W. L. Jorgensen, *J. Am. Chem. Soc.*, 113 (1991) 2810.
139. J. Leszczyński, *Int. J. Quantum. Chem., Quantum Biol. Symp.*, 18 (1991) 9.
140. J. Leszczyński, *J. Phys. Chem.*, 97 (1993) 3520.
141. G.H. Roehrig, N.A. Oyler, and L. Adamowicz, *Chem. Phys. Lett.*, 225 (1994) 265.
142. J.G. Contreras and J.B. Alderete, *Chem. Phys. Lett.*, 232 (1995) 61.
143. L. Adamowicz, *J. Phys. Chem.*, 99 (1995) 14285.
144. T.K. Ha, H.J. Keller, R. Gunde, and H.H. Gunthard, *J. Mol. Struct.*, 376 (1996) 375.
145. N.E. Kassimi, R.J. Doerksen, and A.J. Thakkar, *J. Phys. Chem.*, 99 (1995) 12790.
146. K.E. Laidig, P. Speers, and A. Streiwieser, *Can. J. Chem.*, 74 (1996) 1215.
147. I.S. Han, C.K. Kim, and H.J. Jung, *Theor. Chim. Acta*, 93 (1996) 199.
148. N.E. Kassimi, R.J. Doerksen, and A.J. Thakkar, *J. Phys. Chem.*, 100 (1996) 8752.
149. E.F. Archibong and A.J. Thakkar, *Mol. Phys.*, 81 (1994) 557.
150. C. Moberg, H. Adolfsson, K. Waernmark, P.O. Norrby, K.M. Marstokk, and H. Mollendal, *Chem.-Eur. J.*, 2 (1996) 516.
151. J. Wang and R.J. Boyd, *J. Phys. Chem.*, 100 (1996) 16141.
152. M. Orozco and F.J. Luque, *J. Comput. Chem.*, 11 (1990) 909.
153. C.E. Dykstra, S.Y. Liu, and D.J. Malik, *Adv. Chem. Phys.*, 75 (1989) 37.
154. W. Hüchel and C.M. Salinger, *Chem. Ber.*, 77 (1944) 810.
155. W.C. Schneider, *J. Am. Chem. Soc.*, 70 (1948) 627.
156. J.E. Del Bene, *J. Comput. Chem.*, 2 (1981) 251.

157. M. Kulakowska, M. Geller, B. Lesyng, and K.L. Wierzchowski, *Biochim. Biophys. Acta*, 361 (1974) 119.
158. P. Mauret and J.P. Fayet, *Compt. Rend. Acad. Sci. Paris, Sér. C*, 264 (1967) 2081.
159. C. Giessner-Prettre and A. Pullman, *Theor. Chim. Acta*, 9 (1968) 279.
160. A.R. Katritzky and M. Karelson, *J. Am. Chem. Soc.*, 113 (1991) 1561.
161. W.C. Schneider and I.F. Halverstadt, *J. Am. Chem. Soc.*, 70 (1948) 2626.
162. E.D. Bergmann and H. Weiler-Feilchenfeld, in: E.D. Bergmann and B. Pullman (eds.), *The Purines: Theory and Experiment, Proc. Jerusalem Symp. Chem. Biochem.*, Vol. 4, p. 21, Israel Acad. Sci. Human., Jerusalem, 1972.
163. B. Pullman, H. Berthod, F. Bergmann, Z. Neiman, H. Weiler-Feilchenfeld, and E.D. Bergmann, *Tetrahedron*, 26 (1970) 1483.
164. H. De Voe and I. Tinoco, Jr., *J. Mol. Biol.*, 4 (1962) 500.
165. B. Mély and A. Pullman, *Theor. Chim. Acta*, 13 (1969) 278.
166. H. Miyazaki, *Osaka Daigaku Zasshi*, 11 (1959) 4306.
167. B. Koutek, *Collect. Czech. Chem. Commun.*, 43 (1978) 2368.
168. A. Kawski and L. Bilot, *Acta Phys. Polon.*, 26 (1964) 41.
169. A. Kawski, *Acta Phys. Polon.*, 29 (1966) 507.
170. A. Chamma and P.C. Viallet, *Compt. Rend. Acad. Sci. Paris, Sér. C*, 270 (1970) 1901.
171. N.G. Bakhshiev, *Opt. Spektrosk.*, 16 (1964) 821; *Opt. Spectry.*, 16 (1964) 446.
172. E.G. McRae, *J. Phys. Chem.*, 61 (1957) 562.
173. P. Suppan, *J. Chem. Soc. A*, (1968) 3125.
174. P. Suppan and C. Tsiamis, *Spectrochim. Acta*, 36A (1980) 971.
175. P. Suppan, *Chem. Phys. Lett.*, 94 (1983) 272.
176. A.R. Katritzky, M.C. Zerner, and M.M. Karelson, *J. Am. Chem. Soc.*, 108 (1986) 7213.
177. M. Karelson, A.R. Katritzky, M. Szafran, and M.C. Zerner, *J. Chem. Soc., Perkin Trans. 2*, (1990) 195.
178. M. Karelson, T. Tamm, A.R. Katritzky, M. Szafran, and M.C. Zerner, *Int. J. Quantum Chem.*, 37 (1990) 1.
179. M. Karelson, A.R. Katritzky, and M.C. Zerner, *J. Org. Chem.*, 56 (1991) 134.
180. Yu.G. Siretskii, A.L. Kirillov, and N.G. Bakhshiev, *Dokl. Akad. Nauk SSSR*, 275 (1984) 1463; *Doklady Fiz. Khim. (Eng. Trans.)*, 275 (1984) 369.
181. L.S. Prabhumirashi and S.S. Kunte, *Indian J. Chem.*, 29A (1990) 215.
182. P. Suppan, *J. Photochem. Photobiol. A: Chemistry*, 50 (1990) 293.
183. N. Ghoneim and P. Suppan, *Spectrochim. Acta*, 51A (1995) 1043.
184. N.H. Ayachit and J. Tonannavar, *Spectrochim. Acta*, 47A (1991) 1637.
185. V.Z. Kurbako, V.V. Drboglav, and N.I. Garbuz, *Zh. Prikl. Spektrosk.*, 51 (1989) 851; *Chem Abstr.*, 112 (1990) 89353u.
186. T. Abe, *Bull. Chem. Soc. Jpn.*, 54 (1981) 327.
187. T. Abe and I. Iweibo, *Bull. Chem. Soc. Jpn.*, 58 (1985) 3415.
188. T. Abe, *Bull. Chem. Soc. Jpn.*, 61 (1988) 3797.
189. T. Abe, *Bull. Chem. Soc. Jpn.*, 64 (1991) 3224.
190. A. Imamura, H. Fujita, and C. Nagata, *Bull. Chem. Soc. Jpn.*, 40 (1967) 21.
191. E.D. Bergmann and H. Weiler-Feilchenfeld, in: J. Duchesne (ed.), *Physico-Chemical Properties of Nucleic Acids*, p. 1, Academic Press, London, 1973.

192. F.A. Savin, Yu.V. Morozov, A.V. Borodavkin, V.O. Chekhov, E.I. Budowsky, and N.A. Simukova, *Int. J. Quantum. Chem.*, 16 (1979) 825.
193. S. K. Srivastava and P.C. Mishra, *J. Mol. Struct.*, 65 (1980) 199.

New developments in the analysis of vibrational spectra On the use of adiabatic internal vibrational modes

Dieter Cremer*, J. Andreas Larsson, and Elfi Kraka

Department of Theoretical Chemistry, University of Göteborg,
Kemigården 3, S-41296 Göteborg, Sweden

1. INTRODUCTION

Vibrational spectroscopy is an often used tool to identify and characterize a molecule with the help of its vibrational modes. Depending on its geometry, conformation, and electronic structure, each molecule has typical vibrational spectra which are measured with the help of infrared or Raman spectroscopy [1-9]. For example, an infrared band at 1700 cm^{-1} is typical for a carbonyl stretching frequency or one at $700\text{ to }800\text{ cm}^{-1}$ for a CCl bond stretching frequency. In this way, one can draw conclusions from the measured vibrational spectra with regard to the structure of a compound. One can also calculate vibrational frequencies and force constants in the harmonic approximation and these values are often used for the analyses of measured vibrational spectra [10-12]. This is done to identify and verify the structure of molecules generated in the experiment, which will be of particular usefulness if limitations in the experiment do not make any other experimental investigation possible. For example, molecules trapped at low temperatures in a matrix are elegantly investigated by reproducing the measured infrared spectrum by appropriate calculations. In this way a number of labile species have been identified [13-17].

The amount of information contained in a measured vibrational spectrum is exploited to some, but not full extent. For example, vibrational spectra are never used to characterize all bonds of the molecule and to describe its electronic structure and charge distribution in detail. Of course, aspects of such investigations can be found off and on in the literature, however, both quantum chemists and spectroscopists fail to use vibrational spectra on a routine basis as a source of information on bond properties, bond-bond interactions, bond delocalization or other electronic features. Therefore, it is correct to say that the information contained in the vibrational spectra of a molecule is not fully utilized. This has to do with the fact that the analysis of vibrational spectra is always carried out in a way that is far from chemical thinking. The basic instrument in this respect is the normal mode analysis (NMA), which describes the displacements of the atomic nuclei during a molecular vibration in terms of delocalized normal modes [1-6].

A normal mode is composed of the movement of many or even all atoms of a molecule, which is difficult to visualize. Because of this chemists try to simplify the description of a normal mode by focusing on the motions of just few atoms that seem to dominate the normal mode. This requires an appropriate measure that determines which atomic motion is dominant. Attempts in this direction have been made and it is common practice now to associate certain normal modes of a molecule with chemically interesting fragment modes even though this simplification is usually not justified. Hence the basic problem of vibrational spectroscopy is the transformation of the delocalized normal modes, which are difficult to visualize, to chemically more appealing localized modes that can be associated with particular fragments of a molecule.

In this article, we present a new way of analyzing calculated vibrational spectra in terms of internal vibrational modes associated with the internal coordinates used to describe geometry and conformation of a molecule. The internal modes will be determined by solving the Euler-Lagrange equations for molecular fragments ϕ_n being characterized by internal coordinates q_n . An internal mode will be localized in a molecular fragment by describing the rest of the molecule as a collection of massless points that just define molecular geometry. Alternatively, one can consider the new fragment motions as motions that are obtained after relaxing all parts of the vibrating molecule but the fragment under consideration. Because of this property, the internal modes will be called adiabatic internal modes. Once the adiabatic mode vectors are known, adiabatic force constants k_a , adiabatic frequencies ω_a , and adiabatic masses m_a (corresponding to $1/G_{nn}$ of Wilson's G matrix) will be defined. The adiabatic internal modes are independent of the set of internal coordinates used to describe molecular geometry, comply with the symmetry of the molecule, and lead to a clear separation of mass and electronic effects in the vibrational modes of the molecule.

The new modes are perfectly suited to analyze the vibrational spectra of a molecule in terms of internal coordinate modes, to correlate the vibrational spectra of different molecules, and to extract chemically useful information directly from vibrational spectra. It will be shown that adiabatic stretching frequencies and force constants correlate with the corresponding bond lengths and that this can be used to extend Badger's rule from diatomic to polyatomic molecules. The intensities of adiabatic stretching modes lead to effective atomic charges and bond dipole moments. Generalized adiabatic modes will be defined for reacting molecules located somewhere along the reaction path. They will be used to analyze the direction and curvature of the reaction path and, by this, to obtain a better insight into reaction mechanism and reaction dynamics.

2. THE CONCEPT OF LOCALIZED INTERNAL VIBRATIONAL MODES

Chemists have learned to understand geometry and conformation of a molecule in terms of (localized) internal coordinates such as bond lengths, bond angles, and torsional angles. Therefore, it would be chemically useful to discuss

vibrational spectra in terms of bond stretching modes, angle bending modes or torsional modes, which are the localized counterparts of the delocalized normal modes. Each localized mode should be associated with an internal coordinate that describes a molecular fragment of interest. If the normal modes obtained in the NMA of vibrational spectroscopy could be transformed into these internal vibrational modes, then infrared and Raman spectroscopy could provide for each bond a characteristic stretching mode frequency ω_n and a stretching mode force constant k_n that could be used to describe the properties of the bonds of a molecule and that would complement information obtained from direct bond length measurements. Moreover, by determining the bending mode frequencies and force constants of a molecule a direct insight into bond-bond interactions would be provided by vibrational spectroscopy. One could systematically investigate all two-atom, three-atom, four-atom, etc. units of a molecule and, in this way, obtain a detailed description on bonding and electronic structure of a molecule by just using data obtained from vibrational spectroscopy. In this way, vibrational spectroscopy would become a major source of information on molecules, which it is not at the present time.

Of course, one could say that this information can be gained directly from a molecular geometry determination that provides all bond lengths, bond angles, and torsional angles. However, there are several reasons why a description of bonding and other electronic features of a molecule with the help of its vibrational modes should be of advantage. Apart from the fact that it is often easier to measure a vibrational spectrum than to carry out a geometry determination by microwave, electron diffraction or X-ray methods, there is also the reason that the information derived from the vibrational modes has a different quality than that derived from measured bond lengths, bond angles, and dihedral angles. The value of a bond length depends primarily on the accumulation of electron density in the bonding region. It is not very sensitive with regard to the environment of a bond, i.e. the nature of the atoms and groups attached to the bond in question or the electronic characteristics of neighboring bonds. A bond stretching motion, on the other hand, is clearly influenced by the bond environment and, therefore, the internal stretching frequency should reflect not only the amount of electron density accumulated in the bonding region, but also the bulk and electronic nature of the atoms and groups attached to the bond in question. Hence, the internal stretching frequencies and force constants should be a better measure of the strength of the bonds of a molecule than the measured bond lengths.

The information on the various bonds in a molecule should be hidden somewhere in infrared and Raman spectra and it is only a question how to unravel it from experimentally obtained or calculated vibrational spectra. To obtain this information, one has to specify exactly what kind of internal vibrational mode is needed [18-23]:

An internal mode should be fully characterized by only one internal coordinate q_n . The internal coordinate q_n should be the parameter that leads the internal mode and, therefore, it can be called *the leading parameter* of the internal mode [18]. The internal mode should be localized in the fragment ϕ_n of

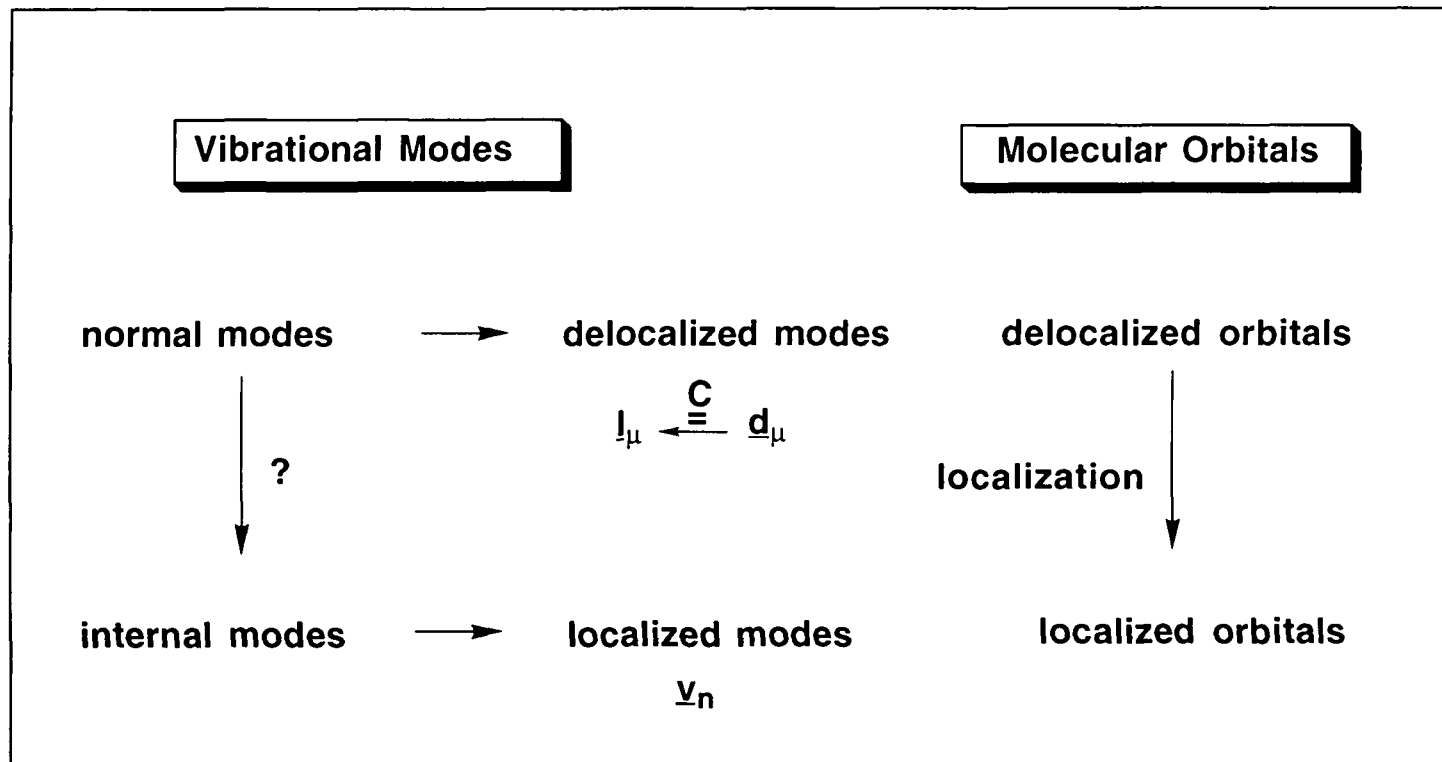


Figure 1. Analogy between delocalized/localized vibrational modes and delocalized/localized molecular orbitals.

the molecule described by the internal coordinate. Of course, this does not necessarily imply that all other atoms are at rest when a fragment is vibrating. On the contrary, to keep all other internal coordinates at their equilibrium values when a particular fragment ϕ_n vibrates, the rest of the molecule has to move with the same frequency than the molecular fragment ϕ_n . However, this is not a contradiction since the motion is still localized in the bond under consideration.

Normal modes are orthogonal to each other and one advantage of this property is that the force constant matrix associated with the normal modes can be given in a diagonal form. Internal modes will no longer be orthogonal, which means that the force constant matrix has no longer a diagonal form. In a way, if the vibrational modes of a molecule are expressed in a form which is closer to chemical thinking, some of these mathematical properties are lost. However, once internal vibrational modes are defined it should be possible to convert back to normal modes and express the latter in terms of internal modes so that it becomes clear to which extent it is justified to interpret normal modes as fragment modes.

An analogy to molecular orbital (MO) theory may help to clarify further what is needed. Chemists prefer to discuss chemical problems in terms of localized MOs rather than in terms of (canonical) delocalized MOs resulting from Hartree-Fock (HF) based quantum chemical calculations. The localized MOs are obtained from the delocalized ones by a transformation ("localization"), which in most cases yields MOs directly related to the bonds of a molecule. The same should be true with regard to localized modes associated with a particular internal coordinate q . The question is only: How can we transform from delocalized normal modes to localized internal modes? To answer this question we will first summarize the basic theory of vibrational spectroscopy.

3. THE BASIC EQUATIONS OF VIBRATIONAL SPECTROSCOPY

The potential energy function $V(\mathbf{x})$ of a molecule with K atoms describes the increase in energy upon a displacement of the atomic nuclei from their equilibrium positions by a Cartesian displacement vector $\mathbf{x} = (x_1, y_1, z_1, \dots, x_K, y_K, z_K)^T$. Expanding the potential energy in a Taylor series and neglecting all higher order terms one obtains for $V(\mathbf{x})$ expression (1) [1-6]:

$$V(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{f} \mathbf{x} \quad (1)$$

where \mathbf{f} is the force constant (Hessian) matrix expressed in Cartesian coordinates at the equilibrium geometry $\mathbf{x}_0 = \mathbf{0}$ because \mathbf{x} represents the displacements from the equilibrium geometry. The kinetic energy $T(\dot{\mathbf{x}})$ of a vibrating molecule is given by Eq. (2)

$$T(\dot{\mathbf{x}}) = \frac{1}{2} \dot{\mathbf{x}}^T \mathbf{M} \dot{\mathbf{x}} \quad (2)$$

where \mathbf{M} is the mass matrix. With Eqs. (1) and (2), the Lagrangian \mathcal{L} of the molecule becomes

$$\mathcal{L}(\mathbf{x}, \dot{\mathbf{x}}) = T(\dot{\mathbf{x}}) - V(\mathbf{x}) \quad (3)$$

and the dynamics of the nuclei of the molecule can be determined by solving the Euler-Lagrange equations (4)

$$\frac{d}{dt} \frac{\partial \mathcal{L}(\mathbf{x}, \dot{\mathbf{x}})}{\partial \dot{x}_i} - \frac{\partial \mathcal{L}(\mathbf{x}, \dot{\mathbf{x}})}{\partial x_i} = 0, \quad i = 1, \dots, 3K \quad (4)$$

The solutions of (4) take the form of (5) [1-6]:

$$\mathbf{x} = \mathbf{I}_\mu Q_\mu \quad (5)$$

where Q_μ is a normal coordinate, which oscillates with the frequency ω_μ according to (6)

$$Q_\mu(t) = a \cos(\omega_\mu t) + b \sin(\omega_\mu t) \quad (6)$$

Inserting (5) and (6) into (4) leads to the basic equation of vibrational spectroscopy [1-6]:

$$\mathbf{f} \mathbf{I}_\mu = \omega_\mu^2 \mathbf{M} \mathbf{I}_\mu, \quad \mu = 1, \dots, N_{\text{vib}} \quad (7)$$

which is used to calculate the $N_{\text{vib}} = 3K - L$ normal mode frequencies of a K -atomic molecule where $L = 5$ or 6 gives the number of zero eigenvalues in (7) resulting from translations and rotations of the molecule.

In Eq. (7), the normal mode vectors are expressed in Cartesian coordinate space. However, it is much more useful to express the motions of a molecule in internal coordinate space using N internal coordinates collected in a column vector $\mathbf{q} = (q_1, \dots, q_N)^+$. Changes in bond lengths, bond angles, and dihedral angles can be used as convenient internal coordinates. To specify the positions of all nuclei in Cartesian space, an additional set of L external coordinates has to be given. The external coordinates are arranged in a column vector $\mathbf{e} = (e_1, \dots, e_L)^+$. Transformation from internal to Cartesian coordinates is given by Eq.(8) [24]:

$$x_i = \sum_{m=1}^N C_{im} q_m + \sum_{\alpha=1}^L C_{0,i\alpha} e_\alpha \quad (8)$$

where C_{im} is an element of the $(3K, N)$ -rectangular matrix \mathbf{C} with column vectors \mathbf{c}_m :

$$\mathbf{C} = \mathbf{M}^{-1}\mathbf{B}^+\mathbf{G}^{-1} \quad (9)$$

and matrix \mathbf{C}_0 has been defined by Neto [24]. Wilson's \mathbf{G} -matrix [1] is given by

$$\mathbf{G} = \mathbf{B}\mathbf{M}^{-1}\mathbf{B}^+ \quad (10)$$

and the elements of the \mathbf{B} matrix are defined by

$$q_n = \sum_{i=1}^{3K} B_{ni} x_i \quad (11)$$

$$B_{ni} = \left(\frac{\partial q_n(\mathbf{x})}{\partial x_i} \right)_{\mathbf{x}=\mathbf{x}_0} \quad (12)$$

with \mathbf{x}_0 denoting the equilibrium geometry of the molecule.

Note that

$$\mathbf{B}\mathbf{C} = \mathbf{B}\mathbf{M}^{-1}\mathbf{B}^+\mathbf{G}^{-1} = \mathbf{G}\mathbf{G}^{-1} = \mathbf{I} \quad (13)$$

and

$$\mathbf{B}\mathbf{C}_0 = \mathbf{0}. \quad (14)$$

Generally, internal and external coordinates couple in the kinetic energy term, however they can be decoupled by inserting (8) into (3) and using (14), which leads to

$$\mathcal{L} = \mathcal{L}_{\text{int}} + \mathcal{L}_{\text{ext}} \quad (15)$$

where \mathcal{L}_{ext} depends on external coordinates and, accordingly, is not relevant for the vibrational problem. The quantity \mathcal{L}_{int} determines the time dependence of the internal coordinates and is given by

$$\mathcal{L}_{\text{int}}(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2} \dot{\mathbf{q}}^+ \mathbf{G}^{-1} \dot{\mathbf{q}} - \frac{1}{2} \mathbf{q}^+ \mathbf{F} \mathbf{q} \quad (16)$$

where \mathbf{F} is the $N \times N$ -dimensional force constant matrix expressed in internal coordinate space :

$$F_{nm} = \mathbf{c}_n^+ \mathbf{f} \mathbf{c}_m \quad (17)$$

Solving the Euler-Lagrange equation (18)

$$p_m = \frac{\partial \mathcal{L}_{\text{int}}(\mathbf{q}, \dot{\mathbf{q}})}{\partial \dot{q}_m} \quad \left. \vphantom{p_m} \right\} m = 1, \dots, N \quad (18a)$$

$$\frac{d}{dt} p_m = \frac{\partial \mathcal{L}_{\text{int}}(\mathbf{q}, \dot{\mathbf{q}})}{\partial q_m} \quad (18b)$$

(p_m is the generalized momentum) leads to Wilson's GF formalism for determining vibrational frequencies ω_μ [1]:

$$\mathbf{F} \mathbf{d}_\mu = \omega_\mu^2 \mathbf{G}^{-1} \mathbf{d}_\mu. \quad (19)$$

$$\omega_\mu^2 = \frac{\mathbf{d}_\mu^* \mathbf{F} \mathbf{d}_\mu}{\mathbf{d}_\mu^* \mathbf{G}^{-1} \mathbf{d}_\mu}. \quad (20)$$

Vector \mathbf{d}_μ represents the normal mode μ in internal coordinate space. It can be transformed to Cartesian coordinate space according to Eq. (21):

$$\mathbf{l}_\mu = \mathbf{C} \mathbf{d}_\mu \quad (21)$$

4. PREVIOUS ATTEMPTS OF DEFINING INTERNAL VIBRATIONAL MODES

The column vectors of \mathbf{C} , \mathbf{c}_n , can be chosen to represent the internal displacement vectors \mathbf{v}_n :

$$\mathbf{v}_n = \mathbf{c}_n \quad (22)$$

for a given internal coordinate q_n and $n = 1, \dots, N_{\text{vib}}$.

The "c-vectors" are implicitly used when expressing normal vibrational modes in terms of internal coordinates or when applying the potential energy distribution (PED) analysis to describe vibrational modes [25-27]. However, they have never been used explicitly to define internal modes of a molecule in the sense of Eq. (22). Since c-vectors are associated with internal coordinates q_n , and each of the latter describes a molecular fragment ϕ_n , they seem to be the natural choice for internal modes. However, it has been shown that $\mathbf{v}_n = \mathbf{c}_n$ is not a satisfactory choice of an internal vibration [19].

One often assumes that certain normal modes and their associated normal mode frequencies represent internal modes and internal mode frequencies (e.g., a normal mode frequency of 1700 cm^{-1} of a ketone as the C=O stretching mode frequency) [1-9]. If a normal mode vector \mathbf{l}_μ is largely localized in the molecular fragment ϕ_n , then the normal mode frequency ω_μ will be similar to the characteristic fragment frequency $\omega(\phi_n) = \omega_n$. It is one of the major goals of vibrational spectroscopy to determine fragment frequencies ω_n , which can be used to identify functional groups in a molecule to be investigated by vibrational

spectroscopy [1-9]. The existence of such frequencies simply results from the fact that functional groups largely retain their properties within different molecular environments. This, in turn, indicates that bonding and electron density distribution of a functional group are largely unaffected by the rest of the molecule and that group characteristic parameters such as internal mode frequency and internal mode force constant are appropriate to describe bonding and electron density distribution of a particular group.

However, using Eq. (21) it is easy to show that such an assumption is strictly valid only for the case where

$$(\mathbf{d}_\mu)_n = \delta_{n\mu} \quad (23)$$

($\delta_{n\mu}$: Kronecker delta) since this leads to

$$\mathbf{l}_\mu = \mathbf{c}_n \quad (24)$$

where it is assumed that $\mu = n$. However, even if displacements along vectors \mathbf{c}_n and \mathbf{c}_m do not couple thus leading to a diagonal \mathbf{F} matrix with $F_{nm} = 0$ (see Eq. 17, no electronic coupling), there is always mass coupling between the \mathbf{c} -vectors because the \mathbf{G} matrix is non-diagonal, which according to Eq. (19) leads to $d_{n\mu} \neq \delta_{n\mu}$ and $\mathbf{l}_\mu \neq \mathbf{c}_n$. Nevertheless, most vibrational spectroscopists will assume a more diagonal character of the \mathbf{G} matrix if there is a large mass difference between the atoms participating in the molecular motions. In some way, the assumptions made in Eqs. (23) and (24) provide the only basis for an experimentalist to discuss measured frequencies in terms of internal mode frequencies.

Hence, both the choice of \mathbf{c} -vectors as internal mode vectors \mathbf{v}_n and the typical assumption $\mathbf{l}_\mu = \mathbf{c}_n$ are not suited to provide an analysis of vibrational spectra in terms of internal modes [18,19]. Therefore, in the next section we will discuss a different approach that is based on a physically reasonable definition of internal modes.

5. DEFINITION OF ADIABATIC INTERNAL MODES

To obtain reasonable internal modes one has to consider that mass coupling prevents the vibrational modes to be localized in a particular molecular fragment. Hence, one has to eliminate mass coupling by an appropriate redefinition of the Euler-Lagrange equations. This is done by simply assuming that in Eqs. (4) all masses but the ones which belong to the atoms of fragment ϕ_n are zero [18]. With this assumption, the equations of motion (4) will lead to a pure internal vibration of fragment ϕ_n . Such an internal vibration expressed in Cartesian coordinate space is of little use for a chemist, who prefers to think in terms of internal coordinates rather than Cartesian coordinates. However, since

$m_i = 0$ ($i \in \phi_m$ with $m \neq n$) implies that the associated generalized momentum p_i is also equal to zero, one can extend the assumption that all atoms not belonging to ϕ_n are massless points just describing the molecular geometry and apply it to internal parameters by assuming that all internal parameters q_m ($m \neq n$) are associated with the generalized momentum $p_m = 0$. With this assumption, the Euler-Lagrange equations (18) take the form of (25) and (26) [18]:

$$p_n = \frac{\partial \mathcal{L}(q, \dot{q})}{\partial \dot{q}_n} \neq 0 \tag{25a}$$

$$p_m = \frac{\partial \mathcal{L}(q, \dot{q})}{\partial \dot{q}_m} = 0 \quad \forall m, m \neq n \tag{25b}$$

$$\dot{p}_n = \frac{\partial V}{\partial q_n} \tag{26a}$$

$$\dot{p}_m = \frac{\partial V}{\partial q_m} = 0 \quad m \neq n \tag{26b}$$

Eqs. (26) can be solved by adding Eq. (27):

$$\dot{p}_n = \lambda \tag{27a}$$

$$\lambda = \frac{\partial V}{\partial q_n} \tag{27b}$$

Eqs. (26b) and (27b) are used to express all internal coordinates q as functions of λ

$$\begin{aligned} q_i &= q_i(\lambda) \\ \dots\dots\dots & \\ q_N &= q_N(\lambda) \end{aligned} \tag{28}$$

Eq. (28) determines the form of internal vibrations v_n because it defines one-dimensional subspaces within the full configuration space. The motion in an one-dimensional subspace can be described by vector v_n , which can be found by linearization (e.g. *via* a Taylor expansion at point $\lambda=0$) of Eq. (28). If needed, the time dependence of λ can be found using generalized momenta

$$p_n = p_n(q, \dot{q}) = p_n(\lambda, \dot{\lambda}) \tag{29}$$

in connection with Eqs. (27a) and (28). In this way, one obtains an internal vibration $v_n = a_n$ for parameter q_n associated with fragment ϕ_n .

A set of equations similar to (27) can be obtained by applying a completely different approach [18]. One can displace parameter q_n from its equilibrium value ($q_n = 0$), keep it frozen and equal to a constant q_n^* . At the same time, all other parameters q_m can relax until the molecular energy attains its minimum. Hence, parameter q_n^* leads the corresponding motion as described by Eq. (30)

$$\mathbf{x}^* = \mathbf{v}_n q_n^* \quad (30)$$

(*leading parameter principle* [18]). For obvious reasons, one can call the vibrations generated by q_n^* as adiabatic vibrations defined by (31):

$$V(\mathbf{q}) = \min \quad (31a)$$

$$q_n = \text{const} = q_n^* \quad (31b)$$

Eq. (31) can be easily solved using the method of Lagrange multipliers:

$$\frac{\partial}{\partial q_m} [V(\mathbf{q}) - \lambda(q_n - q_n^*)] = 0 \quad m = 1, \dots, N \quad (32)$$

which leads to Eqs. (33)

$$\lambda = \frac{\partial V}{\partial q_n} \quad (33a)$$

$$0 = \frac{\partial V}{\partial q_m} \quad \forall m, m \neq n \quad (33b)$$

which are identical with Eqs. (27b) and (26b). Hence, the approximation based on "massless internal parameters q_m " is equivalent to the adiabatic approximation.

In quantum chemical calculations, the vibrational problem is normally described in the harmonic approximation. Assuming that the vibrational problem has been solved, potential energy and each internal parameter q_n can be expressed as function of N_{vib} normal mode coordinates Q_μ , [1-6]

$$V(Q) = \frac{1}{2} \sum_{\mu=1}^{N_{\text{vib}}} k_\mu Q_\mu^2 \quad (34)$$

$$q_n(Q) = \sum_{\mu=1}^{N_{\text{vib}}} D_{n\mu} Q_\mu \quad (35)$$

where matrix \mathbf{D} collects in its columns the normal mode vectors \mathbf{d}_μ expressed in internal coordinate space (compare with Eqs. 19 and 21). Inserting (34) into (31) and using the method of Lagrange multipliers, one obtains

$$\frac{\partial}{\partial Q_\mu} [V(Q) - \lambda(q_n(Q) - q_n^*)] = 0 \quad (36)$$

and

$$Q_\mu^{(n)} = \frac{D_{n\mu}}{k_\mu} \lambda. \quad (37)$$

The superscript n denotes the solution for internal parameter q_n where

$$q_n(Q) = q_n^* \quad (38)$$

as described above. Using equations (35), (37), and (38), λ can be found as function of q_n^*

$$\lambda = \frac{1}{\sum_{\mu=1}^{N_{nb}} \frac{D_{n\mu}^2}{k_\mu}} q_n^* \quad (39)$$

Inserting Eq. (39) into Eq. (37) leads to the normal coordinates as a function of q_n^*

$$Q_\mu^{(n)} = Q_{\mu n}^0 q_n^*, \quad (40)$$

where $Q_{\mu n}^0$ is a constant defined as

$$Q_{\mu n}^0 = \frac{\frac{D_{n\mu}}{k_\mu}}{\sum_{\nu=1}^{N_{nb}} \frac{D_{n\nu}^2}{k_\nu}}. \quad (41)$$

According to Eq. (40), any change in parameter q_n^* leads to a movement of all normal coordinates along the adiabatic vector \mathbf{a}_n , the components of which in normal coordinate space are given by

$$(\mathbf{a}_n)_\mu = Q_{\mu n}^0. \quad (42)$$

With Eq. (42) it is straightforward to transform adiabatic vectors into the space of Cartesian displacements:

$$(\mathbf{a}_n)_i = \sum_{\mu=1}^{N_{nb}} l_{i\mu} (\mathbf{a}_n)_\mu \quad i = 1, \dots, 3K \quad (43)$$

where $l_{i\mu}$ is a component of the normal mode l_μ defined in Eq. (21).

6. DEFINITION OF ADIABATIC INTERNAL FORCE CONSTANT, MASS, AND FREQUENCY

Once vector \mathbf{v}_n that determines the movement of the molecule under the influence of parameter q_n^* is known, one can define a force constant, that corresponds to such a motion, by inserting (30) into the expression for the potential energy of the molecule in the harmonic approximation:

$$V(q_n^*) = \frac{1}{2} k_n (q_n^*)^2 \quad (44)$$

where the internal force constant k_n is given by

$$k_n = \mathbf{v}_n^+ \mathbf{f} \mathbf{v}_n \quad (45)$$

It has been shown [18] that defining an internal mass M_n associated with the internal vibration \mathbf{v}_n by

$$M_n = \mathbf{v}_n^+ \mathbf{M} \mathbf{v}_n \quad (46)$$

which implies an characteristic fragment frequency Ω_n :

$$\Omega_n^2 = \frac{k_n}{M_n} \quad (47)$$

is not a useful choice since all masses of the molecule contribute to the mass M_n , which enters into the definition of Ω_n . In this way, the internal frequency Ω_n becomes sensitive to the environment of molecular fragment ϕ_n . This can lead to nonphysical shifts of internal frequencies as was documented in the literature [19].

Therefore, one has to proceed in a different way to find a typical mass m_n that opposes any change in the internal parameter q_n . In this connection, two conditions should be fulfilled. First, the mass m_n should be extractable from the functional form of the internal coordinate q_n . Secondly, m_n should be directly connected to the vibrational motion v_n caused by a change in q_n .

To fulfil these two conditions, one has to ask how the atoms of the molecule have to move so that the kinetic energy adopts a minimum and the generalized velocity \dot{q}_n becomes identical with \dot{q}_n^* , i.e. the system fulfils Eqs. (48) and (49):

$$T(\dot{\mathbf{x}}) = \frac{1}{2} \dot{\mathbf{x}}^* \mathbf{M} \dot{\mathbf{x}} = \min \quad (48)$$

$$\mathbf{b}_n^* \dot{\mathbf{x}} = \mathbf{b}_n^* v_n \dot{q}_n^* \quad (49)$$

where vector \mathbf{b}_n corresponds to the n th column of the \mathbf{B} matrix [1-6] and $\mathbf{b}_n^* v_n \dot{q}_n^*$ is the generalized velocity of internal coordinate q_n when the system moves according to Eq. (30). Using the Lagrange multiplier λ and combining (48) and (49), one obtains

$$\frac{\partial}{\partial \dot{x}_i} \left(\frac{1}{2} \dot{\mathbf{x}}^* \mathbf{M} \dot{\mathbf{x}} - \lambda (\mathbf{b}_n^* \dot{\mathbf{x}} - \mathbf{b}_n^* v_n \dot{q}_n^*) \right) = 0 \quad (50)$$

and

$$\dot{\mathbf{x}} = \mathbf{M}^{-1} \mathbf{b}_n \lambda \quad (51)$$

By inserting Eq. (51) into Eq. (49), the Lagrange multiplier λ is given by Eq. (52):

$$\lambda = \frac{\mathbf{b}_n^* v_n}{\mathbf{b}_n^* \mathbf{M}^{-1} \mathbf{b}_n} \dot{q}_n^* \quad (52)$$

With Eq. (52), $\dot{\mathbf{x}}$ of (51) can be determined as a function of \dot{q}_n^* . In turn, the kinetic energy of (48) can be written according to (53)

$$T(\dot{q}_n^*) = \frac{1}{2} m_n (\dot{q}_n^*)^2 \quad (53)$$

with the internal mass m_n associated with parameter q_n being given by

$$m_n = \frac{(\mathbf{b}_n^* v_n)^2}{\mathbf{b}_n^* \mathbf{M}^{-1} \mathbf{b}_n} \quad (54)$$

The denominator of Eq. (54) can be recognized as element G_{nn} of the \mathbf{G} matrix.

Once the internal force constant k_n (see Eq. 45) and the internal mass m_n (see Eq. 54) have been derived, the internal frequency ω_n is given by Eq. (55):

$$\omega_n^2 = \frac{\mathbf{v}_n^+ \mathbf{f} \mathbf{v}_n}{(\mathbf{b}_n^+ \mathbf{v}_n)^2} \frac{1}{G_{nn}} \quad (55)$$

Eq. (55) implies that if internal coordinate q_n represents the change in the bond distance of a diatomic molecular fragment AB caused by AB bond stretching, then $1/G_{nn}$ will be exactly equal to the reduced mass defined by $m_A m_B / (m_A + m_B)$. Furthermore, Eq. (55) reveals that in the general case $1/G_{nn}$ can be taken as the reduced mass associated with internal coordinate q_n , no matter which functional form q_n takes.

The term $(\mathbf{b}_n^+ \mathbf{v}_n)^2$ in the denominator of Eq. (55) guarantees proper normalization of vector \mathbf{v}_n . It suggests that the force constant k_n should be calculated according to Eq. (56) rather than Eq. (45):

$$k_n = \mathbf{v}_n'^+ \mathbf{f} \mathbf{v}_n' \quad (56)$$

with \mathbf{v}_n' given by

$$\mathbf{v}_n' = \frac{\mathbf{v}_n}{\mathbf{b}_n^+ \mathbf{v}_n} \quad (57)$$

This means that in Eq. (30) \mathbf{v}_n' rather than \mathbf{v}_n is used:

$$\mathbf{x}^* = \mathbf{v}_n' q_n^* \quad (58)$$

If Eq. (58) is multiplied from the left by \mathbf{b}_n^+ , then one will obtain $q_n = (\mathbf{b}_n^+ \mathbf{v}_n') q_n^*$. Because of Eq. (57) $\mathbf{b}_n^+ \mathbf{v}_n' = 1$, which ensures that q_n and q_n^* are the same during an internal vibration. This is of crucial importance for the calculation of internal force constants. If $\mathbf{v}_n = \mathbf{a}_n$, \mathbf{v}_n will be properly normalized in the sense that $\mathbf{b}_n^+ \mathbf{a}_n = 1$ (see Eq. 31b). The term $(\mathbf{b}_n^+ \mathbf{v}_n)^2$ in the denominator of (55) is important only when q_n is not equal to q_n^* . This is the case for c-vectors calculated with redundant sets of parameters [19].

7. CHARACTERIZATION OF NORMAL MODES IN TERMS OF INTERNAL VIBRATIONAL MODES

In the previous section, we have determined elementary modes of suitable structural units or molecular fragments ϕ_n that are associated with internal coordinates q_n describing these fragments. These so-called internal modes [18,19] play the same role in the understanding of the vibrating molecule as internal coordinates play in the understanding of molecular geometry and conformation, i.e. internal modes add a dynamic part to the static description of molecules with the help of internal coordinates.

The characterization of normal modes in terms of the localized internal modes (CNM analysis) [20] complements the NMA of vibrational spectroscopy and introduces a chemical aspect into vibrational spectroscopy, namely the

description of the dynamic behavior of molecules in terms of the dynamical properties of groups and molecular fragments. For the purpose of the CNM analysis, one has to define an amplitude $A_{n\mu}$, which specifies the contribution of a particular internal mode \mathbf{v}_n to a given delocalized normal mode \mathbf{l}_μ [20]. Utilizing amplitudes $A_{n\mu}$, one can decompose normal modes in terms of internal modes and, in this way, exactly relate the normal modes of a molecule to its structural units. This clearly facilitates the use of vibrational spectroscopy as a structure determining tool and extends its possible uses within chemistry.

Clearly, the assets of a useful, in itself noncontradictory, and physically based CNM analysis are the internal vibrational motions and their properties as well as the amplitudes that relate internal modes to normal modes. As shown in the previous section, the *adiabatic internal modes* \mathbf{a}_n are the appropriate candidates for internal modes. Adiabatic modes are based on a dynamic principle, they are calculated by solving the Euler-Lagrange equations, they are independent of the composition of the set of internal coordinates to describe a molecule, and they are unique in so far as they provide a strict separation of electronic and mass effects [18,19]. Therefore, they fulfil the first requirement for a physically based CNM analysis.

There are no explicit criteria that help to define a suitable amplitude A needed to describe the contribution of internal modes to normal modes and, then, to judge on the quality of this definition. However, there are properties that are implicitly assumed to be associated with amplitudes A . These can be formulated in the following way [20]:

1) Symmetry equivalent internal modes associated with symmetry equivalent internal coordinates must have the same amplitudes in the case that the normal mode being decomposed is symmetric. (*Symmetry criterion*)

2) The results of the CNM analysis should not change significantly if some internal motions with low amplitudes are changed or deleted in the expansion of the normal modes as it might happen when changing a redundant set of internal parameters into another set. (*Stability of results*) This can be checked by calculating amplitudes $A_{n\mu}$ of the same internal motions associated with the same internal coordinates q_n for a sequence of different parameter sets PSA, PSB, etc. The difference in amplitudes $\Delta A_{n\mu} = |A_{n\mu}(\text{PSA}) - A_{n\mu}(\text{PSB})|$ has to be evaluated for those internal motions covered by all parameter sets and summed over all normal modes \mathbf{l}_μ to obtain ΔA as a bar spectrum for the internal coordinates q_n considered. The spectrum ΔA - q_n provides a direct insight into the usefulness of the internal mode vectors \mathbf{v}_n and amplitudes $A_{n\mu}$ within the CNM analysis. (*Stability test of A with regard to variations in the parameter set used*)

3) Since it is not possible to directly evaluate the quality of a given definition of $A_{n\mu}$ one has to do this in an indirect way by comparing a normal mode frequency with suitable reference frequencies associated with internal coordinates q_n . It is physically reasonable to expect that if all normal modes \mathbf{l}_μ are studied for fixed internal modes \mathbf{v}_n (associated with fixed parameters q_n), then the magnitude of amplitudes $A_{n\mu}$ should become the smaller the larger the difference $\Delta\omega_{n\mu}$ between the normal mode frequency ω_μ and the fixed reference frequency ω_n is. Therefore,

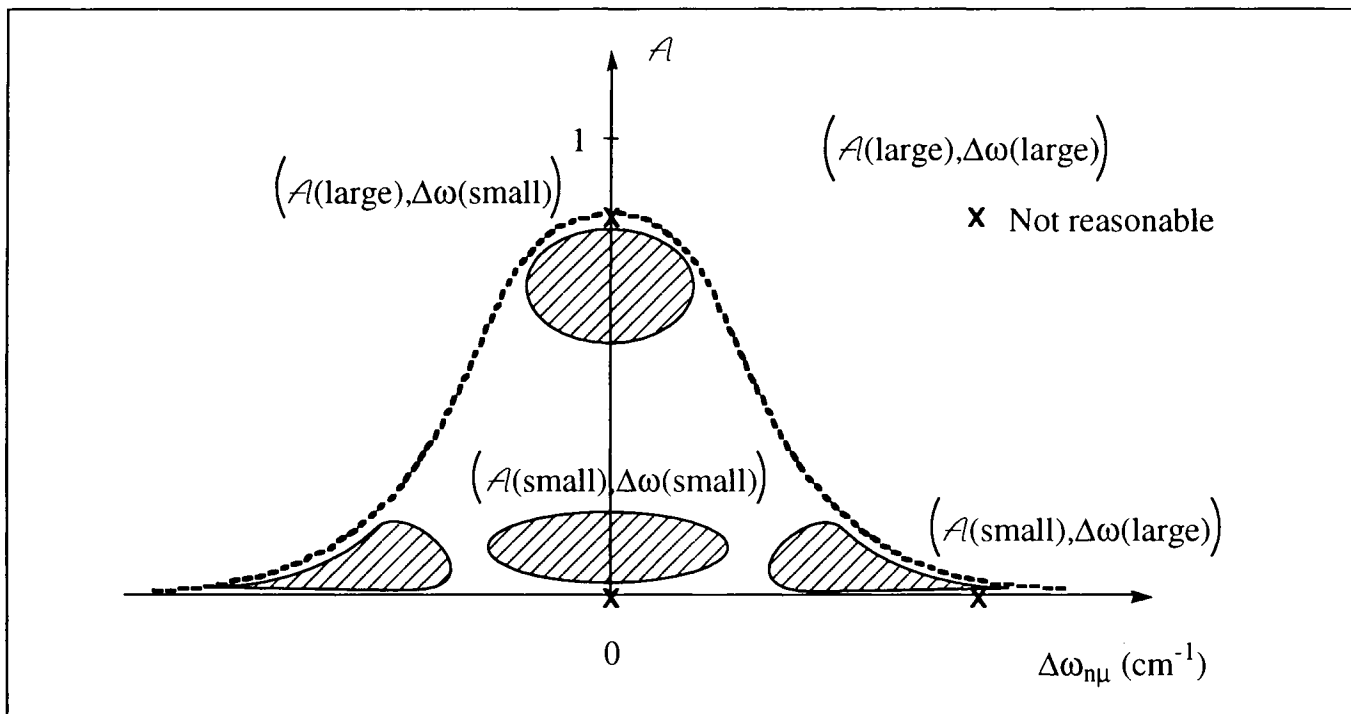


Figure 2. Different possibilities that can occur when plotting amplitudes $A_{n\mu}$ in dependence of the difference $\Delta\omega_{n\mu}$ between normal mode frequencies ω_{μ} and internal mode frequencies ω_n . The dashed line indicates the enveloping Lorentzian (bell-shaped) curve that can be expected in the case of a physically well-defined amplitude.

the distribution of all amplitudes $A_{n\mu}$ in dependence of differences $\Delta\omega_{n\mu} = \omega_n - \omega_\mu$ should be enveloped by a Lorentzian- (bell-)shaped curve as shown in Figure 2. The scattering of $A_{n\mu}$ in dependence of differences $\Delta\omega_{n\mu}$ outside or inside this enveloping curve provides a direct qualitative impression on the usefulness of the chosen amplitude and its underlying dynamical origin. If there are no amplitudes outside the enveloping curve, one can say that the dynamical origin of the normal mode principle will be fulfilled. (*Dynamical origin of normal mode concept*)

4) While 3) provides a crude qualitative test, its quantification is given by the quantity $h_{n\mu}$

$$h_{n\mu} = A_{n\mu} \Delta\omega_{n\mu} \quad (59)$$

which has the dimension of a frequency and can be considered as an uncertainty of the internal mode frequency. It provides a quantitative measurement of the usefulness of amplitude $A_{n\mu}$. In the normal case, the uncertainty $h_{n\mu}$ should have small vanishing values while an accumulation of large $h_{n\mu}$ values indicates deficiencies of amplitudes $A_{n\mu}$. (*Uncertainty test of internal mode frequencies*)

Provided the dynamical origin of the normal mode concept is correctly considered, the amplitude $A_{n\mu}$ will adopt a large value if the frequency difference $\Delta\omega_{\mu n} = \omega_\mu - \omega_n$ is relatively small, which simply means that the internal mode ν_n associated with the internal coordinate q_n dominates the normal mode l_μ and that the normal mode frequency ω_μ indicates the presence of the structural unit ϕ_n characterized by q_n and the internal mode frequency ω_μ :

$$A_{n\mu}(\text{large}) \Rightarrow \Delta\omega_{\mu,n}(\text{small}) \quad (60)$$

Relationship (60) is the basis for the empirical assignment of measured frequencies to structural units or fragments of a molecule.

Similarly, if there is a normal mode frequency ω_μ placed far from an internal mode frequency ω_n associated with fragment ϕ_n , then one will not expect a large amplitude since it is unlikely that the internal mode ν_n dominates the normal mode l_μ .

$$\Delta\omega_{\mu,n}(\text{large}) \Rightarrow A_{n\mu}(\text{small}) \quad (61a)$$

Hence, the case

$$\Delta\omega_{\mu,n}(\text{large}) \Rightarrow A_{n\mu}(\text{large}) \quad (61b)$$

should not occur. Of course, due to strong couplings within the molecule it can happen that, although a normal mode frequency ω_μ possesses a similar value as

the internal mode frequency ω_n , normal mode \mathbf{l}_μ has nothing in common with internal mode \mathbf{v}_n . This will be indicated by a low value of amplitude $A_{n\mu}$ according to

$$\Delta\omega_{\mu,n}(\text{small}) \Rightarrow A_{n\mu}(\text{small}) \quad (62)$$

If amplitudes $A_{n\mu}$ are plotted as a function of $\Delta\omega_{\mu,n}$, then the distribution of amplitude points should be enveloped by the Lorentzian (bell-shaped) curve of Figure 2 similar to the one describing the line shape of spectroscopic bands [9] since this curve complies with expectations (60) - (62).

8. DEFINITION OF INTERNAL MODE AMPLITUDES A

Any procedure to define an amplitude A must guarantee that normal and internal vibrational modes are related in a physically reasonable way [20]. The internal mode vector \mathbf{v}_n describes how the molecule vibrates when internal coordinate q_n that initiates ("leads") the internal motion is slightly distorted from its equilibrium value. From the NMA, one obtains normal mode vectors \mathbf{l}_μ , each of which shows how the atoms of a molecule move when the normal coordinate Q_μ is changed. By comparing the normal mode \mathbf{l}_μ with the internal mode \mathbf{v}_n the amplitude $A_{n\mu}$ is obtained that describes \mathbf{l}_μ in terms of the vibration of the smaller structural unit ϕ_n represented by displacement vector \mathbf{v}_n . Clearly, amplitude $A_{n\mu}$ has to be defined as a function of \mathbf{l}_μ and \mathbf{v}_n :

$$A_{n\mu} = f(\mathbf{l}_\mu, \mathbf{v}_n) \quad (63)$$

The internal mode vector \mathbf{v}_n can be defined with the help of the c-vectors (Eq. 22) as is implicitly assumed within the PED analysis [25-27]. Alternatively, one can use the adiabatic internal modes \mathbf{a}_n which are led by the associated internal parameters q_n as internal vibrational modes. The latter are preferred since they have a better physical justification than vectors \mathbf{c}_n , which should pay off when defining the amplitude $A_{n\mu}$ [18-20].

Once \mathbf{v}_n is chosen, one can compare the normal mode vibration \mathbf{l}_μ , with the vibration \mathbf{v}_n of a structural unit ϕ_n according to Eq. (64) [20]

$$A_{n\mu} = \frac{(\mathbf{l}_\mu, \mathbf{v}_n)^2}{(\mathbf{v}_n, \mathbf{v}_n)(\mathbf{l}_\mu, \mathbf{l}_\mu)} \quad (64)$$

where the symbol $A_{n\mu}$ is used to distinguish between a specific definition of A and the general amplitude $A_{n\mu}$. The denominator in (64) accounts for proper normalization and guarantees that $A_{n\mu}$ will adopt values between 0 and 1.

The scalar product (\mathbf{a}, \mathbf{b}) , which appears in the definition of the amplitude $A_{n\mu}$ (Eq. 64), can be defined in the most general way as

$$(\mathbf{a}, \mathbf{b}) = \sum_{i,j} a_i O_{ij} b_j \quad (65)$$

where O_{ij} is an element of the metric matrix \mathbf{O} and a_i and b_j are components of vectors \mathbf{a} and \mathbf{b} in Cartesian space. For the metric \mathbf{O} , there are three natural choices, namely

$$O_{ij} = \delta_{ij} \quad (66a)$$

$$O_{ij} = M_{ij} \quad (66b)$$

$$O_{ij} = f_{ij} \quad (66c)$$

with M_{ij} and f_{ij} being elements of the mass and force constant matrix, respectively. Eq. (66a) provides an estimate whether the two vectors \mathbf{a} and \mathbf{b} are spatially close, i.e. it measures their "spatial overlap". Eq. (66b) compares the two vectors kinetically ("mass comparison") and Eq. (66c) compares them dynamically ("force comparison"). Eqs. (66b) and (66c) reveal the influence of the atomic masses (*via* mass matrix \mathbf{M}) or that of the electronic structure (*via* force constant matrix \mathbf{f}) on the form of the normal mode l_μ .

The amplitude $A_{n\mu}$ defined in Eq. (64) can be considered as an "absolute amplitude". It is common practice to renormalize amplitudes and to express them as percentages according to Eq. (67):

$$A_{n\mu}^{\%} = \frac{A_{n\mu}}{\sum_m A_{m\mu}} 100 \quad (67)$$

to have a convenient way to compare them. This advantage has to be balanced against the fact that because of Eq. (67) amplitudes are no longer independent of the parameter set chosen.

According to which internal vibrational modes (c- or a-vectors: C_v or A_v) and according to which metric \mathbf{O} is used in Eq. (66) ($\mathbf{O} = \mathbf{S}, \mathbf{M}, \mathbf{f}$), different amplitudes can be defined, which are abbreviated in the following way:

$$\mathbf{O} = \mathbf{S} \begin{cases} A_vAS & A_vPS \\ C_vAS & C_vPS \end{cases} \quad (68)$$

$$\mathbf{O} = \mathbf{M} \begin{cases} A_vAM & A_vPM \\ C_vAM & C_vPM \end{cases} \quad (69)$$

$$\mathbf{O} = \mathbf{f} \begin{cases} A_vAF & A_vPF \\ C_vAF & C_vPF \end{cases} \quad (70)$$

where also the notation for \mathbf{P} matrix based "amplitudes" used in the PED analysis [25-27] and discussed in Ref. 20 have been added.

On purely theoretical grounds as well as on application examples it has been shown [20,21] that only two of the twelve amplitudes given in Eqs. (68), (69), and (70), namely AvAF and AvAM, are suitable for the task of comparing ω_μ with ω_n or decomposing I_μ in terms of v_n . The six Cv... amplitudes based on c-vectors are largely unstable with regard to changes in the internal coordinates chosen to describe a molecule and, therefore, they are not suited for a comparison of normal modes and internal modes. On theoretical grounds, the A-type amplitudes are clearly superior to the P-type amplitudes of the PED analysis, [25-27] which excludes the six P-based amplitude definitions of Eqs. (68), (69), and (70). A spatial comparison of two vectors or functions, although a common practice when one considers dipole moments, orbitals, etc., provides little information in the case of the dynamic process of vibrating molecules. Therefore, it is more useful to use as metric matrix either the mass matrix M (kinematic comparison) or the force constant matrix f (dynamic comparison), which leaves of the twelve possible amplitudes just AvAM and AvAF as amplitudes suitable for a comparison of normal modes and internal modes within the CNM analysis.

A short summary of these results is provided in Figure 3, which shows frequency uncertainty tests in form of $A_{n\mu-\Delta\omega_{\mu n}}$ diagrams, in which normalized amplitudes $A_{n\mu}$ are plotted as a function of frequency differences $\Delta\omega_{\mu n}$ for the benzocyclobutadiene molecule. Amplitudes and frequencies were calculated at the HF/6-31G(d,p) level of theory for both a nonredundant set of internal coordinates (Figures 3a - 3d) and a strongly redundant set of internal coordinates (Figures 3e - 3h), which are described in Ref. 21. Amplitudes AvAF, AvPF, CvAF, and CvPF are employed in connection with adiabatic internal frequencies and c-vector frequencies.

In Figures 3a-3d, there are relatively large differences between the correlation patterns for Av- and Cv-type amplitudes where the former lead to clearly better results. In view of an expected Lorentzian-shaped correlation pattern $A_{n\mu-\Delta\omega_{n\mu}}$, the worst result is obtained in the case of the CvPF amplitudes of the PED analysis, which indicates that the PED approach is a rather poor basis for carrying out a CNM investigation. Replacing the P-type amplitude by the corresponding A-type amplitude as in the CvAF diagram improves the situation somewhat, however, there are still severe shortcomings of the description, which is obviously a result of the shortcomings of the c-vectors [19].

Clearly, the best correlation pattern complying exactly with the expected Lorentzian form is obtained in the case of the AvAF amplitudes in connection with a comparison of frequencies ω_μ with adiabatic internal frequencies ω_a . Adiabatic internal modes, the amplitude definition of Eq. (64) and the force constant matrix f as a suitable metric for comparison provide the right ingredients for a physically well-founded CNM analysis.

Using a redundant internal coordinate set as in the case of Figures 3e - 3h, a significant improvement of all correlation patterns can be observed. This has to do with the fact that with increasing size of the redundant parameter set c-vectors adopt more the form of a-vectors [19]. For example, in the case of the nonredundant internal coordinate set the average overlap between adiabatic and c-vectors is 0.69, which means that the two types of internal mode vectors are

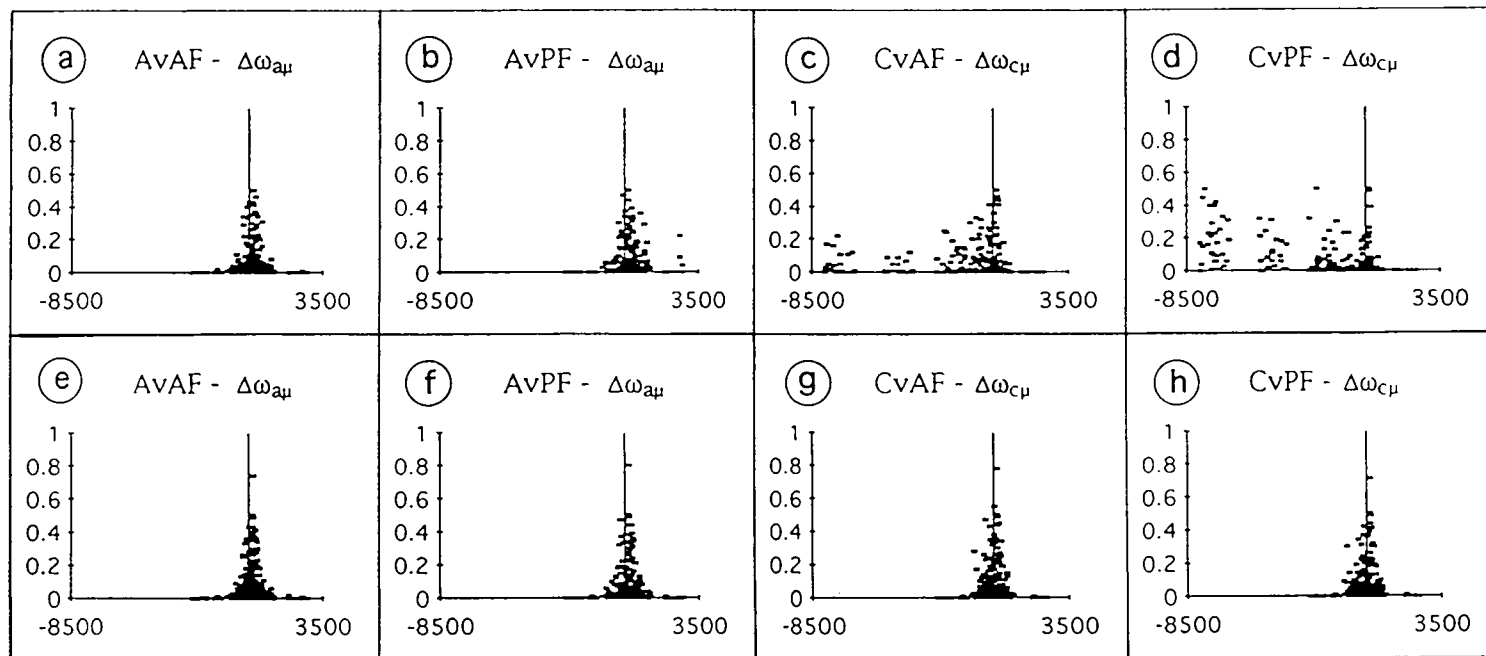


Figure 3. Frequency uncertainty test for benzocyclobutadiene according to HF/6-31G(d,p) calculations. The correlation diagrams correspond to correlations between normalized amplitudes $A_{n\mu}$ and frequency differences $\Delta\omega_{n\mu} = \omega_n - \omega_\mu$ with ω_n being a normal mode frequency of a molecular fragment ϕ_n and ω_μ being a normal mode frequency. Amplitudes AvAF, AvPF, CvAF, CvPF are employed in connection with adiabatic internal frequencies ω_a and c-vector frequencies ω_c using a nonredundant set of internal coordinates (a - d) or a strongly redundant set (e - h). In all cases, points that have $\Delta\mathcal{A} = 0$ for all tests within a given row of diagrams are removed.

indeed significantly different. In the case of the redundant coordinate set, the average overlap has increased to 0.84 without changing the form of the \mathbf{a} -vectors from that of the nonredundant coordinate set (\mathbf{a}_n is completely independent of the set of internal coordinates chosen), i.e. with increasing number of internal coordinates, \mathbf{c} -vectors will approach more and more the form of adiabatic vectors \mathbf{a}_n , which accordingly should be considered as the physically most reasonable internal vibrational mode vectors.

The 8 diagrams of Figure 3 clearly demonstrate the superiority of adiabatic internal mode vectors: They are independent of the choice of internal coordinates and determined just by the electronic structure of the molecule investigated. Obviously, one can improve in critical cases the usefulness of \mathbf{c} -vectors by using larger and larger redundant parameter sets, however this solution cannot be generalized so that in general \mathbf{a} vectors will always be superior to \mathbf{c} vectors.

9. ANALYSIS OF VIBRATIONAL SPECTRA IN TERMS OF ADIABATIC INTERNAL MODES

A characterization of vibrational normal modes in terms of adiabatic internal modes is straightforward with the definitions given in the previous sections. As an example, the vibrational modes of cyclopropane [28] will be discussed. They have been calculated at the HF/6-31G(d,p) level of theory and they are compared with experimental frequencies in Table 1.

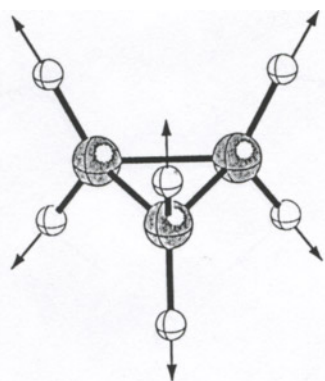
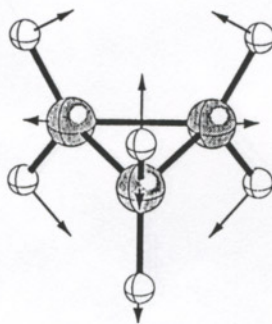
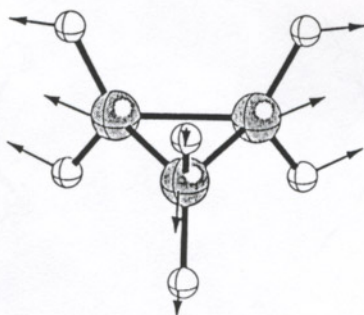
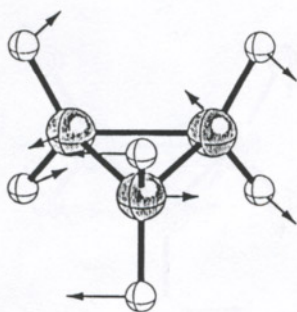
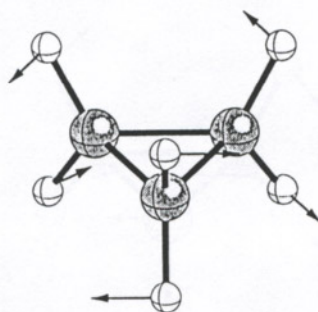
The normal modes of cyclopropane (see Figure 4) are easy to characterize because most of them involve motions associated with the same type of internal coordinate as, e.g., all six CH bond lengths (mode #1) or all three CH₂ twisting parameters (mode #5). Strong coupling between different types of internal parameters can only be found for modes #2, #10, #13 and #14 (Table 1). In the first two cases, CC stretching motions are mixed in, which becomes obvious from the pictorial representation of these modes given in Figure 4. However, these representations are sometimes misleading as can be seen from mode #3. According to the pictorial representation, one might expect that the ring breathing motion is connected with a CH scissoring or CH stretching motion, but the adiabatic analysis shows that mode #3 does not involve CH₂ scissoring or CH stretching. The arrows at the H atoms are simply a consequence of the movement of the C atoms. Modes #13 and #14 are a result of strong coupling between adiabatic CH₂ rocking and CH₂ twisting motions that is quantitatively described in the adiabatic mode analysis of Table 1.

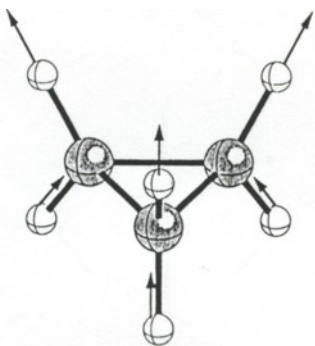
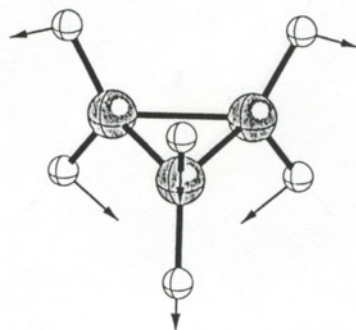
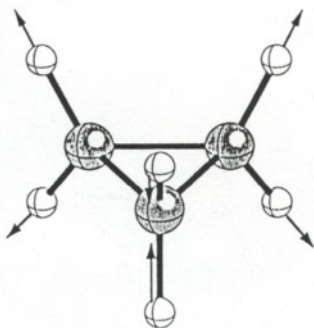
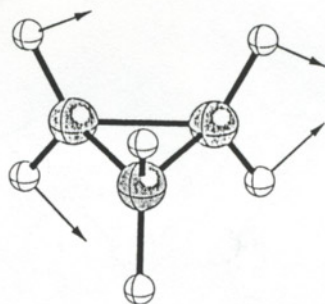
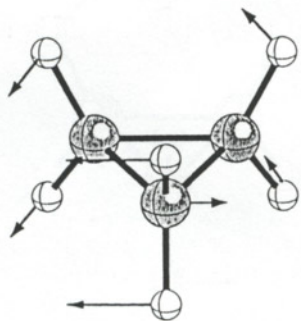
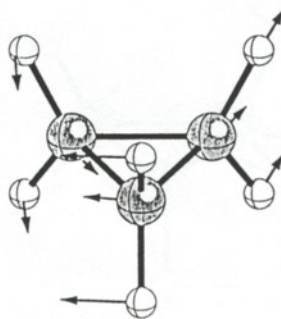
Adiabatic frequencies of cyclopropane are compared in Table 2 with those of some other hydrocarbons [28]. The adiabatic CC frequency is about 40 and the adiabatic CH stretching frequency about 130 cm⁻¹ larger than the corresponding values for cyclohexane. Compared to ethene, the adiabatic CH stretching frequencies are almost identical, which is in line with the high dissociation energy of the CH bond of cyclopropane [29]. The same observation has been made

Table 1
Analysis of the normal modes of cyclopropane using adiabatic internal modes. ^a

#	Sym	Frequencies			Characterization	Number of internal parameters
		exp.	HF, sc.	MP2, sc.		
1	a ₁ '	3038	3130	3077	CH stretch (96%)	CH: (6 x 16%)
2		1479	1570	1496	CH ₂ scissoring def (81%) + CC stretch (18%)	CH ₂ : (3 x 27%) + CC (3 x 6%)
3	a ₂ '	1188	1238	1182	CC stretch (90%)	CC stretch: (3 x 30%)
4		1070	1158	1043	CH ₂ wag (99%)	CH ₂ wag (3 x 33%)
5	a ₁ "	1126	1197	1129	CH ₂ twist (99%)	CH ₂ twist (3 x 33%)
6	a ₂ "	3102	3211	3184	CH stretch (100%)	CH stretch (6 x 16.7%)
7		854	873	846	CH ₂ rock (99%)	CH ₂ rock (3 x 33%)
8	e'	3024	3117	3067	CH stretch (100%)	CH stretch (4 x 25%)
9		1438	1515	1439	CH stretch (66%) CH ₂ def (98%)	CH stretch (2 x 33%) CH ₂ def (66% + 17% + 16%)
10	e"	1028	1113	1043	CH ₂ def (98%) CH ₂ wag (86%) + CC stretch (9%)	CH ₂ def (2 x 49%) CH ₂ wag (48% + 38%) + CC (9%)
11		868	911	872	CH ₂ wag (86%) + CC stretch (14%) CC stretch (88%)	CH ₂ wag (57 + 9 + 20%) + CC stretch (8 + 6%) CC stretch (64% + 24%)
12	e"	3082	3189	3168	CC stretch (96%) CH stretch (98%)	CC stretch (55% + 41%) CH stretch (2 x 25% + 2 x 24%)
13		1188	1271	1177	CH stretch (66%) CH ₂ rock (34%) + CH ₂ twist (49%)	CH stretch (2 x 33%) CH ₂ rock (34%) + CH ₂ twist (25 + 24%)
14	e"	739	744	734	CH ₂ rock (51%) + CH ₂ twist (33%) CH ₂ twist (56%) + CH ₂ rock (30%)	CH ₂ rock (26 + 25%) + CH ₂ twist (33%) CH ₂ twist (2 x 28%) + CH ₂ rock (30%)
					CH ₂ rock (44%) + CH ₂ twist (37%)	CH ₂ rock (2 x 22%) + CH ₂ twist (37%)

^a All frequencies in cm⁻¹. Scaled HF/6-31G(d,p) and MP2/(9s5p1d/4s1p)[4s2p1d/2s1p] frequencies; scaling factors are 0.87 (HF) and 0.95 (MP2), respectively. Each normal mode is dissected into adiabatic internal vibrations [28]. The notation CH: (6 x 16%) implies that all six CH stretching modes (each with 16%) of cyclopropane contribute to the normal mode #1.

#1, a_1' , in phase CH_2 sym. stretch#2, a_1' , CH_2 sym. scissor def#3, a_1' , C-C ring breath#4, a_2' , CH_2 wag#5, a_1'' , CH_2 twist

#6, a_2'' , in phase CH_2 asym. stretch#7, a_2'' , CH_2 rock#8, e' , out of phase CH_2 sym. stretch#9, e' , CH_2 asym scissor def#10, e' , CH_2 wag#11, e' , C-C ring def

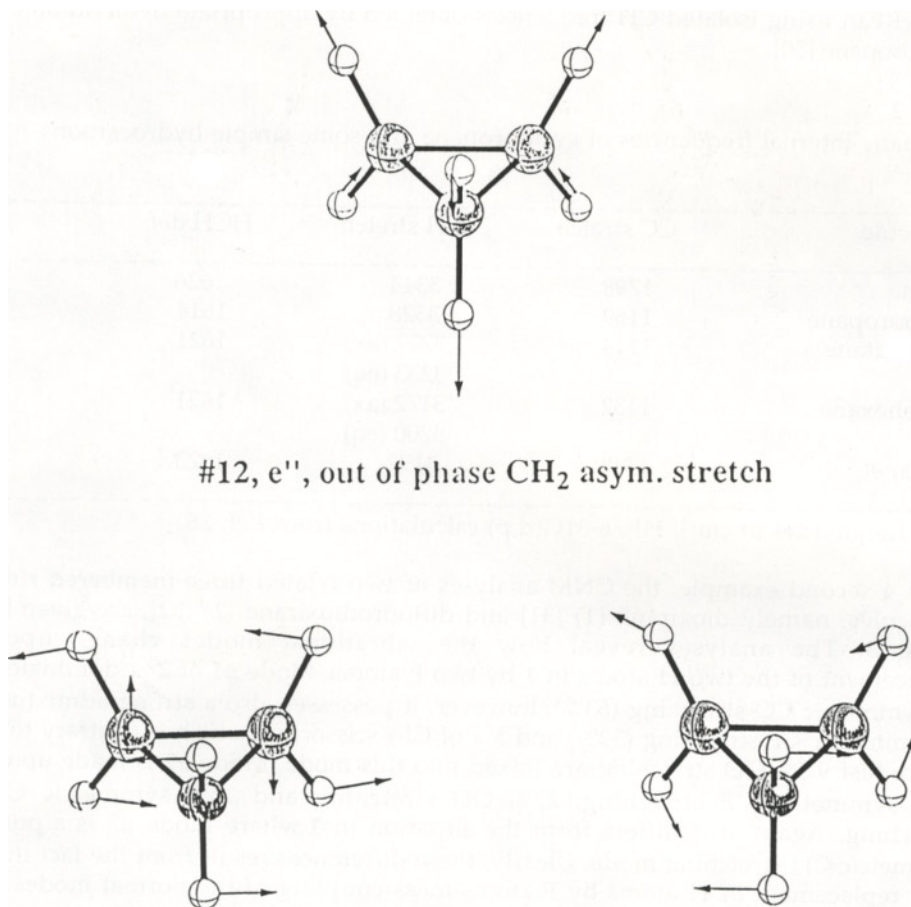


Figure 4. Vibrational modes of cyclopropane as obtained at the HF/6-31G(d,p) level of theory. Arrows indicate the direction and amplitude of each atomic motion. Symmetry assignments and a characterisation of each mode is also given in line with the notations used in Table 1.

by McKean using isolated CH frequencies obtained by appropriate deuteration of cyclopropane [30].

Table 2

Adiabatic internal frequencies of cyclopropane and some simple hydrocarbons. ^a

Molecule	CC stretch	CH stretch	HCH def
Ethene	1798	3344	1626
Cyclopropane	1169	3328	1614
Cyclobutane	1114	3222 (ax) 3233 (eq)	1621
Cyclohexane	1132	3172 (ax) 3200 (eq)	1621
Propane	1143	3192	1623

^a All frequencies in cm^{-1} . HF/6-31G(d,p) calculations from Ref. 28.

As a second example, the CNM analyses of two related three-membered ring molecules, namely dioxirane (**1**) [31] and difluorodioxirane (**2**) [32], are given in Table 3. The analyses reveal how the vibrational modes change upon replacement of the two H atoms in **1** by two F atoms. Mode #1 of **2** is dominated by symmetric CO stretching (63%), however, it possesses also a strong admixture of symmetric CF stretching (30%) and 5% of CF_2 scissoring, which is contrary to **1** where just 9.5% OO stretching are mixed into this mode. Mode #2 is made up of 60% symmetric CF stretching, 27% OO stretching and 13% symmetric CO stretching. Again, this differs from the situation in **1** where mode #2 is a pure symmetric CH stretching mode. Clearly, these differences result from the fact that by a replacement of H atoms by F atoms mass coupling of the normal modes is increased.

Modes #3 (84% OO stretching), #4 (88% CF_2 scissoring), #5 (100% CF_2 twisting), and #7 (92% CF_2 rocking) of **2** are less coupled with remaining contributions (see Table 3) being < 10%. Again, strong coupling is found in modes #6 (78% asymmetric CF stretching, 22% CF_2 rocking), #8 (67% asymmetric CO stretching, 33% CF_2 wagging), and #9 (73% CF_2 wagging and 27% asymmetric CO stretching). In the case of **1**, just two of the nine normal modes, namely the two symmetrical ring stretching modes, modestly couple with each other (admixture < 10%) while all other modes are almost uncoupled. This strikingly shows the influence of mass in mode coupling.

The calculated adiabatic frequencies reveal that the CO stretching modes increase by 70 cm^{-1} upon geminal F-substitution while the uncoupled OO stretching mode frequency decreases by just 44 cm^{-1} . Compared to oxirane (adiabatic CO stretching ω : 1130 cm^{-1}), the adiabatic CO stretching frequency of **1** (1121 cm^{-1} , Table 3) is normal while it is considerably increased for **2** (1189 cm^{-1}).

Table 3

Characterization of normal modes in terms of adiabatic internal modes for difluorodioxirane and dioxirane. ^a

#	Sym	exp.	Difluorodioxirane (2)		Dioxirane (1)		Adiabatic Frequencies	
			CCSD(T),sc.	Characterization	CCSD(T)	Characterization	2	1
1	a ₁	1467	1470	62.6% CO sym. str. (29.3% CF sym. str.; 5.5% CF ₂ scissor)	1311	89.4% CO sym. str. (9.5% OO str.)	1189, CO	1121, CO
2	a ₁	918	910	60.3% CF sym. str. (27.2% OO str.; 12.5% CO sym. str.)	3109	99.5% CH sym. str.	1200, CF	3180, CH
3	a ₁	658	658	84% OO str. (8.9% CF sym. str.; 7% CO sym. str.)	759	93.4% OO str. (6.1% CO sym. str.)	800, OO	844, OO
4	a ₁	511	512	88% CF ₂ scissor 5.6% CO sym. str.	1578	93.8% CH ₂ scissor	688, FCO	1267, HCO
5	a ₂	416	389	100% CF ₂ twist	1050	100% CH ₂ twist	688, FCO	1267, HCO
6	b ₁	1260	1259	78.2% CF asym. str. (21.8% CF ₂ rock)	3187	99.9% CH asym. str.	1200, CF	3180, CH
7	b ₁	557	559	92.5% CF ₂ rock (7.5% CF asym. str.)	1200	99.9% CH ₂ rock	688, FCO	1267, HCO
8	b ₂	1062	1068	66.7% CO asym. str. (33.3% CF ₂ wag)	931	99.6% CO asym. str.	1189, CO	1121, CO
9	b ₂	621	617	73.2% CF ₂ wag (26.8% CO asym. str.)	1292	96.1% CH ₂ wag	688, FCO	1267, HCO

^a Normal mode frequencies [cm⁻¹] for **1** from Ref. 31 (CCSD(T)/cc-VTZ2P+f,d), for **2** from Ref. 32 (CCSD(T)/cc-VTZ2P+f) and from Ref. 33 (exp.). Decomposition of normal modes in %. Second and third contributions are given in parentheses to facilitate reading. Adiabatic frequencies in cm⁻¹ according to MP2(full)/cc-VTZ2P calculations are given with regard to internal coordinates specified after each frequency as CO for CO stretching frequency, FCO as FCO bending frequency, etc.

This indicates typical changes in the CO bond strength upon geminal F substitution in **1**.

As indicated for **1** and **2**, the CNM analysis in terms of adiabatic internal modes makes it rather simple to correlate the vibrational spectra of related molecules and to discuss the influence of substituents, heteroatoms, and structural changes in terms of the internal mode frequencies. In the following section, we will provide further examples how vibrational spectra of different molecules can be correlated with the help of the CNM analysis.

10. CORRELATION OF VIBRATIONAL SPECTRA OF DIFFERENT MOLECULES

The CNM analysis in terms of adiabatic internal modes has been carried out to correlate the calculated vibrational spectra of the three dehydrobenzenes, namely *ortho*- (**3**), *meta*- (**4**) and *para*-benzyne (**5**), with the vibrational spectrum of benzene (**6**). Investigation of dehydrobenzenes with the help of infrared spectroscopy is of considerable interest at the moment since these molecules have been found to represent important intermediates in the reaction of enediyne anticancer drugs with DNA molecules [34-37]. Both **4** and **5** are singlet biradicals and, therefore, they are so labile that they can only be trapped at low temperatures in an argon matrix upon photolytic decomposition of a suitable precursor [38-40].

A positive identification of the dehydrobenzenes in the matrix requires, besides an expert set up of the experiment, high level *ab initio* calculations of the infrared spectra of the compounds trapped so that comparison between measured and calculated spectra becomes meaningful. In this way, both **4** and **5** have been identified and investigated in the matrix [38,39]. To further understand the electronic nature and the relationship of the three dehydrobenzenes, a correlation of their calculated vibrational spectra is desirable.

Kraka and co-workers [41] have calculated the vibrational spectra of **3**, **4**, and **5** at the GVB(1)/6-31G(d,p) level of theory where in each case the biradical nature of the dehydrobenzenes was described by the two-configuration approach of GVB. In Tables 4, 5, and 6, a CNM analysis of the calculated spectra based on calculated adiabatic internal modes is presented.

With the CNM analyses presented in Tables 4, 5, and 6 and a similar analysis for benzene, it is straightforward to correlate the vibrational spectra of the three benzyne with each other and with that of benzene. This is done in Tables 7, 8, and 9, which should be read considering that benzene has 30 normal modes while the benzyne have only 24. Hence, not all normal modes of benzene can be correlated with normal modes of the benzyne.

Tables 7, 8, and 9 are the basis for the correlation diagram shown in Figure 5 that compares the calculated infrared spectra of the three benzyne with that of benzene. Only the normal modes with infrared intensities larger than 0.1 are considered in this comparison. The numbers in parentheses denote the mode numbers used in Tables 4-9. Dashed lines connect infrared bands that are related according to the CNM analyses presented in Tables 4, 5, and 6.

Table 4
 CNM analysis of the vibrational spectrum of 1,2-didehydrobenzene (*o*-benzyne, 3)
 calculated at the GVB(1)/6-31G(d,p) level of theory. ^a

#	Sym	Freq	Characterization	Detailed characterization
24	a ₁	3387	HC(88%)	HC(2*44)
23	b ₂	3384	HC(94%)	HC(2*47)
22	a ₁	3358	HC(86%)	HC(2*43)
21	b ₂	3341	HC(94%)	HC(2*47)
20	a ₁	1942	CC(66%)	CC(66%)
19	b ₂	1688	CC(40%)+HCC(30%)	CC(2*20)+HCC(2*8+2*7)
18	a ₁	1607	HCC(58%)+CC(30%)	HCC(2*15+2*14)+CC(20+10)
17	b ₂	1559	HCC(62%)+CC(30%)	HCC(2*24+2*7)+CC(2*15)
16	a ₁	1398	HCC(46%)+CC(37%)	HCC(2*16+2*7)+CC(2*15+7)
15	b ₂	1362	HCC(58%)+CC(34%)	HCC(2*29)+CC(2*17)
14	a ₁	1234	CC(53%)+HCC(44%)	CC(17+2*14+8)+HCC(2*22)
13	b ₂	1210	HCC(50%)+CC(38%)	HCC(2*16+2*11)+CC(2*19)
12	a ₂	1106	HCCC(78%)+CCCC(19%)	HCCC(2*28+2*11)+CCCC(19)
11	a ₁	1104	CC(68%)+HCC(8%)	CC(2*34)+HCC(2*4)
10	a ₁	1086	CC(76%)+HCC(10%)	CC(44+2*16)+HCC(2*5)
9	b ₁	1057	HCCC(90%)	HCCC(2*28+2*17)
8	a ₂	974	HCCC(96%)	HCCC(2*33+2*15)
7	b ₂	969	CCC(98%)	CCC(2*28+2*14+2*7)
6	b ₁	833	HCCC(98%)	HCCC(2*25+2*24)
5	a ₂	692	CCCC(91%)	CCCC(20+2*17+2*13+11)
4	a ₁	657	CCC(68%)	CCC(2*34)
3	b ₂	593	CCC(92%)	CCC(2*36+2*10)
2	a ₂	493	CCCC(91%)	CCCC(37+20+2*17)
1	b ₁	448	CCCC(90%)	CCCC(2*26+2*19)

^a Frequencies in cm⁻¹. The following notation is used: CC: CC stretching; HC: HC stretching; CCC: CCC bending; HCC: HCC bending; CCCC: ring torsion; HCCC(op): hydrogen out-of-plane bending. The last column gives a detailed analysis of each normal mode in terms of adiabatic modes. E.g., normal mode #18 is described by 58% HCC in-plane bending and 30% CC stretching. The HCC in-plane bending comprises four different HCC bending modes, two with 15% and two with 14%, and the CC stretching mode comprises two different CC stretching modes, one with 20% and one with 10%.

Table 5
 CNM analysis of the vibrational spectrum of 1,3-didehydrobenzene (*m*-benzyne, 4)
 calculated at the GVB(1)/6-31G(d,p) level of theory. ^a

#	Sym	Freq	Characterization	Detailed characterization
24	a ₁	3420	HC(99%)	HC(99%)
23	a ₁	3377	HC(99%)	HC(42+41+16)
22	b ₂	3371	HC(100%)	HC(2*50)
21	a ₁	3342	HC(82%)	HC(82%)
20	a ₁	1792	CC(68%)	CC(2*24+2*10)
19	b ₂	1685	CC(42%)+HCC(26%)	CC(2*21)+HCC(26)
18	a ₁	1552	HCC(49%)+CC(42%)	HCC(25+24)+CC(2*12+2*10)
17	b ₂	1544	HCC(64%)	HCC(20+17+16+11)
16	a ₁	1384	HCC(89%)	HCC(61+2*14)
15	b ₂	1313	CC(60%)+HCC(27%)	CC(2*16+2*14)+HCC(27)
14	b ₂	1201	CC(64%)+HCC(31%)	CC(2*17+2*15)+HCC(2*9+8+5)
13	a ₁	1183	HCC(46%)+CC(26%)	HCC(2*23)+CC(2*13)
12	a ₁	1109	CC(68%)	CC(2*18+2*16)
11	b ₁	1102	HCCC(83%)	HCCC(57+2*13)
10	b ₂	1068	CC(70%)	CC(2*24+2*11)
9	a ₁	977	CCC(90%)	CCC(2*24+22+2*10)
8	a ₂	966	HCCC(92%)	HCCC(2*46)
7	b ₁	953	HCCC(80%)	HCCC(76+4)
6	b ₁	857	HCCC(71%)	HCCC(23+2*18+12)
5	b ₁	685	CCCC(94%)	CCCC(4*16+2*15)
4	b ₂	655	CCC(96%)	CCC(2*26+2*22)
3	a ₂	532	CCCC(60%)	CCCC(2*30)
2	a ₁	531	CCC(96%)	CCC(39+21+2*18)
1	b ₁	447	CCCC(82%)	CCCC(2*22+2*19)

^a See footnote in Table 4.

Table 6

CNM analysis of the vibrational spectrum of 1,4-didehydrobenzene (*p*-benzyne, 5) calculated at the GVB(1)/6-31G(d,p) level of theory. ^a

#	Sym	Freq	Characterization	Detailed characterization
24	a _g	3379	HC(100%)	HC(4*25%)
23	b _{2u}	3378	HC(100%)	HC(4*25%)
22	b _{3g}	3362	HC(100%)	HC(4*25%)
21	b _{1u}	3361	HC(100%)	HC(4*25%)
20	b _{3g}	1785	CC(72%)	CC(4*18%)
19	a _g	1646	CC(36%)+HCC(24%)	CC(2*18%)+HCC(4*6%)
18	b _{1u}	1604	HCC(64%)	HCC(4*16%)
17	b _{2u}	1456	HCC(40%)+CC(22%)	HCC(4*10%)+CC(2*11%)
16	b _{3g}	1397	HCC(92%)	HCC(4*23%)
15	b _{2u}	1297	CC(64%)+HCC(28%)	CC(4*16%)+HCC(4*7%)
14	a _g	1250	HCC(76%)	HCC(4*19%)
13	b _{1u}	1124	CC(52%)+HCC(28%)	CC(4*13%)+HCC(4*7%)
12	b _{2u}	1117	CC(64%)+HCC(16%)	CC(2*32%)+HCC(4*4%)
11	a _g	1097	CC(98%)	CC(2*17%+4*16%)
10	a _u	1080	HCCC op(88%)	HCCC op(4*22%)
9	b _{1u}	1050	CCC(100%)	CCC(2*20%+4*15%)
8	b _{2g}	1042	HCCC op(76%)	HCCC op(4*19%)
7	b _{1g}	897	HCCC op(100%)	HCCC op(4*25%)
6	b _{3u}	853	HCCC op(88%)	HCCC op(4*22%)
5	b _{2g}	715	CCCC(96%)	CCCC(6*16%)
4	a _g	662	CCC(60%)	CCC(2*30%)
3	b _{3g}	640	CCC(96%)	CCC(4*24%)
2	b _{3u}	492	CCCC(100%)	CCCC(4*25%)
1	a _u	464	CCCC(62%)	CCCC(2*31%)

^a See footnote in Table 4.

Table 7
Correlation of the normal modes of benzene (6) with those of 1,2-didehydrobenzene (*ortho*-benzyne, 3).^a

Benzene			<i>ortho</i> -Benzyne		
#	Sym	Characterization	#	Sym	Characterization
30	a _{1g}	sym HC st	24	a ₁	HC(88%)
29	e _{1u}	asym HC st	23	b ₂	HC(94%)
28	e _{1u}		-	-	-
27	e _{2g}	asym HC st	-	-	-
26	e _{2g}		22	a ₁	HC(86%)
25	b _{1u}	asym HC st	21	b ₂	HC(94%)
24	e _{2g}	ring st	20	a ₁	CC(66%)
23	e _{2g}		19	b ₂	CC(40%)+HCC(30%)
22	e _{1u}	ring def	-	-	-
21	e _{1u}		-	-	-
20	a _{2g}	HCC def	17	b ₂	HCC(62%)+CC(30%)
19	b _{2u}	HCC def	16	a ₁	HCC(46%)+CC(37%)
18	e _{2g}	HCC def	14	a ₁	CC(53%)+HCC(44%)
17	e _{2g}		13	b ₂	HCC(50%)+CC(38%)
16	b _{2u}	ring st	-	-	-
15	e _{1u}	HCC def	11	a ₁	CC(68%)+HCC(8%)
14	e _{1u}		-	-	-
13	b _{2g}	H twist	12	a ₂	HCCC(78%)+CCCC(19%)
12	e _{2u}	H twist	-	-	-
11	e _{2u}		9	b ₁	HCCC(90%)
10	b _{1u}	asym ring breath	7	b ₂	CCC(98%)
9	a _{1g}	sym ring breath	10	a ₁	CC(76%)+HCC(10%)
8	e _{1g}	HC wagging	-	-	-
7	e _{1g}		8	a ₂	HCCC(96%)
6	b _{2g}	chair o.p	5	a ₂	CCCC(91%)
5	a _{2u}	HC wagging	6	b ₁	HCCC(98%)
4	e _{2g}	ring def	3	b ₂	CCC(92%)
3	e _{2g}		4	a ₁	CCC(68%)
2	e _{2u}	twist boat o.p	2	a ₂	CCCC(91%)
1	e _{2u}	boat o.p	1	b ₁	CCCC(90%)

^a For an explanation of the notation used, see Table 4.

Table 8
Correlation of the normal modes of benzene (6) with those of 1,3-didehydrobenzene (*m*-benzyne, 4).^a

Benzene			<i>meta</i> -Benzyne		
#	Sym	Characterization	#	Sym	Characterization
30	a _{1g}	sym HC st	23	a ₁	HC(99%)
29	e _{1u}	asym HC st	23	a ₁	HC(99%)
28	e _{1u}		22	b ₂	HC(100%)
27	e _{2g}	asym HC st	22	b ₂	HC(100%)
26	e _{2g}		21	a ₁	HC(82%)
25	b _{1u}	asym HC st	21	a ₁	HC(82%)
24	e _{2g}	ring st	20	a ₁	CC(68%)
23	e _{2g}		19	b ₂	CC(42%)+HCC(26%)
22	e _{1u}	ring def	18	a ₁	HCC(49%)+CC(42%)
21	e _{1u}		17	b ₂	HCC(64%)
20	a _{2g}	HCC def	16	a ₁	HCC(89%)
19	b _{2u}	HCC def	15	b ₂	CC(60%)+HCC(27%)
18	e _{2g}	HCC def	13	a ₁	HCC(46%)+CC(26%)
17	e _{2g}		14	b ₂	CC(64%)+HCC(31%)
16	b _{2u}	ring st	10	b ₂	CC(70%)
15	e _{1u}	HCC def	13	a ₁	HCC(46%)+CC(26%)
14	e _{1u}		-	-	-
13	b _{2g}	H twist	11	b ₁	HCCC(83%)
12	e _{2u}	H twist	8	a ₂	HCCC(92%)
11	e _{2u}		11	b ₁	HCCC(83%)
10	b _{1u}	asym ring breath	9	a ₁	CCC(90%)
9	a _{1g}	sym ring breath	12	a ₁	CC(68%)
8	e _{1g}	HC wagging	8	a ₂	HCCC(92%)
7	e _{1g}		7	b ₁	HCCC(80%)
6	b _{2g}	chair o.p	5	b ₁	CCCC(94%)
5	a _{2u}	HC wagging	6	b ₁	HCCC(71%)
4	e _{2g}	ring def	2	a ₁	CCC(96%)
3	e _{2g}		-	-	-
2	e _{2u}	twist boat o.p	3	a ₂	CCCC(60%)
1	e _{2u}	boat o.p	1	b ₁	CCCC(82%)

^a For an explanation of the notation used, see Table 4.

Table 9
Correlation of the normal modes of benzene (6) with those of 1,4-didehydrobenzene (*p*-benzyne, 5).^a

Benzene			<i>para</i> -Benzyne		
#	Sym	Characterization	#	Sym	Characterization
30	a _{1g}	sym HC st	24	a _g	HC(100%)
29	e _{1u}	asym HC st	21	b _{1u}	HC(100%)
28	e _{1u}		23	b _{2u}	HC(100%)
27	e _{2g}	asym HC st	22	b _{3g}	HC(100%)
26	e _{2g}		24	a _g	HC(100%)
25	b _{1u}	asym HC st	21	b _{1u}	HC(100%)
24	e _{2g}	ring st	19	a _g	CC(36%)+HCC(24%)
23	e _{2g}		20	b _{3g}	CC(72%)
22	e _{1u}	ring def	18	b _{1u}	HCC(64%)
21	e _{1u}		17	b _{2u}	HCC(40%)+CC(22%)
20	a _{2g}	HCC def	16	b _{3g}	HCC(92%)
19	b _{2u}	HCC def	15	b _{2u}	CC(64%)+HCC(28%)
18	e _{2g}	HCC def	14	a _g	HCC(76%)
17	e _{2g}		16	b _{3g}	HCC(92%)
16	b _{2u}	ring st	12	b _{2u}	CC(64%)+HCC(16%)
15	e _{1u}	HCC def	13	b _{1u}	CC(52%)+HCC(28%)
14	e _{1u}		13	b _{1u}	CC(52%)+HCC(28%)
13	b _{2g}	H twist	8	b _{2g}	HCCC op(76%)
12	e _{2u}	H twist	10	a _u	HCCC op(88%)
11	e _{2u}		6	b _{3u}	HCCC op(88%)
10	b _{1u}	asym ring breath	9	b _{1u}	CCC(100%)
9	a _{1g}	sym ring breath	11	a _g	CC(98%)
8	e _{1g}	HC wagging	7	b _{1g}	HCCC op(100%)
7	e _{1g}		-	-	-
6	b _{2g}	chair o.p	5	b _{2g}	CCCC(96%)
5	a _{2u}	HC wagging	6	b _{3u}	HCCC op(88%)
4	e _{2g}	ring def	4	a _g	CCC(60%)
3	e _{2g}		3	b _{3g}	CCC(96%)
2	e _{2u}	twist boat o.p	1	a _u	CCCC(62%)
1	e _{2u}	boat o.p	2	b _{3u}	CCCC(100%)

^a For an explanation of the notation used, see Table 4.

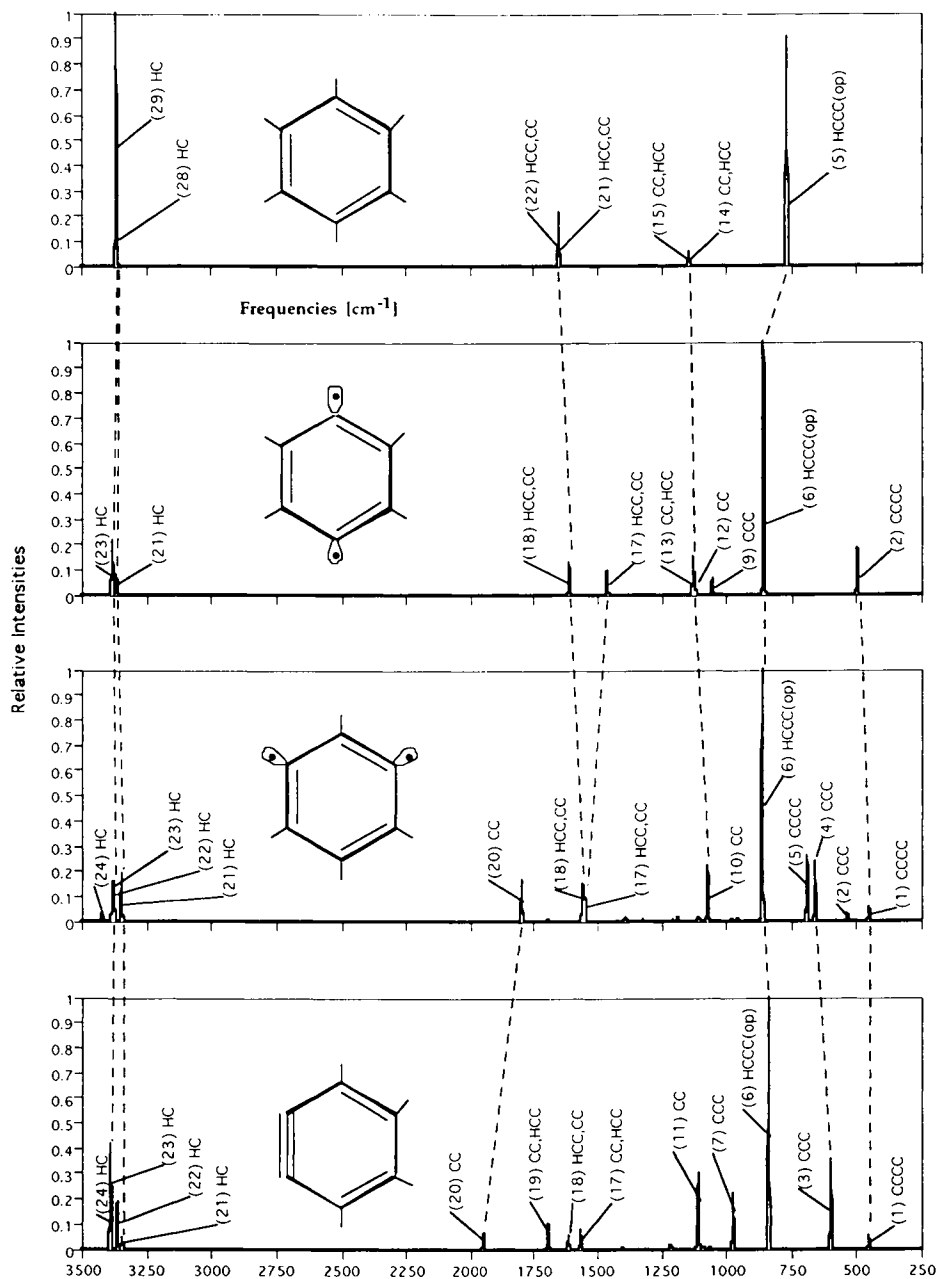
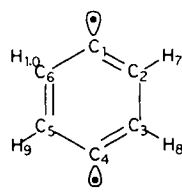
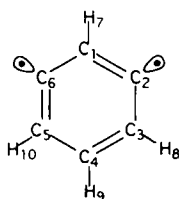
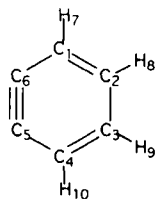


Figure 5. Correlation of the calculated infrared spectra of benzene (6), *para*-benzyne (5), *meta*-benzyne (4), and *ortho*-benzyne (3). For each infrared band, the corresponding mode number (in parentheses) and an appropriate characterization according to the CNM analysis of Tables 4 - 6 is given. Dashed lines correlate the infrared bands of different molecules.

Table 10

Adiabatic internal frequencies (in cm^{-1}) of *ortho*- (3), *meta*- (4) and *para*-benzyne (5) calculated at the GVB(1)/6-31(d,p) level of theory.



<i>ortho</i> -benzyne		<i>meta</i> -benzyne		<i>para</i> -benzyne	
Parameter	Freq	Parameter	Freq	Parameter	Freq
C5C6	1787	C1C2	1356	C1C2	1415
C2C3	1213	C1C6	1356	C1C6	1415
C1C6	1386	C2C3	1419	C3C4	1415
C4C5	1386	C5C6	1419	C4C5	1415
C1C2	1391	C3C4	1395	C2C3	1317
C3C4	1391	C4C5	1395	C5C6	1317
H7C1	3378	H7C1	3414	H7C2	3366
H10C4	3378	H9C4	3343	H8C3	3366
H8C2	3348	H8C3	3367	H9C5	3366
H9C3	3348	H10C5	3367	H10C6	3366
C1C6C5	739	C2C1C6	735	C1C2C3	953.8
C4C5C6	739	C3C4C5	963	C2C3C4	953.8
C2C1C6	919	C1C2C3	786	C4C5C6	953.8
C3C4C5	919	C1C6C5	786	C1C6C6	953.8
C1C2C3	1023	C2C3C4	963	C2C1C6	954.1
C2C3C4	1023	C4C5C6	963	C3C4C5	954.1
H7CC ip	1423	H7CC ip	1391	H7CC ip	1394
H10CC ip	1423	H9CC ip	1429	H8CC ip	1394
H8CC ip	1434	H8CC ip	1403	H9CC ip	1394
H9CC ip	1434	H10CC ip	1403	H10CC ip	1394
C1C6C5C4	560	C3C2C1C6	618	C3C2C1C6	638
C1C2C3C4	684	C2C1C6C5	618	C2C3C4C5	638
C2C1C6C5	581	C1C2C3C4	644	C3C4C5C6	638
C3C4C5C6	581	C1C6C5C4	644	C2C1C6C5	638
C3C2C1C6	646	C2C3C4C5	633	C1C2C3C4	639
C2C3C4C5	646	C3C4C5C6	633	C1C6C5C4	639
H7CCC op	957	H7CCC op	910	H7CCC op	945
H10CCC op	957	H9CCC op	1004	H8CCC op	945
H8CCC op	956	H8CCC op	920	H9CCC op	945
H9CCC op	956	H10CCC op	920	H10CCC op	945

The adiabatic internal frequencies calculated for the three benzyne are listed in Table 10 together with the associated internal coordinates. They have to be compared with the corresponding adiabatic frequencies of benzene obtained at the HF/6-31G(d,p) level of theory: CC 1406, HC 3348, CCC 997, HCC in-plane 1441, CCCC 653, and HCCC out-of-plane 969 cm^{-1} .

With the help of Figure 5, it is possible to identify the three benzyne and to discuss their electronic features. For example, **3** is best identified by its CC triple bond stretching frequency close to 1942 cm^{-1} (after scaling at 1690 cm^{-1}), which has a low intensity, but nevertheless should be observable since no other infrared bands appear in this region. Similarly, the boat-type ring torsion mode of **5** possesses contrary to the other molecules a stronger intensity in a region where no other infrared bands should appear (see Figure 5). In the case of **4**, it is the pattern of ring distortion modes in the region between 500 and 1500 cm^{-1} that facilitates its identification [38,39,41].

Similar correlations of vibrational spectra have been carried out with the help of the CNM analysis and the adiabatic frequencies for a number of molecules [40,41]. They all confirm the value of the CNM analysis that extends beyond a simple comparison of geometries. For example, in the case of the benzyne (Table 10), the adiabatic CC stretching frequencies do not correlate with the calculated CC equilibrium bond lengths. This has to do with the fact that, unlike to the equilibrium bond lengths, the adiabatic stretching frequencies are sensitive to the environment of the CC bonds. In **4**, bond C1C2 has a lower adiabatic frequency than bond C3C4 since a C1C2 stretching vibration leads to an increase of CH bond eclipsing strain and, therefore, this bond is stiffer. Bond C2C3, which one might expect to be comparable with bond C1C2, possesses an even higher adiabatic stretching frequency indicating in this way CC bond strengthening by through-bond interactions between the radical center C2 and the σ^* -orbital of bond C3C4 [36,37]. In a similar way, the other CC stretching frequencies listed in Table 10 can be discussed.

11. DERIVATION OF BOND INFORMATION FROM VIBRATIONAL SPECTRA

A serious attempt of associating measured normal mode frequencies with characteristic fragment frequencies was undertaken by McKean who investigated the stretching mode of the CH group in various hydrocarbons [30]. This author solved the problem of mode-mode coupling within the molecules investigated by D-substitution of all H atoms but the one considered thus increasing mass differences and reducing the amount of intramolecular mode-mode coupling. His approach led to characteristic CH stretching frequencies in different molecules and, by this, to a unique insight into the nature of the CH bond under different situations [30]. McKean could set up a linear relationship between the isolated CH stretching frequencies he measured and experimentally known CH bond lengths where both r_0 and r_s values had to be used. McKean suggested to employ this relationship for the determination of unknown CH bond lengths by infrared spectroscopy using measured CH stretching frequencies where he

predicted that this could be done with an accuracy of $\pm 0.0005 \text{ \AA}$ which is better than the accuracy achieved when determining CH bond lengths by microwave spectroscopy. In Figure 6, the linear relationship between measured C-H stretching frequencies taken from the work of McKean and equilibrium C-H bond lengths calculated by Larsson and Cremer [42] is shown. Calculated rather than measured C-H bond lengths are used since they provide a more consistent description of the relationship between frequencies and bond lengths than the r_0 and r_s values used by McKean.

Certainly, it is possible to obtain other characteristic fragment frequencies in a systematic way although an enormous amount of synthetic work is involved to get suitable isotopomers in each case. In addition, the measured fragment frequencies will always be contaminated by some residual coupling. Therefore, one can predict that it is hardly possible to solve, just by experimental means, the problem of determining fragment-specific frequencies.

In this situation, an attractive alternative is provided by the adiabatic internal frequencies. For example, the McKean relationship between C-H stretching frequencies and equilibrium C-H bond lengths can easily be reproduced with the help of adiabatic CH stretching frequencies calculated at the HF/6-31G(d,p) level of theory as is shown in Figure 7. The r^2 coefficient obtained is 0.998, which is clearly better than the r^2 coefficient for the correlation of the experimental frequencies (0.991, Figure 6). Eq. (71) gives the relationship between C-H equilibrium bond lengths and internal frequencies, which can be used

$$r_e(C-H) = -8.0155 \times 10^{-5} \omega_e(C-H) + 1.3442 \quad (71)$$

to calculate CH bond lengths once experimental values of isolated CH stretching frequencies are known. From a computational point of view, Eq. (72) is more useful since it provides C-H vibrational frequencies

$$\omega_e(C-H) = 16770 - 12476 r_e(C-H) \quad (72)$$

once the geometry of the molecule has been calculated.

Figure 7 confirms the McKean relationship [30] and, furthermore, suggests that calculated adiabatic internal frequencies are as useful or even more useful as the measured "isolated" C-H stretching frequencies. However, the real advantage of adiabatic frequencies will become obvious if one attempts to set up McKean relationships also for other bonds.

This question has been checked in the case of the CC bond [42]. In a molecule with more than one CC bond, individual CC stretching motions spread over several normal modes and there is mostly considerable coupling between the individual modes. To obtain "isolated" CC stretching frequencies similar as in the case of the CH bonds is impossible both for experimental and mass reasons. Synthesizing isotopomers, for which the C atoms of neighboring CC bonds are replaced by heavier isotopes just to "isolate" the CC bond under investigation would be a synthetically difficult and at the same time fruitless enterprise since a replacement of ^{12}C by ^{13}C or even ^{14}C isotopes means a too small change in the relative masses to achieve any effective mass decoupling. If one takes on the

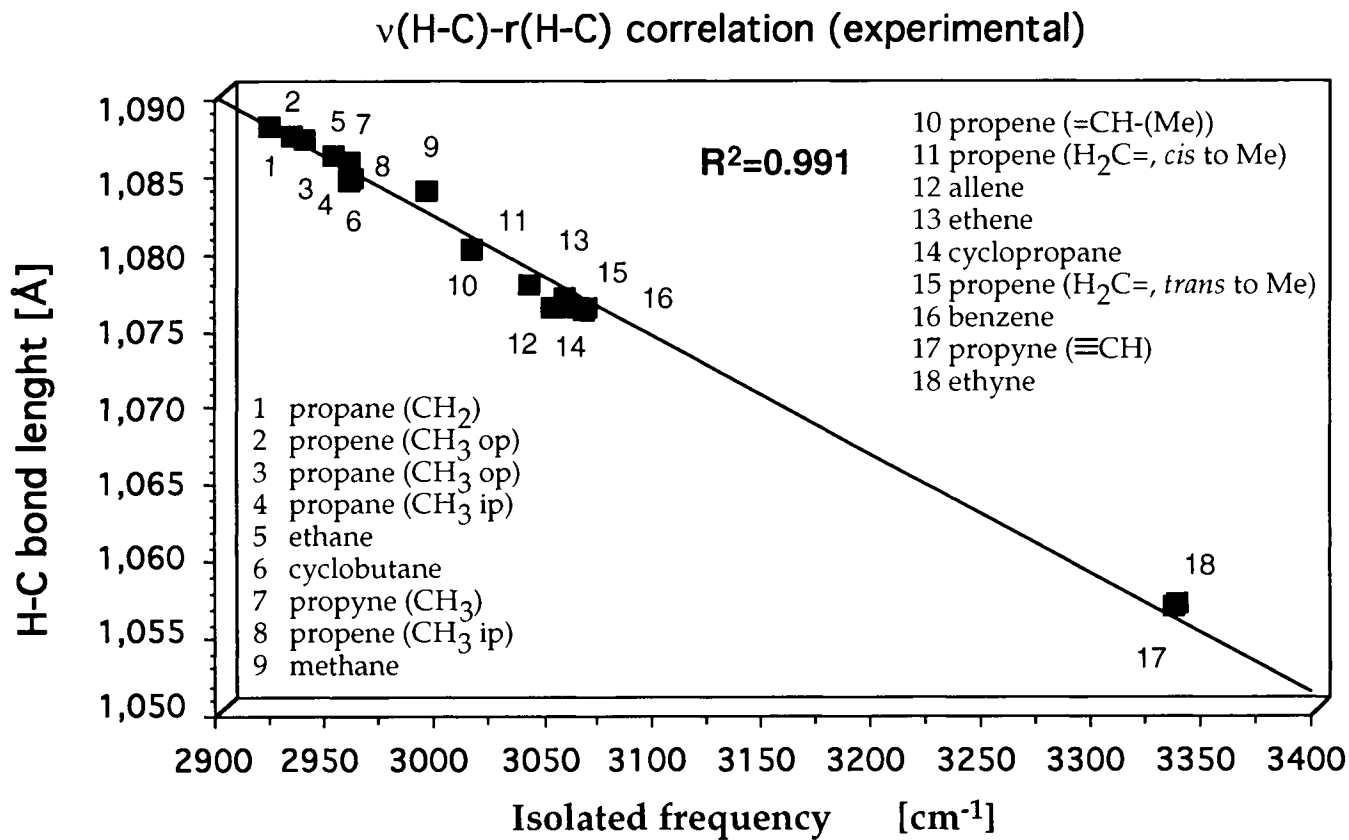


Figure 6. Linear correlation between CH bond lengths calculated at the HF/6-31G(d,p) level of theory [42] and the measured "isolated infrared frequencies" of McKean [30].

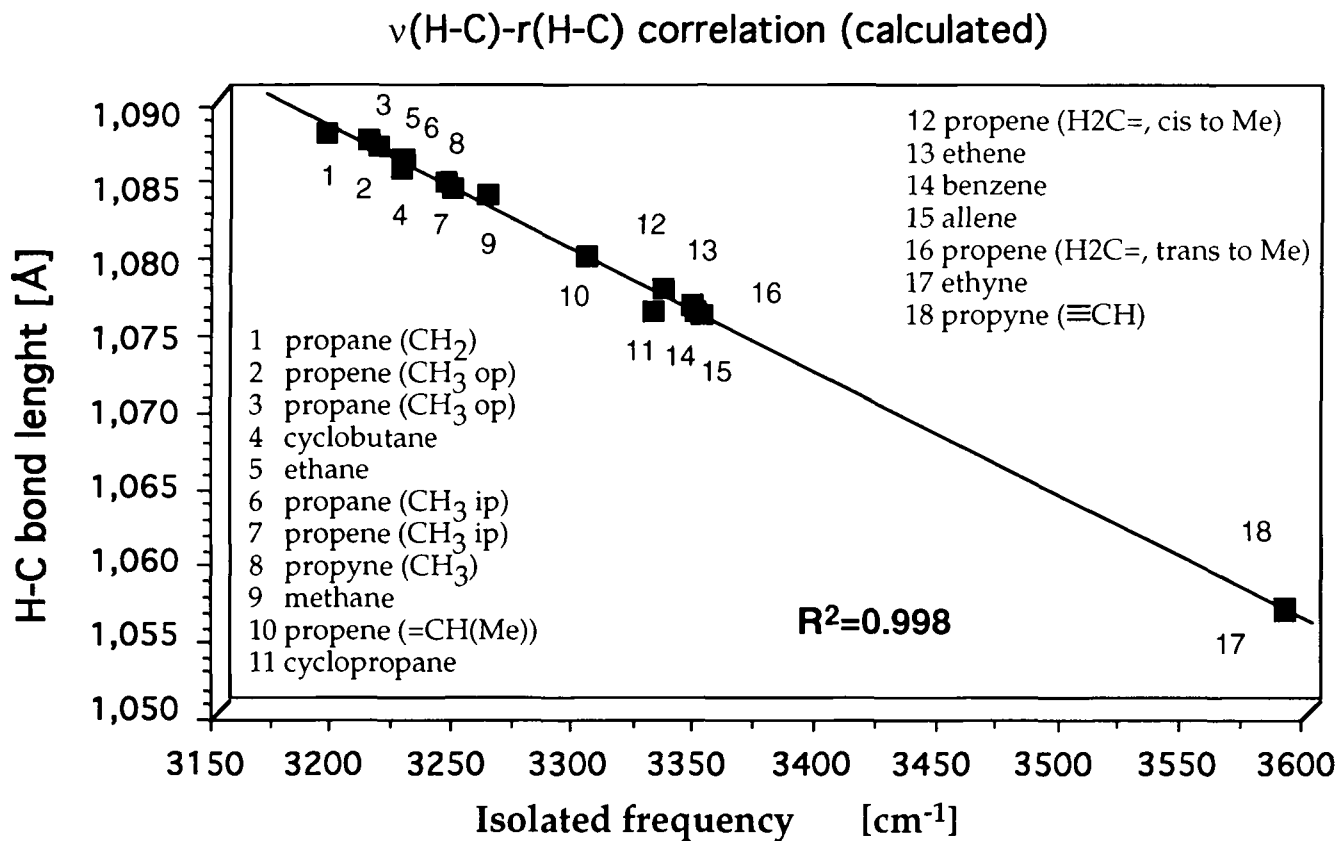


Figure 7. Linear correlation between CH bond lengths calculated at the HF/6-31G(d,p) level of theory and adiabatic CH stretching frequencies [42].

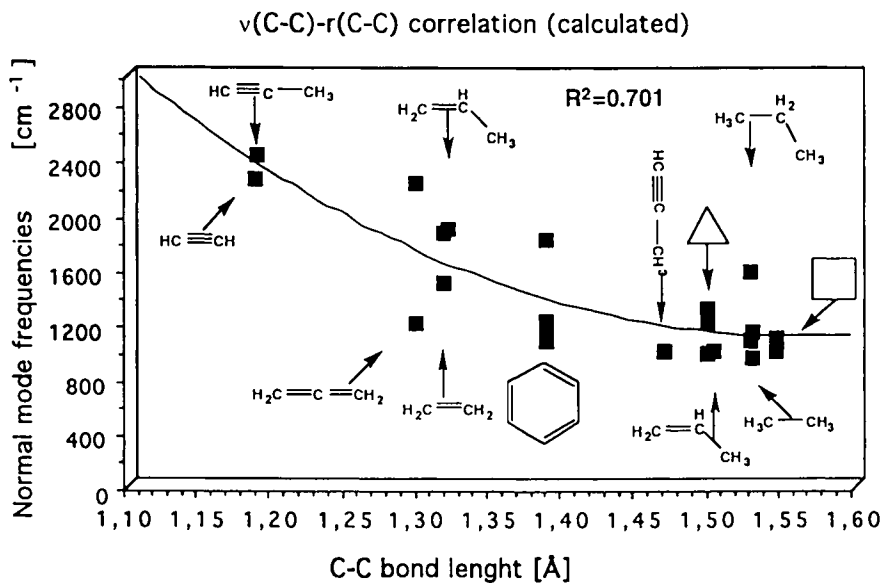


Figure 8. Quadratic correlation between normal mode frequencies generally considered to represent CC stretching frequencies and CC equilibrium bond lengths both calculated at the HF/6-31G(d,p) level of theory [42].

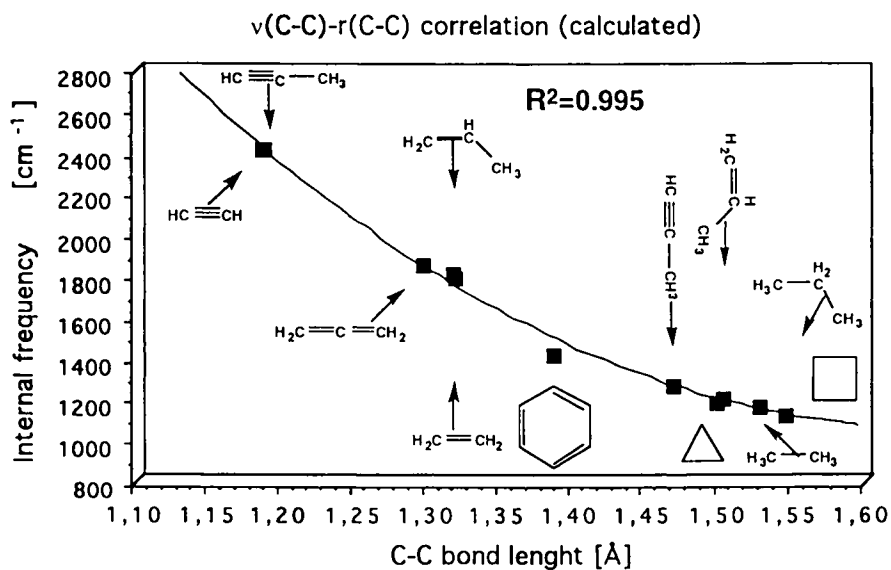


Figure 9. Quadratic correlation between adiabatic CC stretching frequencies and CC equilibrium bond lengths both calculated at the HF/6-31G(d,p) level of theory [42].

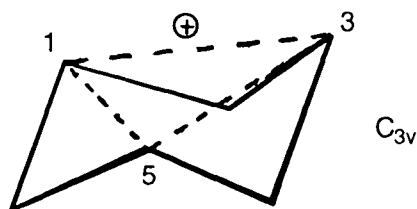
other hand the CC stretching modes in the way they are calculated in the normal mode analysis, then there is little correlation between CC stretching frequencies and equilibrium CC bond lengths as can be seen from Figure 8. There is a considerable scattering of CC stretching frequencies ($r^2 = 0.701$) around a ω_e/r_e relationship, which has basically quadratic form.

The problem can be solved by using adiabatic CC stretching frequencies derived from the normal mode frequencies of Figure 8 [42]. In Figure 9, a correlation between adiabatic CC stretching frequencies and CC equilibrium bond lengths is shown, which is best represented by the quadratic Eq. (73):

$$\omega_c(CC) = 18196 - 20742 r_c(CC) + 6268 [r_c(CC)]^2 \quad (73)$$

which leads to a correlation coefficient $r^2 = 0.995$.

One can use stretching frequency/bond length relationships to predict geometrical features of molecules from measured infrared spectra. This should be first checked for a case where verification of the prediction is possible due to an available geometry obtained from an *ab initio* calculation.



The trishomotropanylium cation contains a homoaromatic 2-electron-3-center system with long 1,3 distances [43]. If one calculates the adiabatic C1C3 stretching mode for this system at the MP2/6-31G(d,p) level of theory, one gets a value of 556 cm^{-1} . Using the quadratic relationship between stretching frequencies and CC distances shown in Figure 10, one obtains a C1C3 distance of 1.82 \AA , which is almost identical with the calculated MP2/6-31G(d,p) distance [43]. Hence, the example shows that distances can be reliably predicted once the value of the internal frequency is known.

12. ADIABATIC INTERNAL MODES FROM EXPERIMENTAL FREQUENCIES

Ab initio frequencies of normal vibrational modes and, by this also, adiabatic frequencies suffer from the harmonic approximation used in the calculation. Even when applying efficient scaling procedures, there is no guarantee that *ab initio* frequencies accurately reproduce the exact fundamental frequencies of the experiment. Therefore, one has to ask whether the adiabatic internal frequencies might not be much more meaningful if they would be based on experimental frequencies rather than frequencies calculated within the harmonic approximation.

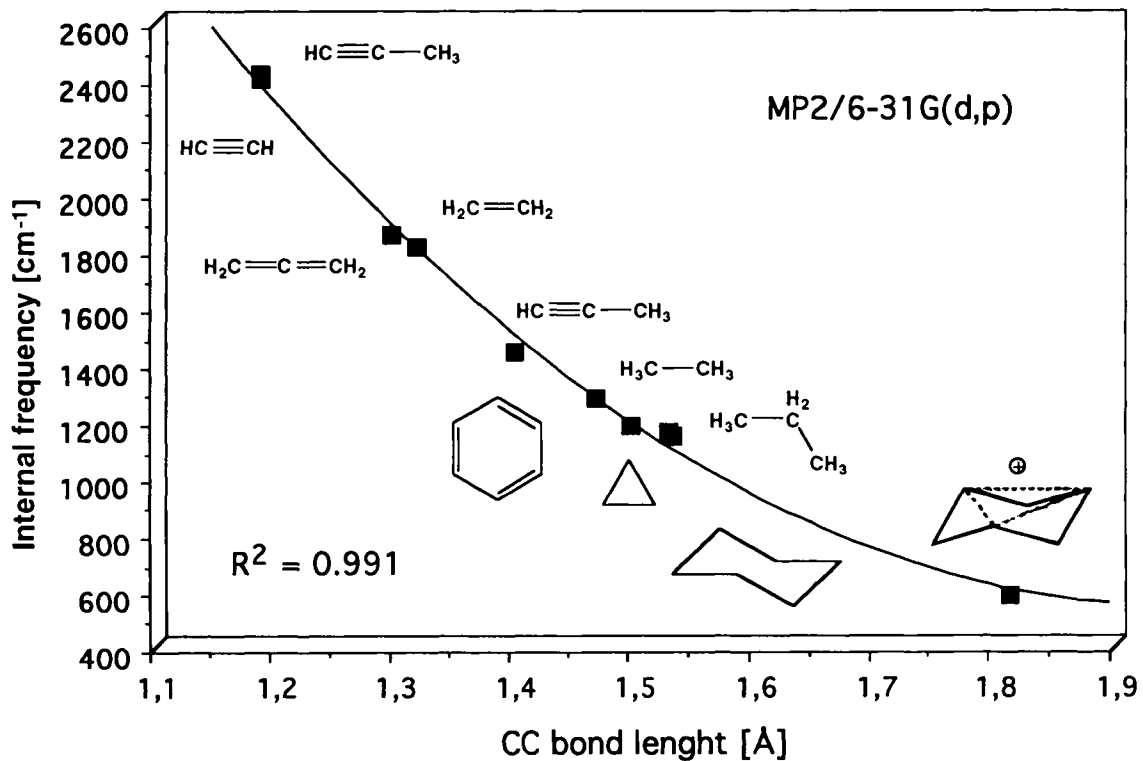


Figure 10. Determination of the C1C3 distance in the trishomotropenylium cation with the help of a quadratic correlation between adiabatic CC stretching frequencies and CC equilibrium bond lengths both calculated at the MP2/6-31G(d,p) level of theory [42].

An adiabatic mode analysis of measured vibrational spectra is possible with a simple perturbation theory approach that was already published in the sixties [44]. The basic equation of vibrational spectroscopy (compare with Eq. 19) can be written in matrix form according to Eq. (74)

$$\mathbf{F}\mathbf{D} = \mathbf{G}^{-1}\mathbf{D}\underline{\underline{\Lambda}} \quad (74)$$

where the matrix $\underline{\underline{\Lambda}}$ collects the squares of the frequencies ω_{μ} on its diagonal. Problem (74) can be solved as soon as \mathbf{F} is known from an appropriate *ab initio* calculation based on the harmonic approximation. One can assume that the calculated normal mode vectors \mathbf{d}_{μ} (expressed in terms of internal coordinates and collected in matrix \mathbf{D}) represent a reasonable approximation to the true normal mode vectors \mathbf{d}_{μ}' so that $\mathbf{D} \approx \mathbf{D}'$. This assumption makes some sense in view of the fact that reasonable experimental frequencies can be reproduced from calculated harmonic frequencies by simple scaling procedures. As a matter of fact, all scaling procedures are based on the assumption that \mathbf{D} is close to the true \mathbf{D}' .

Once the experimental frequencies are known, it is possible to derive an improved version of Eq. (74) for the experimental situation utilizing the calculated \mathbf{D} :

$$(\mathbf{F} + \Delta\mathbf{F})\mathbf{D} = \mathbf{G}^{-1}\mathbf{D}(\underline{\underline{\Lambda}} + \Delta\underline{\underline{\Lambda}}) \quad (75)$$

in which the correction matrix $\Delta\mathbf{F}$ has to be determined with the help of \mathbf{F} , \mathbf{D} , \mathbf{G} and $\underline{\underline{\Lambda}}$ obtained from the *ab initio* calculation and $\Delta\underline{\underline{\Lambda}}$ from experimental frequencies. This can be done by solving

$$\Delta\mathbf{F}\mathbf{D} = \mathbf{G}^{-1}\mathbf{D}\Delta\underline{\underline{\Lambda}} \quad (76)$$

For this purpose, one defines the matrices defined in Eqs. (77a) and (78):

$$\tilde{\mathbf{D}} = \mathbf{G}^{-1/2}\mathbf{D} \quad (77a)$$

$$\tilde{\mathbf{D}}^+\tilde{\mathbf{D}} = \tilde{\mathbf{D}}\tilde{\mathbf{D}}^+ = \mathbf{I} \quad (77b)$$

$$\Delta\tilde{\mathbf{F}} = \mathbf{G}^{1/2}\Delta\mathbf{F}\mathbf{G}^{1/2} \quad (78)$$

so that the eigenvalue problem (79) can be formulated

$$\Delta\tilde{\mathbf{F}}\tilde{\mathbf{D}} = \tilde{\mathbf{D}}\Delta\underline{\underline{\Lambda}} \quad (79)$$

By diagonalization, $\Delta\tilde{\mathbf{F}}$ and $\Delta\mathbf{F}$ can be determined:

$$\Delta\tilde{\mathbf{F}} = \tilde{\mathbf{D}} \Delta\tilde{\mathbf{A}} \tilde{\mathbf{D}}^+ \quad (80a)$$

$$\Delta\mathbf{F} = \mathbf{G}^{-1/2} \Delta\tilde{\mathbf{F}} \mathbf{G}^{-1/2} \quad (80b)$$

Hence, the experimental situation is described by a change $\Delta\mathbf{F}$ of the force constant matrix and $\Delta\tilde{\mathbf{A}}$ in the square of the frequencies relative to the calculated force constant matrix and frequencies at some level of *ab initio* theory. Since $\Delta\tilde{\mathbf{A}}$ is known from the differences between experimental and calculated frequencies, it is straightforward to calculate $\Delta\mathbf{F}$ and the true force constant matrix. Once the true force constant matrix is determined, one can apply the adiabatic mode analysis in the same way as it is applied for calculated vibrational spectra.

As a simple example, experimental and calculated adiabatic mode frequencies of ethene and methane are shown in Figure 11. The two sets of adiabatic frequencies differ on the average by 200 cm^{-1} . It is interesting to compare what one normally considers as the typical CC stretching frequency in ethene (1623 cm^{-1} [45]) and the adiabatic stretching frequency (1566 cm^{-1} , Figure 10). The difference of about 60 cm^{-1} results from coupling of the normal mode dominated by the CC stretching motion with other normal modes and, of course, from some of the assumptions included in the calculation of the experimentally derived adiabatic frequencies.

One can determine experimentally-based adiabatic CC and CH stretching frequencies and correlate them with bond lengths as discussed in the previous section. We have checked the McKean correlations between the isolated CH frequencies and the CH bond lengths in this fashion [42]. Also, we have used experimentally based adiabatic internal CH stretching frequencies to correlate them with other bond properties such as dissociation energies. The existence of correlations between stretching frequencies and dissociation energies has been discussed in the literature [46,47]. The stretching frequency (and its force constant) gives a measure for the curvature of the potential energy surface in the direction of a dissociation reaction. A large (small) stretching frequency suggests a strong (weak) curvature and a large (small) dissociation energy (see Figure 12).

Figure 13 gives the correlation between dissociation energy D_e [48] and adiabatic internal frequency for a number of CH bonds. It can be expressed by Eq. (81)

$$D_e = (\omega_e - 2080.3)/11.379 \quad (81)$$

where dissociation energies D_e are given in kcal/mol and the correlation coefficient r^2 is 0.969. Clearly, there is a direct relationship between the curvature of the potential energy surface in the direction of the CH bond dissociation and the energy difference between molecule and dissociation products, which can be used to predict D_e values.

These examples confirm that the adiabatic mode analysis can be extended with advantage to experimental vibrational spectra provided all experimental frequencies are known. However, even in the case, in which the set of

Adiabatic Internal Frequencies [cm^{-1}]

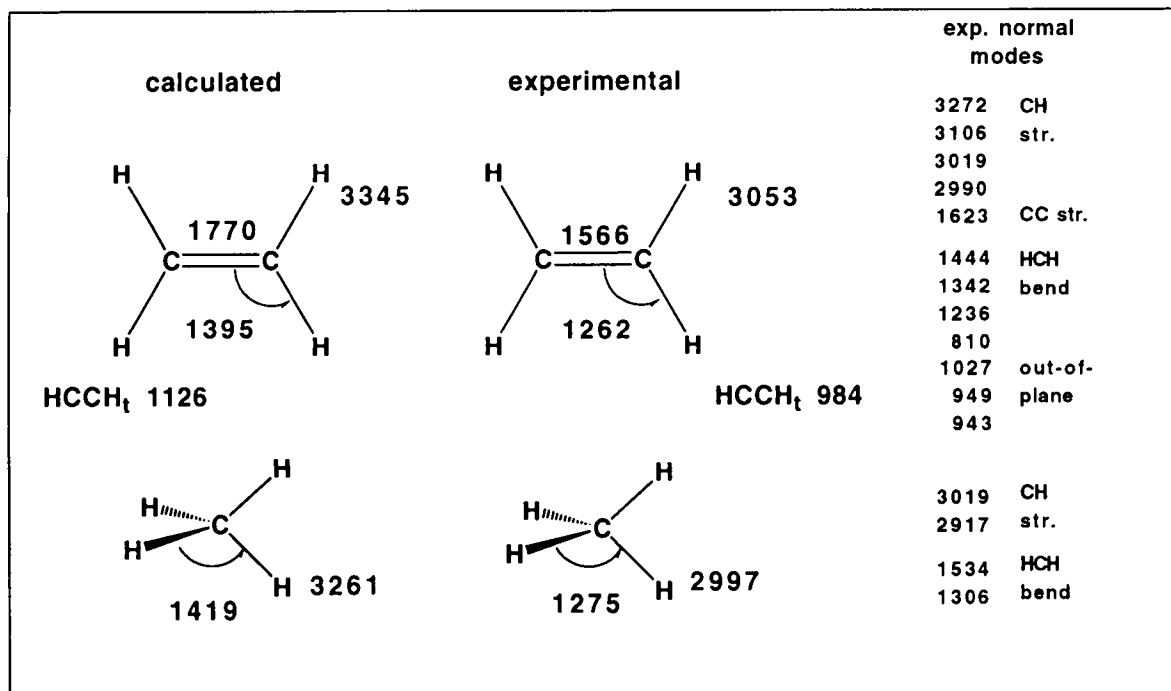


Figure 11. Comparison of directly calculated adiabatic frequencies (HF/6-31G(d,p) calculations) and adiabatic frequencies derived from experimental vibrational frequencies (see text).

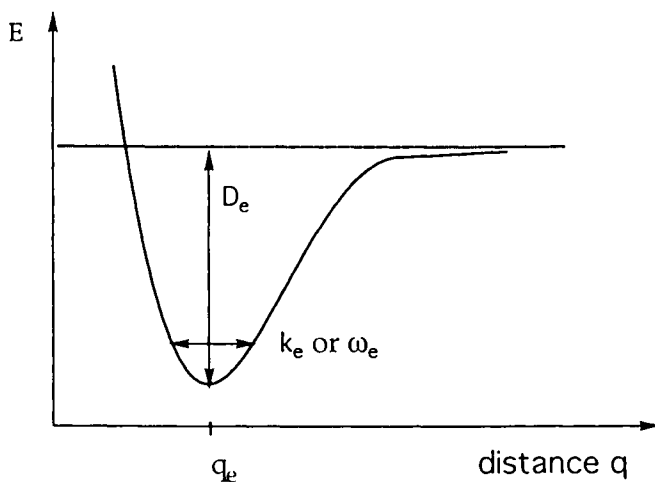


Figure 12. Relationship between dissociation energy and curvature of the potential energy function at the equilibrium represented by force constant k_e or vibrational frequency ω_e .

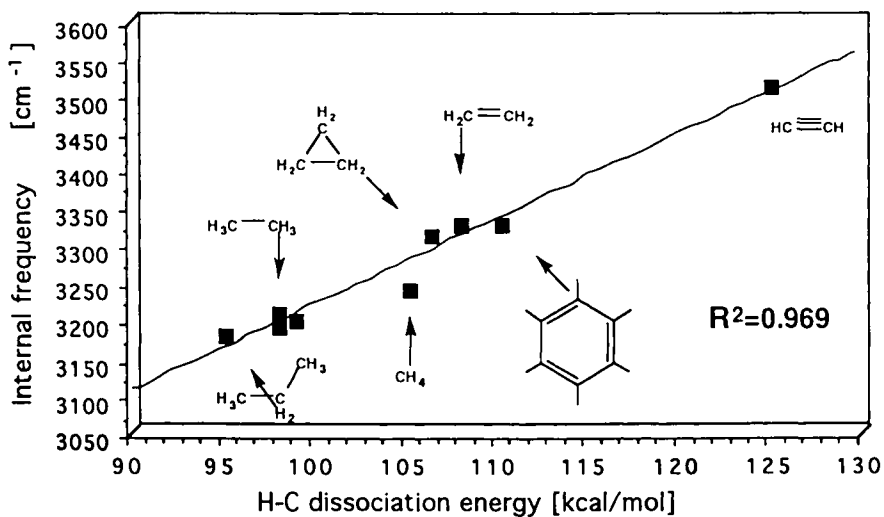


Figure 13. Linear correlation between adiabatic CH stretching frequencies (obtained from experimental frequencies and HF/6-31G(d,p) normal mode vectors) and experimental CH dissociation energies.

experimental frequencies is incomplete, appropriately scaled calculated frequencies can be used to complement the set of experimental frequencies and to carry out the adiabatic mode analysis.

13. A GENERALIZATION OF BADGER'S RULE

Already in the 30s of this century, diatomic molecules were investigated to correlate vibrational spectroscopic constants with bond lengths. The most successful of these relations was the Badger's rule (82) [49]

$$k_e \left(r_e - d_{ij} \right)^3 = C_{ij} \quad (82a)$$

or

$$r_e = (C_{ij} / k_e)^{1/3} + d_{ij} \quad (82b)$$

In Eq. (82), r_e is the equilibrium bond length of a diatomic molecule, k_e the associated bond stretching force constant, and C_{ij} as well as d_{ij} are constants, which depend on the rows i and j of the periodic table containing the atoms linked by the bond. Badger's rule has often been used to determine r_e values from spectroscopic constants. Today, it is applied in *ab initio* quantum chemical programs to provide an estimate of the Hessian matrix for the starting point of a geometry optimization, i.e. all distances of the starting geometry are known and appropriate k values have to be estimated to set up the Hessian matrix.

Badger found linear relationships between $k_e^{1/3}$ and r_e in the case of diatomic molecules [49]. Several attempts to generalize these relationships for polyatomic molecules failed because appropriate force constant values k_e for diatomic subunits within a polyatomic molecule were not available. This problem can now be solved with the help of the adiabatic modes. In Figures 14 and 15, $k_e^{1/3}$ versus r_e correlations are shown for the CH and CC adiabatic modes of Figures 7 and 9.

Linear relations (correlation coefficients r^2 for CH: 0.997 and CC: 0.993) are obtained for the diatomic subunits of polyatomic molecules similar to those investigated for diatomic molecules by Badger. This author suggested that the $k_e^{1/3}$ versus r_e relationships for diatomic molecules constituted of atoms from different rows of the periodic table could be reproduced by a series of parallel lines. Badger anticipated similar $k_e^{1/3}$ versus r_e relationships for polyatomic molecules, which can now be checked with the help of adiabatic stretching force constants. In Figure 16, $k_e^{1/3}$ - r_e correlations are given for eight different bond types involving H and first row atoms. For all bond types, a linear relationship is obtained with a correlation coefficient $r^2 > 0.98$ or even 0.99. One can distinguish between AH bonds and AB bonds. Within each class correlation lines are parallel or at least almost parallel where Badger's assumption is better fulfilled for the bonds AH with A = C, N, O, and B rather than the AB bonds (Figure 16) [42].

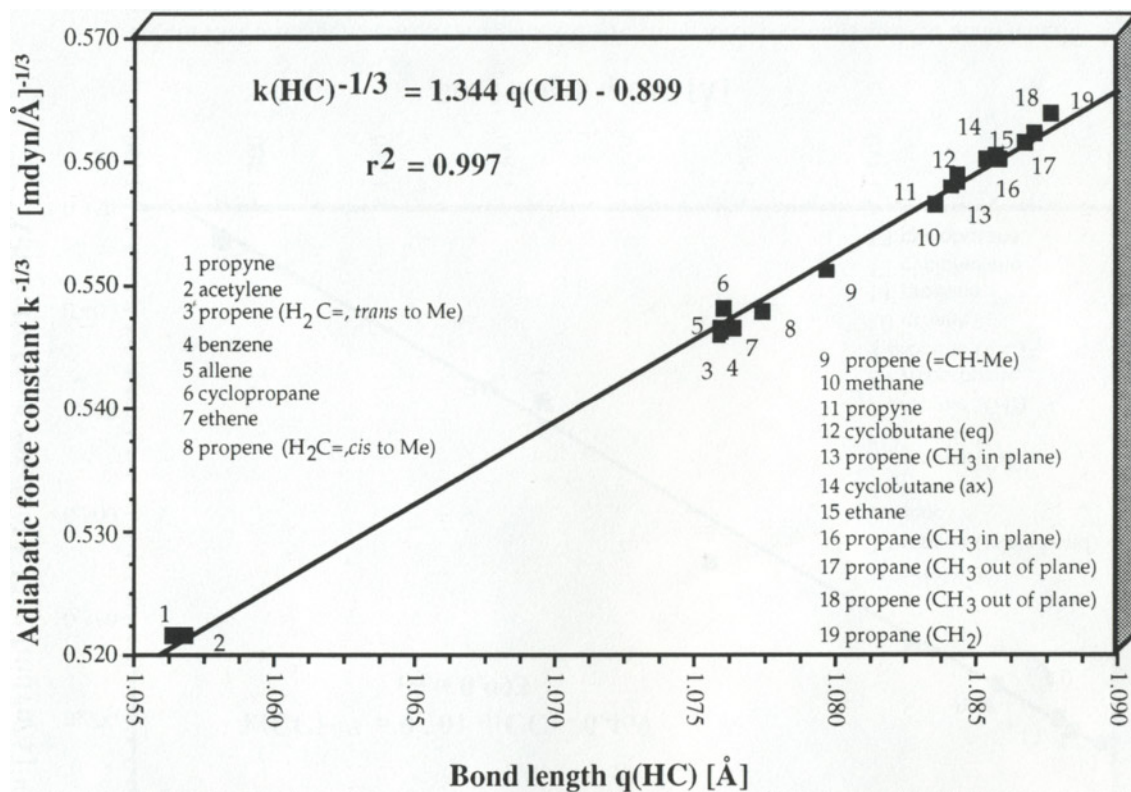


Figure 14. Correlation between adiabatic CH stretching force constants and CH bond lengths according to Badger. (HF/6-31G(d,p) calculations)

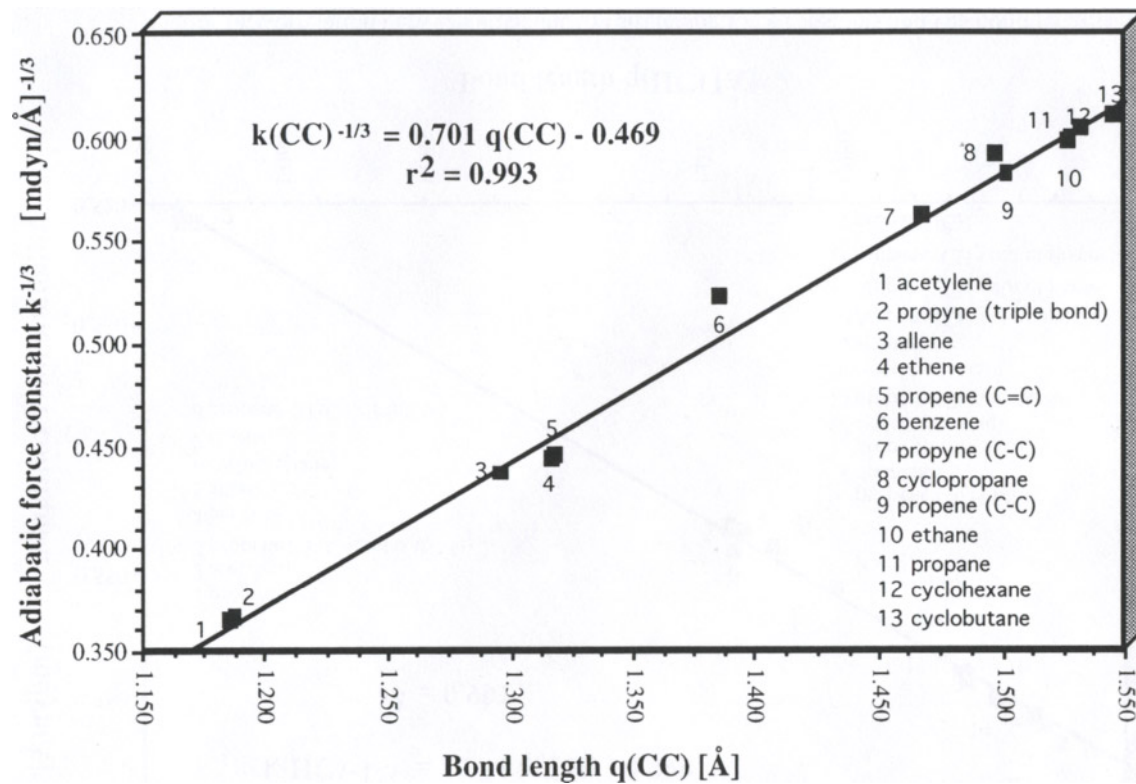


Figure 15. Correlation between adiabatic CC stretching force constants and CC bond lengths according to Badger. (HF/6-31G(d,p) calculations)

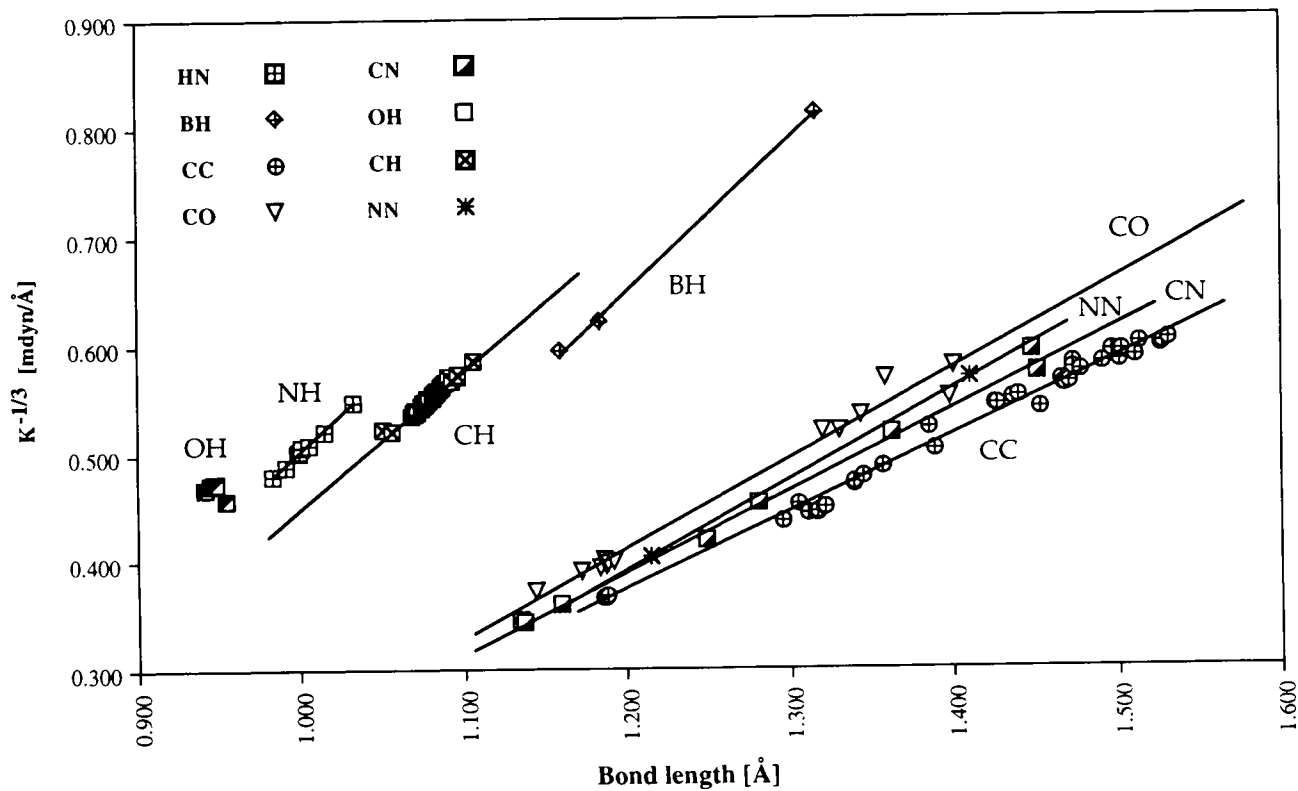


Figure 16. Correlation between adiabatic bond stretching force constants and bond lengths for polyatomic molecules. The correlation lines are parallel for AH bonds and AB bonds as anticipated by Badger. (HF/6-31G(d,p) calculations)

However, Badger's rule is fully confirmed by the adiabatic force constants of polyatomic molecules, which explains its usefulness even for today's research.

14. INTENSITIES OF ADIABATIC INTERNAL MODES

Once internal modes are defined, it is also possible to define the infrared intensity of these modes. For a normal mode, the infrared intensity is calculated with the help of the dipole moment derivatives. The dipole derivatives $\partial\mu/\partial x$ with regard to Cartesian coordinates can be determined in the course of an *ab initio* calculation of vibrational frequencies. The corresponding dipole derivatives with regard to normal coordinates Q_μ are obtained by the transformation

$$\left(\frac{\partial\mu}{\partial Q_\mu}\right) = \left(\frac{\partial\mu}{\partial x}\right) I_\mu \quad (83)$$

where the normal mode vector I_μ relates Cartesian coordinates x to the normal coordinate Q_μ according to Eq. (5).

The matrix of dipole derivatives with regard to normal coordinates $\partial\mu/\partial Q$ contains the derivatives of the dipole moment components with regard to each normal coordinate. The infrared intensity I_μ of the normal mode I_μ is calculated according to Eq. (84)

$$I_\mu = C \sum_{x=1}^3 \left(\frac{\partial\mu_x}{\partial Q_\mu}\right)^2 = g_\mu \frac{\pi N_0}{3 c^2} \sum_{x=1}^3 \left(\frac{\partial\mu_x}{\partial Q_\mu}\right)^2 \quad (84)$$

where C is a conversion factor from atomic units to km/mol that is given by the degeneracy g_μ of normal mode I_μ , the Avogadro number N_0 and the speed of light c .

In a similar way as dipole derivatives with respect to normal coordinates are obtained from dipole derivatives with regard to Cartesian coordinates, one can also obtain dipole derivatives with respect to the internal coordinates associated with the adiabatic internal modes:

$$\left(\frac{\partial\mu}{\partial q_n}\right) = \left(\frac{\partial\mu}{\partial x}\right) a_n \quad (85a)$$

$$\left(\frac{\partial\mu}{\partial q}\right) = \left(\frac{\partial\mu}{\partial x}\right) \mathbf{A} \quad (85b)$$

where \mathbf{A} contains the adiabatic mode vectors a_n from Eq. (43) and connects the internal coordinates q with the Cartesian coordinates x by

$$x = \mathbf{A} q \quad (86)$$

Once adiabatic dipole derivatives are known, the infrared intensity of an adiabatic mode \mathbf{a}_n associated with the internal coordinate q_n is calculated in a similar way as that for a normal mode I_μ :

$$I_n = C \sum_{x=1}^3 \left(\frac{\partial \mu_x}{\partial q_n} \right)^2 \quad (87)$$

The dipole moment of a bond is given by $\mu = q p$, where in this case q is the equilibrium bond length and $\pm p$ defines the partial charges at the atoms connected by the bond. Hence, the derivative of μ with regard to the bond length q should lead to the partial charges p at the atoms linked by the bond, i.e. the infrared intensity of the internal mode should provide a measure for the partial charges of the atoms of a molecule. However, as has been discussed by Zerbi and co-workers [50], one has to consider also the charge flux $\partial p_\alpha / \partial q_n$ toward or away from atom α caused by the stretching of the bond length q_n during a vibration of the bond. If $\partial p_\alpha / \partial q_n < 0$, the flux is directed toward atom α , otherwise away from atom α during a bond stretching vibration that according to the discussion in the previous chapters is best described by \mathbf{a}_n . The quantity $\partial p_\alpha / \partial q_n$ measures the deformability of the charge and also provides insight into the electronic nature of the bond in question.

For example, the intensity of an adiabatic CH stretching mode in a hydrocarbon is related to charge p_α and charge flux $\partial p_\alpha / \partial q_n$ according to Eq. (88):

$$I_n(\text{CH}) \propto p_H + \frac{\partial p_H}{\partial q(\text{CH})} q(\text{CH}) \quad (88)$$

If infrared intensities of bond stretching vibrations are known either from experiment or from theory, atomic charges can be derived. In Table 11, intensity based C and H charges of some simple hydrocarbons are compared with the corresponding Mulliken and virial charges [51]. Also, average intensities per CH bond that have been used by Zerbi to apply Eq. (88) are compared with adiabatic mode intensities. An average intensity per CH bond, e.g., for ethane is obtained by summing the intensities of the three infrared active vibrational modes of ethane and, then, dividing the sum of the intensities by the number of CH bonds in ethane. In Table 11, experimental intensities I_μ/CH obtained in this way are listed together with the corresponding calculated values. The latter as well as all other computed values have been obtained at the HF/6-31G(d,p) level of theory while the experimental data are from Zerbi's work [50].

Before the data of Table 11 are shortly discussed, one has to stress that the partial charges derived from adiabatic infrared intensities are not related to Mulliken charges, virial charges or most other atomic charges used in *ab initio* theory. The partial charges p are effective charges which in addition to the atomic monopole contribution, cover the atomic dipole contribution as well. They are

Table 11
Comparison of infrared intensity based, Mulliken, and virial partitioning based partial atomic charges. ^a

Quantity		CH ₄	C ₂ H ₆	c-C ₃ H ₆	C ₂ H ₄	C ₂ H ₂	Ref
Intensities [km/mol]							
I _μ /CH, exp.		17.4	28.5	11.5	9.6	35.2	50
I _μ /CH, cal.		9.9	28.7	15.2	15.2	46.0	t.w.
I _n /CH, diab.		23.4	39.3	20.0	13.3	42.8	t.w.
CH Bond lengths [Å]							
q(CH), cal.		1.0835	1.0858	1.0760	1.0764	1.0568	t.w.
Charges [electron]							
exp. Intensity	C	-0.260	-0.135	-0.170	-0.268	-0.208	50
	H	0.065	0.045	0.085	0.134	0.208	50
adiab. Intensity	C	-0.290	-0.102	-0.187	-0.240	-0.185	t.w.
	H	0.072	0.034	0.094	0.120	0.185	t.w.
Mulliken	C	-0.472	-0.335	-0.261	-0.254	-0.233	28
	H	0.118	0.112	0.130	0.127	0.233	28
virial charges	C	0.244	0.237	0.104	0.082	-0.121	51
	H	-0.061	-0.079	-0.052	-0.041	0.121	51

^a Calculated values based on HF/6-31G(d,p)//HF/6-31G(d) calculations.

related to those effective charges which have been determined by Zerbi and co-workers [51] from measured intensities. This is confirmed by the fact that the effective charges determined by Zerbi are parallel to the charges based on adiabatic mode intensities (Table 11). Clearly, the average intensity per CH bond is not equal to the adiabatic mode intensity where the differences can be 10-15 km/mol. It is easy to see that an averaging of CH intensities cannot provide reliable intensity values for the determination of atomic charges and that adiabatic mode intensities provide an attractive alternative to average intensities.

It is well known that the electronegativity of a C atom increases with increasing *s*-character, which is nicely reflected by the virial charges listed in Table 11. The only problem is that virial charges suggest a C⁺-H⁻ bond polarity while Mulliken charges and intensity based charges predict a C⁻-H⁺ bond polarity. The H charges derived from (both experimental and calculated) infrared intensities seem to confirm the increase in the electronegativity of the C atom with increasing *s*-character. However, the corresponding C charges reveal that the electronegativity change from ethene to acetylene is not correctly described and that a large electronegativity difference between cyclopropane and ethene is predicted. This is not necessarily an indication that the intensity based charges are ill-defined.

As mentioned above, they absorb the effects of (true) atomic charges and atomic dipole moments, where the latter result from the anisotropy of the electron density at an atom. In the virial partitioning method, atomic charges and atomic dipole moments (multipole moments) are separately calculated and their values may cancel largely in the expression for the bond dipole moment. Hence, effective atomic charges and true atomic charges can differ considerably where of course it should be more difficult to discuss effective charges since they contain the cumulative effect of at least two quantities. It is interesting to note that Mulliken charges also do not reproduce the increase in the C electronegativity when going from ethene to acetylene. This might result from an equal splitting of overlap populations to get Mulliken charges, which may mix into the atomic charges higher multipole contributions and, accordingly, may give Mulliken charges the character of effective rather than pure atomic charges.

It is interesting to note that adiabatic intensity based charges in agreement with Mulliken and virial charges suggest similar hybridizations for cyclopropane and ethene as far as the CH hybrid orbitals are concerned. This is in line with other observations, e.g., made for CH dissociation energies. Effective charges derived from average values of experimental intensities fail to describe the close relationship of the CH bonds in ethene and cyclopropane.

We conclude that the adiabatic mode intensities and effective charges derived from them are the localized counterparts of those effective charges derived from measured intensities. They should be more appropriate for the description of the properties of individual bonds. In particular, they should lead to chemically more meaningful effective charges where future work has to show how effective charges, atomic monopole and dipole contributions, and the charge flux are related.

15. INVESTIGATION OF REACTION MECHANISM WITH THE HELP OF THE CNM ANALYSIS

While the adiabatic mode analysis was discussed in the previous sections exclusively for molecules in their equilibrium geometry, we will show in this section that adiabatic vibrational modes are also useful when describing molecules during a chemical reaction. For this purpose, we extend the procedure previously described for constructing adiabatic modes at equilibrium points of the potential energy surface to points along the reaction path [22,23].

The reaction path is defined by the line $\bar{\mathbf{x}}(s)$ where $\bar{\mathbf{x}}(s)$ is a column vector of 3K mass-weighted Cartesian coordinates \mathbf{x}_i . The reaction path is given parametrically in terms of its arc length s defined by the differential

$$ds^2 = d\mathbf{x}^T \mathbf{M} d\mathbf{x} = d\bar{\mathbf{x}}^T d\bar{\mathbf{x}} \quad (89)$$

with \mathbf{M} being the diagonal matrix of nuclear masses. The reaction path can be calculated using standard *ab initio* methods and reaction path following algorithms [52]. One starts at the transition state and follows in the forward and backward direction the path downhill to products and reactants, respectively, by evaluating at fixed points along the path gradient and Hessian matrix, which are used to determine at these path points the path direction. It is of advantage to calculate the 3K-L-1 vibrational modes orthogonal to the path direction and use them to describe the reaction valley. This is done by diagonalizing the mass-weighted projected force constant matrix $\tilde{\mathbf{K}}^g(s)$ given by Eq. (90) [53,54]:

$$\tilde{\mathbf{K}}^g(s) \tilde{\mathbf{I}}_\mu^g(s) = \omega_\mu^2(s) \tilde{\mathbf{I}}_\mu^g(s), \quad (90)$$

where $\tilde{\mathbf{K}}^g(s)$ is defined by Eq. (91).

$$\tilde{\mathbf{K}}^g(s) = (\mathbf{I} - \tilde{\mathbf{P}}(s)) \tilde{\mathbf{f}}(s) (\mathbf{I} - \tilde{\mathbf{P}}(s)) \quad (91)$$

In Eq. (91), $\tilde{\mathbf{f}}(s)$ is the mass-weighted Cartesian coordinate force constant matrix, and $\mathbf{I} - \tilde{\mathbf{P}}(s)$ is a projector onto the (3K-L)-1-dimensional subspace of the normal mode vibrations orthogonal to the reaction path mode [53,54]. These modes are called generalized normal modes and describe a "harmonic" reaction valley according to Eq. (92) (compare with Eq. 34).

$$V(s, \mathbf{Q}) = V(s) + \frac{1}{2} \sum_{\mu=1}^{N_{\text{nm}}} k_\mu^g(s) [Q_\mu^g(s)]^2 \quad (92)$$

where $k_\mu^g(s)$ is the generalized normal mode force constant, $Q_\mu^g(s)$ the generalized normal mode coordinate and $V(s)$ the energy profile along the reaction path.

To describe energy transfer along the reaction path, curvature vector $\mathbf{K}(s)$, curvature coupling elements $B_{\mu,s}(s)$ and mode-mode coupling elements $B_{\mu,\nu}(s)$ have to be calculated [53,54], of which only the former will be discussed here. The mass-weighted curvature vector $\tilde{\mathbf{K}}(s)$ is defined by (93a) and its Euclidean norm by (93b).

$$\tilde{\mathbf{K}}(s) = \frac{d^2 \tilde{\mathbf{x}}(s)}{ds^2} \quad (93a)$$

$$\kappa(s) = \sqrt{\tilde{\mathbf{K}}(s)^* \tilde{\mathbf{K}}(s)} \quad (93b)$$

The curvature coupling elements $B_{\mu,s}(s)$, which represent coefficients of the expansion of the curvature vector in terms of generalized normal modes $\tilde{\mathbf{I}}_{\mu}^*(s)$, are defined by Eq. (94):

$$B_{\mu,s}(s) = \tilde{\mathbf{K}}(s)^* \tilde{\mathbf{I}}_{\mu}^*(s); \quad (94)$$

It is common practice to graphically present the norm of the curvature vector, $\kappa(s)$, and to discuss energy transfer along the reaction path in terms of the maxima of $\kappa(s)$ [55]. Maximal values of $\kappa(s)$ indicate those points on the path where energy can flow from the motion along the reaction path into one of the transverse normal vibrational modes or vice versa thus decreasing or increasing the reaction rate. The curvature coupling coefficients $B_{\mu,s}(s)$ of Eq. (94) determine how much energy is transferred into (retrieved from) normal mode $\tilde{\mathbf{I}}_{\mu}^*(s)$. Due to the delocalized character of normal modes, it is difficult to identify substituents or molecular fragments, which by their vibrations are predominantly responsible for energy transfer from the reaction path mode into vibrational modes (rate reduction) or alternatively can be used to channel external energy *via* vibrational modes into the reaction path mode (rate enhancement). Therefore, it is desirable to express the curvature coupling coefficients $B_{\mu,s}(s)$ of Eq. (94) in terms of vibrational modes that can be directly associated with chemically relevant molecular fragments or structural units. Such modes are the generalized adiabatic internal modes \mathbf{a}_n , that can be defined by requiring that the harmonic part of the energy in Eq. (92) has to be minimized with regard to displacements in the (3K-L)-1-dimensional vibrational space (rather than the (3K-L)-dimensional space as originally defined) while relaxing all internal parameters but one [22].

Eq. (95) gives the conditions for obtaining generalized adiabatic internal modes $\mathbf{a}_n^g(s)$ [22]:

$$V(\mathbf{Q},s) = \min \quad (95a)$$

$$s = \text{const} \quad (95b)$$

$$q_n(s,\mathbf{Q}) = q_n^* \quad (95c)$$

where in first order the leading parameter q_n is some linear function of the normal mode coordinates, i.e. in the limit of infinitesimal displacements it is defined by Eq. (96):

$$q_n(s, \mathbf{Q}) = \sum_{\mu=1}^{N_{nb}} D_{n\mu}(s) Q_{\mu}(s) \quad (96)$$

$D_{n\mu}(s)$ denotes an element of a Wilson B-type matrix \mathbf{D} that connects normal coordinates with internal coordinates. Solving Eq. (95), generalized adiabatic internal modes and related force constants $k_n^a(s)$, mass $m_n^a(s)$, and frequency $\omega_n^a(s)$ are obtained by Eqs. (97) [22],

$$\left(\mathbf{a}_n(s)\right)_{\mu} = \frac{\frac{D_{n\mu}(s)}{k_{\mu}^g(s)}}{\sum_{\nu=1}^{N_{nb}} \frac{D_{n\nu}(s)^2}{k_{\nu}^g(s)}} \quad (97a)$$

$$k_n^a(s) = \frac{1}{\sum_{\nu=1}^{N_{nb}} \frac{D_{n\nu}(s)^2}{k_{\nu}^g(s)}} \quad (97b)$$

$$m_n^a(s) = \frac{1}{G_{nn}(s)} \quad (97c)$$

$$\omega_n^a(s) = \sqrt{\frac{k_n^a(s)}{m_n^a(s)}} \quad (97d)$$

Generalized adiabatic modes can be transformed from normal mode space into Cartesian coordinate space according to Eq. (98)

$$\left(\mathbf{a}_n^g(s)\right)_i = \sum_{\mu=1}^{N_{nb}} \left(\mathbf{I}_{\mu}(s)\right)_i \left(\mathbf{a}_n^g(s)\right)_{\mu} \quad i = 1, \dots, 3N, \quad (98)$$

where $(\mathbf{I}_{\mu})_i$ is component i of normal mode vector \mathbf{I}_{μ} in Cartesian coordinates.

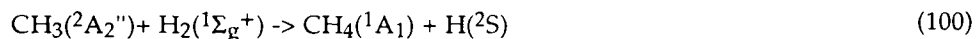
Eqs. (97) indicate that there is no difference in applying the adiabatic mode concept to an equilibrium geometry or to a point along a reaction path. In the latter case, the adiabatic modes are defined in a (3K-L)-1- rather than a 3K-L-dimensional space and all adiabatic properties are a function of the reaction coordinate s . Obviously, the adiabatic mode concept and the leading parameter principle have their strength in the fact that they can generally be applied to equilibrium geometries as well as any point on the reaction path.

Once generalized adiabatic modes $\mathbf{a}_n^g(s)$ have been defined, the normal modes and curvature vector can be analyzed utilizing the CNM approach of Section 7 [20,21]. For this purpose, the amplitude $A_{n,s}$ is defined [22]

$$A_{n,s} = \frac{\mathbf{K}(s)^+ \mathbf{M}(s) \mathbf{a}_n^g(s)}{\sqrt{\mathbf{a}_n^{g+}(s) \mathbf{M}(s) \mathbf{a}_n^g(s)}} \quad (99)$$

which characterizes the curvature vector $\mathbf{K}(s)$ in terms of generalized adiabatic modes associated with internal coordinates used to describe the reaction system. It corresponds to the A-type amplitude AvAM (with metric \mathbf{M} , see Eq. 69), which was found to present the best choice for kinetically characterizing normal modes in terms of adiabatic modes in the case of molecules in their equilibrium geometries [20,21]. Eq. (62) ensures that $A_{n,s}$ has the same dimension as $B_{\mu,s}$ and, for $l_{\mu}^g = \mathbf{a}_n^g$, amplitude $A_{n,s}$ and coefficient $B_{\mu,s}$ are equal. Both curvature vector and normal modes orthogonal to the reaction path can be characterized in terms of generalized adiabatic internal modes, however for the latter the A-type amplitude AvAF (metric \mathbf{f} , see Eq. 70) is used since for these modes the dynamic characterization is more important than a kinetic one.

The generalized adiabatic internal modes are essential for the unified reaction valley analysis (URVA) developed by Konkoli, Kraka, and Cremer to investigate reaction mechanisms and reaction dynamics [22,52]. As an example for the application of the generalized adiabatic internal modes, the hydrogenation reaction of the methyl radical is shortly discussed here:



which has recently been investigated at the MP2/6-31G(d,p) level of theory [22,52]. In Figure 17, the internal coordinates q_n used to describe the reaction complex are given. The most important internal coordinates are $q_1 = R_1$ and $q_2 = R_2$, which describe the length of the breaking HH bond and the length of the CH distance to be formed, respectively. The calculated dependence of the normal mode frequencies $\omega_{\mu}(s)$ in dependence of s is shown in Figure 18. The latter reveals that the strongest changes in the mode frequencies are observed for modes #11 and #8, which accordingly should closely be connected with the bond breaking/bond forming process of reaction (100). Noteworthy is an avoided crossing point at $s = -0.3 \text{ amu}^{1/2} \text{ Bohr}$ involving the a_1 -symmetrical modes #11 and #8 (notation 11/8) and a reaction path bifurcation point at $s = 0.4 \text{ amu}^{1/2} \text{ Bohr}$ that leads to zero and, then, imaginary values of the frequencies of the $1e$ -symmetrical modes #1 and #2 (see Figure 18).

In Figure 19a, the reaction path curvature $\kappa(s)$ is shown as a function of the reaction coordinate s . There are two distinct peaks of $\kappa(s)$ in the transition state (TS; the location of the TS is defined by $s = 0$) region at $s = -0.1$ and $0.7 \text{ amu}^{1/2} \text{ Bohr}$ (peaks κ_2 and κ_3), which are associated with the normal modes 11/8 (i.e. #11 before and #8 after the avoided crossing at $s = -0.3$) and to some smaller extend with modes #5 and 8/11 as the decomposition of $\kappa(s)$ in terms of normal mode contributions reveals. If energy is stored in mode 11/8, it will be channelled into the reaction path mode and lead to rate acceleration. Dissipation of energy into mode 8/11 is small since the avoided crossing between modes #11 and #8 at $s =$

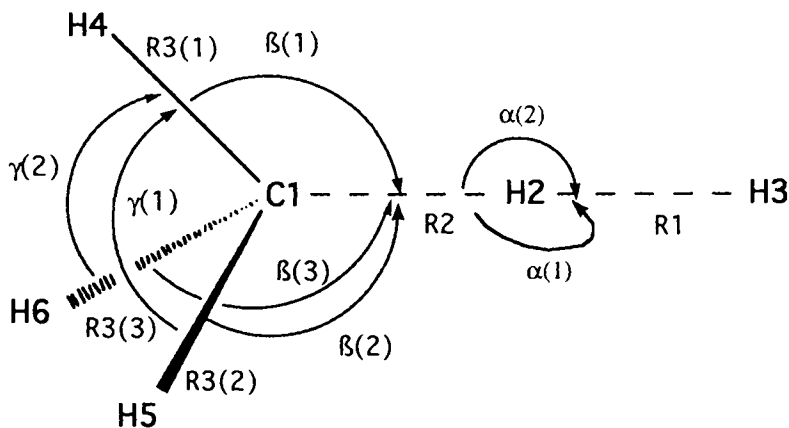


Figure 17. Internal coordinates used to describe the reaction complex of the hydrogenation reaction $\text{CH}_3(^2A_2'') + \text{H}_2(^1\Sigma_g^+) \rightarrow \text{CH}_4(^1A_1) + \text{H}(^2S)$.

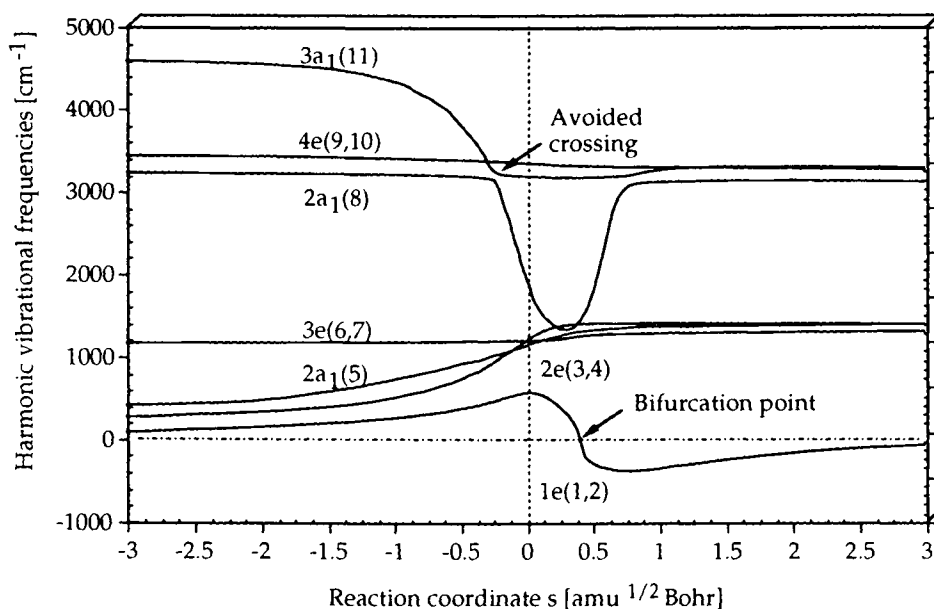


Figure 18. Representation of normal mode frequencies $\omega_\mu(s)$ along the reaction path. Symmetry symbols and numbering of normal modes are given according to the order of normal modes calculated for the reactants. The position of the transition state corresponds to $s = 0 \text{ amu}^{1/2} \text{ Bohr}$ and is given by a vertical line. The value $\omega_{1e}(s) = 0$ indicates the location of a bifurcation point ($s = 0.4 \text{ amu}^{1/2} \text{ Bohr}$) of the reaction path. Imaginary $1e$ frequencies calculated for $s > 0.4 \text{ amu}^{1/2} \text{ Bohr}$ are given as negative numbers.

$-0.3 \text{ amu}^{1/2} \text{ Bohr}$ is strongly localized and, therefore, the exchange of energy rather limited.

In the case of reaction (100), it is easy to determine the nature of modes #11 (HH bond stretching) and #8 (CH bond stretching) and, in this way, to relate the peaks of the curvature vector with the corresponding changes in electronic structure. However, in general this way of analysis is difficult and, therefore, a decomposition of the curvature vector in terms of generalized adiabatic internal modes gives a chemically more meaningful insight into the nature of the reaction path curvature as can be seen from Figure 19b.

The analysis of the reaction path curvature in terms of generalized adiabatic modes reveals that peak κ_2 close to the TS is strongly dominated by the R1 adiabatic HH stretching mode led by internal coordinate R1 while peak κ_3 after the TS results from the adiabatic CH stretching mode led by internal coordinate R2. Since the motions associated with R1 and R2 are closely related in the reaction, each of the two peaks κ_2 and κ_3 also depends to some smaller degree on the adiabatic partner mode of the pair R1-R2. This duality is indicative for the fact that HH bond breaking and CH bond forming are closely connected in the reaction. When the R1-peak of $\kappa(s)$ (peak κ_2) starts to develop the HH bonds begins to break and the CH bond to be formed (Figure 19b). The positive R1-contribution to κ_2 is accompanied by a negative, but much smaller R2 contribution, which can be interpreted as indication that the reaction system resists a further decrease in R2 needed for the formation of the CH bond.

From the second to the third curvature maximum at $s = 0.6 \text{ amu}^{1/2} \text{ Bohr}$ (peak κ_3) the R2 and R1 amplitudes $A_{n,s}$ exchange their role, i.e. the R2 amplitude becomes dominant and positive while the R1 amplitude is relatively small and negative. Peak κ_3 identifies the point where the CH bond forming process is basically finished if the reaction $\text{CH}_3 + \text{H}_2$ is considered; for the reverse reaction $\text{CH}_4 + \text{H}$, it is the point where bond C1H2 starts to be broken accompanied by the resistance of the electronic structure to form as a new bond the HH bond associated with R1.

If one investigates the changes in $\kappa(s)$ (Figure 19b), a clear picture of the mechanism of the HH bond breaking and CH bond forming process emerges: These processes occur simultaneously in the region of the curvature peaks κ_2 and κ_3 ($-0.1 < s < 0.6$) as indicated by maxima or minima of the amplitudes associated with the internal parameters R1 and R2 describing HH and CH bond. Hence, the generalized adiabatic modes help to understand the mechanism of bond breaking and bond forming and, therefore, they are essential for UVAR [22]. Their real value becomes obvious when investigating larger reaction systems. For example, the application of UVAR to the Diels-Alder reaction between ethene and butadiene implies an analysis of reaction path direction and reaction path curvature in terms of 42 normal modes, which represents a rather complicated and chemically complex task difficult to interpret [53]. However, use of generalized adiabatic modes directly clarifies, which structural changes determine direction and curvature of the reaction path [53].

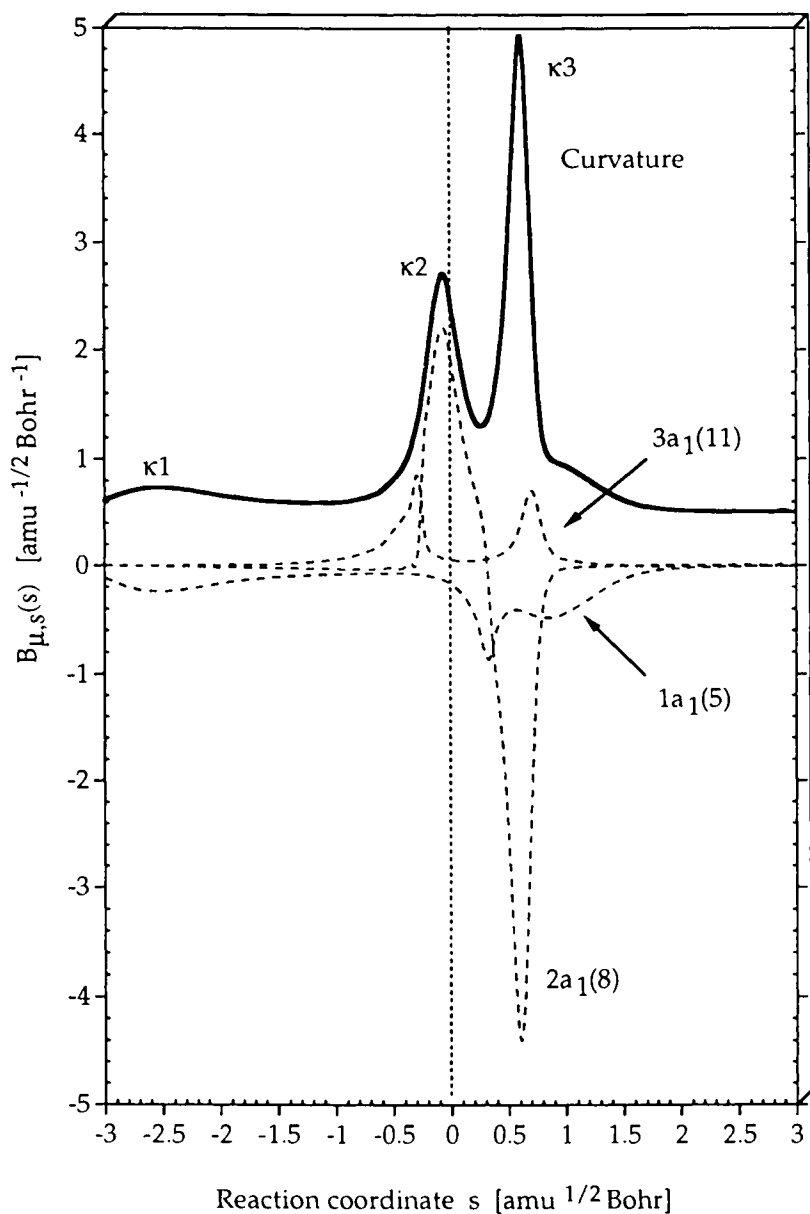


Figure 19a. Characterization of the reaction path curvature $\kappa(s)$ (thick solid line) in terms of normal mode-curvature coupling coefficients $B_{\mu,s}(s)$ (dashed lines). The curve $\kappa(s)$ has been shifted by 0.5 units to more positive values to facilitate the distinction between $\kappa(s)$ and $B_{\mu,s}(s)$. For a definition of the internal coordinates, compare with Figure 17. The position of the transition state corresponds to $s = 0$ $\text{amu}^{1/2} \text{Bohr}$ and is indicated by a vertical line.

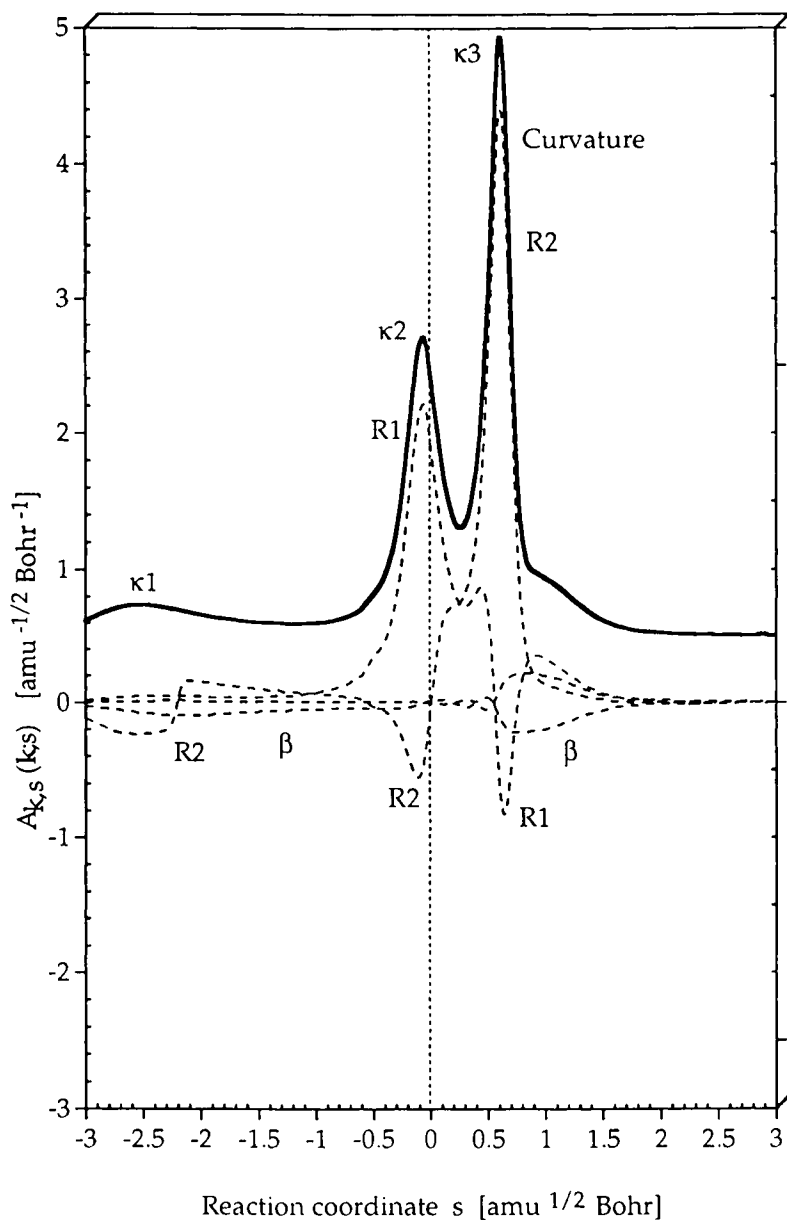


Figure 19b. Characterization of the reaction path curvature $\kappa(s)$ (thick solid line) in terms of adiabatic mode-curvature coupling amplitudes $A_{n,s}(s)$ (dashed lines). The curve $\kappa(s)$ has been shifted by 0.5 units to more positive values to facilitate the distinction between $\kappa(s)$ and $A_{n,s}(s)$. For a definition of the internal coordinates, compare with Figure 17. The position of the transition state corresponds to $s = 0 \text{ amu}^{1/2} \text{ Bohr}$ and is indicated by a vertical line.

16. CONCLUSIONS

One of the major goals of vibrational spectroscopy is to associate measured frequencies with structural features of a molecule and, thereby, to facilitate its identification. These efforts have led to a number of rules that concern the similarity and transferability of force constants and frequencies from one molecule to another provided they contain similar structural units [1-9]. To provide a mathematical basis for the comparison of measured vibrational frequencies and force constants, the adiabatic internal vibrational modes were defined [18], which enable one to investigate molecular fragments in terms of their internal vibrations defined by the pair (q_n, \mathbf{v}_n) .

The derivation of the adiabatic vectors has been motivated by the observation that the masses of the atoms of a molecule effectively hinder the appearance of localized internal vibrations \mathbf{v}_n associated with fragments ϕ_n . However, localized internal vibrations \mathbf{v}_n can be obtained by setting the generalized momenta associated with those internal parameters not used for the description of fragment ϕ_n to zero and solving the Euler-Lagrange equations under this condition. This approach is equivalent to exciting the internal motion \mathbf{v}_n by a constant perturbation q_n^* of the leading parameter associated with ϕ_n and, then, relaxing the distortions of all other internal coordinates q_m until a minimum of the energy is obtained.

Once adiabatic internal modes \mathbf{a}_n are defined (see Section 5), it is straightforward to derive an appropriate adiabatic force constant k_n , an adiabatic mass m_n , and an adiabatic mode frequency ω_n (see Section 6). The choice of m_n as an appropriate fragment mass is confirmed by the fact that it represents a generalized reduced mass $1/G_{nn}$. Furthermore, it guarantees that a fragment frequency does not depend on the masses of those atoms that do not belong to ϕ_n and, therefore, it is typical of ϕ_n and its properties. The dynamics of a vibrating molecular fragment ϕ_n is uniquely characterized by the internal frequency ω_n , the internal mass m_n , and the internal force constant k_n and this makes it possible to compare different molecular fragments of one or many polyatomic molecules in a systematic way.

There are immediately a number of applications of adiabatic internal modes that lead to a new dimension in the analysis of vibrational spectra. For example, the adiabatic vectors \mathbf{a}_n are perfectly suited to present a set of localized internal modes that can be used to analyze delocalized normal modes. This has led to the CNM analysis (Sections 7 and 8) of calculated vibrational spectra of molecules as was discussed in Section 9. With the CNM analysis it is rather easy to correlate the vibrational spectra of different molecules (Section 10). With the help of perturbation theory and calculated normal modes, the determination of adiabatic modes and the CNM analysis can be extended to experimental spectra (Section 12).

Once adiabatic modes are known either from calculations or experimental data, adiabatic frequencies can be used to characterize chemical bonds. For example, it is easy to verify a McKean relationship [30] between adiabatic CH or CC stretching frequencies and the corresponding bond lengths (Section 11). It can

be shown that with the adiabatic force constants k_n Badger's rule can be extended from diatomic to polyatomic molecules (Section 13). In addition, it is possible to determine adiabatic mode intensities, which can be utilized to derive effective charges for the atoms of a molecule (Section 14). Most important, generalized adiabatic vibrational modes can be defined for reacting molecules so that a detailed analysis of the direction and the curvature of the reaction path becomes possible. This is the basis of the UVAR approach [22,23], which leads to a detailed analysis of mechanism and dynamics of chemical reactions (Section 15). A chemically easy to understand description of energy transfer and energy dissipation, quantum mechanical tunneling, structural and electronic changes, etc. occurring along the reaction path can be made, which provides new and very detailed insights into chemical reactions.

ACKNOWLEDGEMENT

This work was supported by the Swedish Natural Science Research Council (NFR). All calculations were done on a CRAY YMP/416 and C94 of the Nationellt Superdatorcentrum (NSC), Linköping, Sweden. The authors thank the NSC for a generous allotment of computer time. Useful comments by Zoran Konkoli are appreciated.

REFERENCES

1. E.B.J. Wilson, J.C. Decius, and P.C. Cross, *Molecular Vibrations, The Theory of Infrared and Raman Vibrational Spectra*, McGraw-Hill, London, 1955.
2. G. Herzberg, *Infrared and Raman Spectra of Polyatomic Molecules*, Van Nostrand, New York, 1945.
3. P. Gans, *Vibrating Molecules*, Chapman and Hall, London, 1971.
4. L.A. Woodward, *Introduction to the Theory of Molecular Vibrations and Vibrational Spectroscopy*, Clarendon Press, Oxford, 1972.
5. S. Califano, *Vibrational States*, Wiley, New York, 1976.
6. D.A. Long, *Raman Spectroscopy*, McGraw-Hill, London, 1977.
7. K. Nakanishi, and P.H. Solomon, *Infrared Absorption Spectroscopy*, Holden-Day, San Francisco, 1977.
8. N.B. Colthup, L.N. Daly, and S.E. Wilberley, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, New York, 1990.
9. J.M. Hollas, *High Resolution Spectroscopy*, Butterworths, London, 1982.
10. P. Pulay, in *Ab initio Methods in Quantum Chemistry, Part II*, K.P. Lawley (ed.), Wiley, New York, 1987, p. 241.
11. B.A. Hess and L.J. Schaad, *Chem Rev.*, 86 (1986) 709.
12. J. Gauss and D. Cremer, *Adv. Quant. Chem.*, 23 (1992) 205.
13. L. Andrews and M. Moskovits (eds.), *Chemistry and Physics of Matrix-Isolated Species*, North-Holland, Amsterdam, 1989.
14. W. Sander, *Angew. Chem., Int. Ed. Engl.*, 29 (1989) 344.

15. (a) A. Patyk, W. Sander, J. Gauss, and D. Cremer, *Angew. Chem.*, 101 (1989) 920. (b) A. Patyk, W. Sander, J. Gauss, and D. Cremer, *Chem. Ber.*, 123, (1990) 89.
16. (a) W. Sander, G. Bucher, F. Reichel, and D. Cremer, *J. Am. Chem. Soc.*, 113 (1991) 5311. (b) M. Trommer, W. Sander, C.-H. Ottosson, and D. Cremer, *Angew. Chem.*, 107 (1995) 999 .
17. (a) S. Wierlacher, W. Sander, C. Marquardt, E. Kraka, and D. Cremer, *Chem. Phys. Lett.*, 222 (1994) 319. (b) R. Albers, W. Sander, H. Ottosson and D. Cremer, *Chem. Eur. J.*, 2 (1996) 967.
18. Z. Konkoli and D. Cremer, *Int. J. Quant. Chem.*, submitted.
19. Z. Konkoli, J.A. Larsson, and D. Cremer, *Int. J. Quant. Chem.*, submitted.
20. Z. Konkoli and D. Cremer, *Int. J. Quant. Chem.*, submitted.
21. Z. Konkoli, J.A. Larsson, and D. Cremer, *Int. J. Quant. Chem.*, submitted.
22. Z. Konkoli, E. Kraka, and D. Cremer, *J. Phys. Chem.*, in press.
23. Z. Konkoli, E. Kraka, and D. Cremer, *J. Comp. Chem.*, in press.
24. (a) N. Neto, *Chemical Physics*, 91 (1984) 89; 101. (b) N. Neto, *Chemical Physics*, 87 (1984) 43.
25. Y. Morino and K. Kuchitsu, *J. Chem. Phys.*, 20 (1952) 1809.
26. P. Pulay and F. Török, *Acta Chim. Hung.*, 47 (1966) 273.
27. G. Keresztury and G. Jalsovszky, *J. Mol. Structure*, 10 (1971) 304.
28. D. Cremer, E. Kraka and K.J. Szabo, in: *The Chemistry of Functional Groups, The Chemistry of the Cyclopropyl Group, Vol 2*, Z. Rapport (ed), J. Wiley, Ltd., New York, 1995, p.43.
29. (a) D. F. McMillen and D. M. Golden, *Am. Rev. Phys. Chem.*, 33 (1982) 493. (b) M. H. Baghal-Vayjooee and S. Benson, *J. Am. Chem. Soc.*, 101 (1979) 2840. (c) W. Tsang, *J. Am. Chem. Soc.*, 107 (1985) 2872.
30. (a) D.C. McKean, *Chem. Soc. Rev.*, 7 (1978) 399. (b) D.C. McKean, *Int. J. Chem. Kinet.*, 71 (1984) 445.
31. S.-J. Kim, H.F. Schaefer, E. Kraka and D. Cremer, *Mol. Phys.*, 88 (1996) 93.
32. E. Kraka, D. Cremer, J. Fowler, and H.F. Schaefer, *J. Am. Chem. Soc.* 118 (1996) 10595.
33. B. Casper, D. Christen, H.-G. Mack, H. Oberhammer, G. A. Argiello, B. Jülicher, M. Kronberg, and H. Willner, *J. Phys. Chem.*, 100 (1996) 3983.
34. (a) K. C. Nicolaou and W. M. Dai, *Angew. Chem. Int. Ed. Engl.*, 30 (1991) 1387. (b) K. C. Nicolaou and A. L. Smith, *Acc. Chem. Res.*, 25 (1992) 497.
35. (a) R. Gleiter and D. Kratz *Angew. Chem.*, 105 (1993) 884. (b) P. Chen, *Angew. Chem.*, 108 (1996) 1584.
36. Kraka, E. and D. Cremer, *Chem. Phys. Lett.*, 216 (1993) 333.
37. Kraka, E. and D. Cremer, *J. Am. Chem. Soc.*, 116 (1994) 4929.
38. R. Marquardt, W. Sander, and E. Kraka, *Angew. Chem. Int. Ed. Engl.*, 35 (1996) 746.
39. R. Marquardt, W. Sander, E. Kraka, and D. Cremer, *Angew. Chem.*, submitted.
40. (a) G. Bucher, W. Sander, E. Kraka, and D. Cremer *Angew. Chem. Int. Ed Engl.*, 31 (1992) 1230. (b) E. Kraka, D. Cremer, G. Bucher, and W. Sander,

- Chem. Phys. Lett., in press. (c) W. Sander, G. Bucher, H. Wandel, A. Kuhn, E. Kraka, and D. Cremer, *J. Am. Chem. Soc.*, submitted.
41. E. Kraka, J.A. Larsson, and D. Cremer, *J. Phys. Chem.*, to be published.
 42. J.A. Larsson and D. Cremer, to be published.
 43. K. J. Szabo, E. Kraka, and D. Cremer, *J. Org. Chem.*, 61 (1996) 2783.
 44. See, e.g., the discussion in Ref. 3.
 45. For a collection of experimental frequencies, see W.J. Hehre, L. Radom, P.v.R. Schleyer, and J.A. Pople, *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986.
 46. A.A. Zavitsas, *J. Phys. Chem.*, 91 (1987) 5573.
 47. D.J. Swanton, and B.R. Henry, *J. Chem. Phys.*, 86 (1987) 4801.
 48. *Handbook of Chemistry and Physics*, 64th edition, Weast, 1984.
 49. (a) R.M. Badger, *J. Chem. Phys.*, 2 (1934) 128. (b) R.M. Badger, *J. Chem. Phys.*, 3 (1935) 552. (c) R.M. Badger, *Phys. Rev.*, 48 (1935) 284.
 50. (a) C. Castiglioni, M. Gussoni and G. Zerbi, *J. Mol. Struct.*, 141 (1986) 341. (b) M. Gussoni, C. Castiglioni and G. Zerbi, *J. Mol. Struct. THEOCHEM*, 138 (1986) 203. (c) M. Gussoni, C. Castiglioni, M. N. Ramos, M. Rui and G. Zerbi, *J. Mol. Struct.*, 224, (1990) 445 and references cited therein.
 51. (a) R. F. W Bader and T. T. Nguyen-Dang, *Adv. Quantum Chem.*, 14 (1981) 63. (b) R. F. W. Bader, T. T. Nguyen-Dang and Y. Tal, *Rep. Prog. Phys.*, 44 (1981) 893. (c) R. F. W. Bader, *Atoms in Molecules - A Quantum Theory*, Oxford University Press, Oxford, 1990. (d) R. F. W. Bader, P. L. A. Popelier and T. A. Keith, *Angew. Chem.*, 106 (1994) 647
 52. For a recent review, see E. Kraka in *Encyclopedia of Computational Chemistry*, H.F. Schaefer III (ed.), Wiley, submitted.
 53. W.H. Miller, N.C. Handy, and J.E. Adams, *J. Chem. Phys.*, 72 (1980) 99.
 54. M. Page and J.W.J. McIver, *J. Chem. Phys.*, 88 (1988) 922.
 55. S. Kato and K. Morokuma, *Chem. Phys.*, 73 (1980) 3900.
 56. T. Johnsson, Z. Konkoli, E. Kraka, and D. Cremer, to be published.

This Page Intentionally Left Blank

Atomistic Modeling of Enantioselection: Applications in Chiral Chromatography

Kenny B. Lipkowitz

Department of Chemistry, Indiana University-Purdue University at Indianapolis
Indianapolis, IN 46202, USA, e-mail: lipkowitz@chem.iupui.edu

INTRODUCTION

Computational chemistry is a multidisciplinary area of science transcending mathematics, chemistry and physics. Although it is a relatively new field of study [1] it has had a major impact on most subdisciplines of chemistry [2]. Moreover, structural biologists, pharmacologists and toxicologists as well as materials scientists are now beginning to use computational chemistry in a routine manner to help rationalize research discoveries and, more importantly, to make predictions. An example of the utility of computational chemistry is computer assisted molecular design [3] (CAMD). Having its roots in the pharmaceutical industry, drug design groups were very successful in bringing new products to market quickly based on computation and theory [4]. Their successes have been noted by other industries and computational design groups now exist in both large and small companies alike, with products as diverse as agrochemicals, plastics, and ceramics. Clearly computational chemistry has established itself as a viable discipline and its uses and applications are expanding rapidly.

One area of research where computational chemistry is anticipated to be of benefit is in separation science. In particular, one could envision molecular modeling as a practical tool for enumerating and evaluating the complex interactions taking place between analyte and chromatographic stationary phase as analytes migrate through a chromatographic column, providing detailed information at the atomic level, and perhaps from first principles, about how chromatographic systems work. In this chapter we review the applications of computational chemistry in chromatography. We show the diversity of computational techniques that have been implemented to rationalize experimental retention orders and to discern where and how molecular recognition takes place on stationary phases used in chromatography. To provide a focus and make this task manageable, this chapter considers the controversial yet timely topic of chiral chromatography [5-13].

In the first section of this chapter a brief review of stereochemistry is provided along with a justification for why scientists need to separate enantiomers. The following section provides a brief review of the principles of chromatography with an emphasis on chiral chromatography. In the next section we provide a working definition of what molecular modeling means followed by a section describing the different kinds of commercially available stationary phases and how they work. The

remainder of the chapter is an examination of what has been accomplished in atomistic molecular modeling of the stationary phases used in chiral chromatography.

1. STEREOCHEMISTRY

The jargon and nomenclature associated with stereochemistry, an immense field of study unto itself, can be daunting to most computational chemists. Accordingly, only the most salient definitions and points of significance required for understanding the remainder of the chapter are summarized here. Isomers are different molecules with the same molecular formulae, and in this regard, there exist two major categories. The first are constitutional isomers. They are comprised of the same building block atoms, but with different topologies. In other words, they have the same number and kinds of atoms but are connected differently as in, say, $\text{CH}_3\text{CH}_2\text{OH}$ and CH_3OCH_3 . Both are $\text{C}_2\text{H}_6\text{O}$ but contain decidedly different functionalities.

The second major category of isomers and the focus of this chapter are stereoisomers. Being isomers they too have the same number and kind of building block atoms, but, unlike constitutional isomers, they have identical topologies. Stereoisomers, in turn, are divided into two groups: enantiomers and diastereomers. Enantiomers are isomers that are not superimposable on their mirror images. And, by definition, diastereomers are all other stereoisomers that are not enantiomers.

Complicating matters somewhat are conformational isomers and atropisomers. Conformational isomers, or conformers, are different forms of the same molecule arising from rotations around single bonds. While this seems conceptually simple and innocent, we point out that many conformational isomers are in fact nonsuperimposable mirror images of one another, and consequently, constitute enantiomeric pairs. An example of this is butane, existing in three distinct conformations: *anti*, with the CCCC dihedral angle 180 degrees, and *gauche* + and *gauche* - having dihedrals of +60 degrees and -60 degrees respectively. The g^+ and g^- conformers are enantiomers.

In spite of this, butane does not display optical activity because the two forms interconvert rapidly leading to a racemic mixture at room temperature. If one could lower the temperature to prevent this interconversion one could isolate the individual enantiomers but this is difficult to do in practice because excessively low temperatures are required. Alternatively, raising the barrier to rotation around the single bond would accomplish the same goal and in nature this happens. Enantiomers arising from hindered rotations around single bonds are called atropisomers. In Figure 1 we provide an example of a molecule displaying this type of hindered rotation [14]. This figure also illustrates that the enantiomeric forms can be separated by chiral chromatography at low temperatures but that at elevated temperatures the molecule racemizes.

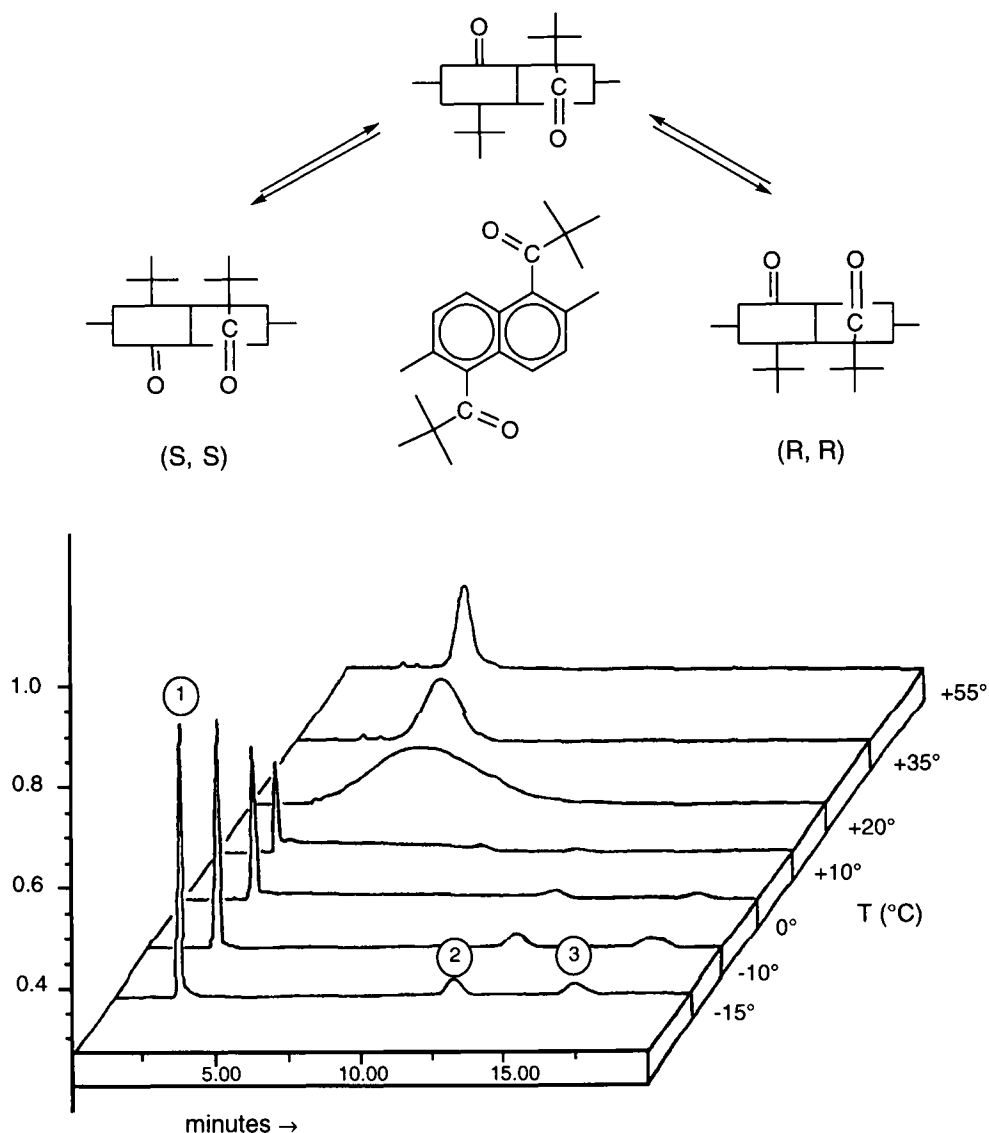


Figure 1. Dynamic HPLC illustrating the separation of atropisomeric naphthyl ketones on a chiral chromatographic column. The *meso* conformer (peak 1) and the conformational isomers (peaks 2 and 3) coalesce into a single peak at higher temperatures. Reprinted with permission from ref. 14.

It is imperative to recall that enantiomers, albeit different molecules with different IUPAC names, behave identically in a symmetric, nonchiral environment, but behave differently in an asymmetric, chiral environment. This nonequivalence is the key ingredient that makes chiral chromatography work and, as we shall see below, the

enantiomers flowing through a chromatographic column designed for chiral separations are experiencing an asymmetric environment of one type or another.

We now address the question: Why separate enantiomers? What is the compelling reason for resolving racemic mixtures anyway? For many years there was little impetus to separate enantiomers other than for the sake of knowledge; scientists were simply interested in studying the behavior of these stereoisomers in different environments. In the industrial sector there was no reason to separate enantiomers. Even in the pharmaceutical business where it was recognized that the enantiomeric forms of chiral drugs can, and usually do, behave differently there was little focus on chiral synthesis and even less on chiral separations until recently.

While there is no single episode that can be described as being responsible for the "chiral revolution" now taking place in industry, the thalidomide story underscores our need for optically pure drugs [15]. Thalidomide was a drug administered in several countries in the 1960's as a sedative to reduce morning sickness in pregnant women. The drug, like most, was given as a racemic mixture and it worked very well. Unfortunately it has a serious side effect. While one enantiomer is in fact a sedating agent, the other enantiomer causes fetal mutations and a large number of children were born with malformed or missing arms and legs (in addition to other life-threatening abnormalities) as a consequence of the "wrong" drug. Most pharmaceutical companies now consider each enantiomer as a different drug and they are poignantly aware of the consequence of having even the smallest amount of the wrong enantiomer in their product. Parenthetically, some enantiomers can act synergistically to enhance drug efficacy while others can serve as agonists or antagonists and yet others may be inert or even have a completely different biological effect. To test each enantiomer as a separate drug candidate one must have the pure enantiomers to work with. Because so few synthetic methodologies have high enough specificity to generate optically pure material, bench chemists are forced to separate a mixture of enantiomers and nowadays they rely on chiral chromatography for that purpose.

2. CHROMATOGRAPHY

As analyte molecules traverse a chromatographic column they are subjected to a variety of complex forces arising from subtle nonlinear pressure drops, a multitude of hydrodynamic effects giving rise to nonlamellar flows and so on. More important is the fact that as molecules move along the column they are being partitioned between an immobile stationary phase and a mobile carrier phase that may be a gas, a liquid or even a supercritical fluid. The type of mobile phase defines the kind of chromatography, e.g., gas chromatography, liquid chromatography or supercritical fluid chromatography. In each, the choice of stationary phases is limitless and the role of the separation scientist is to select a suitable asymmetric, chiral environment so that the two enantiomers have different retention times on the column.

In this regard one could simply use a chiral carrier fluid but this is rarely implemented because of their high costs. A less expensive approach is to add a chiral solvating agent to the nonchiral liquid making it, in effect, a chiral mobile phase. In this chiral mobile phase, then, the enantiomers interact differently with the additive, forming diastereomeric complexes that, by definition, always have different physical properties and behave like normal diastereomers that can be separated on nonchiral stationary phases like silica gel. An example of this would be to add a chiral macrocyclic host molecule such as cyclodextrin (a cyclic oligomer of glucose) that

selectively binds one of the two enantiomers as a guest-host complex. The more tightly bound enantiomer is less likely to adsorb to the silica gel and is carried more quickly through the column than is the less tightly bound enantiomer. The use of cyclodextrins as both a chiral mobile phase additive and as a covalently linked stationary phase will be described later in the chapter.

An alternative approach is to use a stationary phase that is itself chiral. While the concept of using a chiral stationary phase (CSP) for chiral separations has been known for many years, it is only since the development of high resolution chromatographic machinery that chiral resolutions have become possible on a routine basis. There are now over 200 commercially available chiral stationary phases and thousands of others are also in use for a variety of separations. We categorize these stationary phases below. But first, it is important to understand the relationship between analyte retention on such columns and chemical thermodynamics.

The distribution (partition) coefficient of the two enantiomers between the mobile phase and the chiral stationary phase, $K_{x,sm}$, is defined as:

$$K_{x,sm} = C_{x,s} / C_{x,m} \quad (1)$$

where C_x is the molar concentration of enantiomer x and s and m refer to the stationary and mobile phases, respectively. The free energy change associated with transferral from the mobile phase to the stationary phase is:

$$\Delta G^\circ_{x,sm} = -RT \cdot \ln K_{x,sm} \quad (2)$$

where R is the gas constant and T is the temperature in Kelvins.

The capacity factor, k'_x , is similar to the distribution coefficient but rather than relating concentrations, it relates moles of enantiomer in each phase, or retention time:

$$k'_x = n_{x,s} / n_{x,m} = (t_r - t_m) / t_m \quad (3)$$

Here t_r is the total retention time of an enantiomer from the moment of injection to moment of detection and t_m is the time for some unretained solute to reach the detector.

The relationship between K_x and k'_x is defined as:

$$K_{x,sm} = k'_x / \Phi_{sm} \quad (4)$$

where $\Phi_{sm} = V_m / V_s$ is the volume phase ratio between the stationary and mobile phases. Substitution of eq. 4 into the free energy equation (2) provides an expression for the free energy in terms of k'_x and Φ_{sm} :

$$\Delta G^\circ_{x,sm} = -RT \cdot \ln k'_x + RT \cdot \ln \Phi_{sm} \quad (5)$$

Because the Gibbs-Helmholtz equation, $G = H - TS$ relates free energy, G , to enthalpy, H , and entropy, S , equation (5) can be solved for $\ln k'_x$ as a function of T in terms of enthalpy and entropy to yield:

$$\ln [k'_x(T)] = (-\Delta H_{x,sm}/R) \cdot (1/T) + [(\Delta S_{x,sm}/R) + \ln \Phi_{x,sm}] \quad (6)$$

This equation serves as the mathematical basis for van't Hoff plots (plot of the natural logarithm of the capacity factor versus the reciprocal of the column temperature) and is commonly used for deriving enthalpies and entropies of analyte binding to a CSP. A linear van't Hoff plot provides $\Delta H_{x,sm}$ from the slope of the line, but, without an independent determination of Φ_{sm} , the intercept of the line will not provide $\Delta S_{x,sm}$. The difficulty in determining Φ_{sm} is assessing how much mobile phase exists between the injector and the detector. This is referred to as the dead or void volume. Because there is ambiguity in measuring this dead volume, chromatographers sidestep the problem by measuring the elution volume of a "non-retained" solute. However, such ideal compounds do not actually exist thereby creating another set of problems.

While such problems are not insurmountable, not everyone can nor does determine the phase ratios. An alternative approach involves determining the separation factor, α , where $\alpha = k'_R/k'_S$. Here the subscripted R and S refer to the Kahn-Prelog-Ingold stereochemical descriptors for enantiomers. Equation (6) can now be converted to a linear equation in terms of α as a function of T and differential enthalpies and entropies of binding, $\Delta\Delta H$ and $\Delta\Delta S$, as:

$$\ln [\alpha(T)] = [-\Delta\Delta H/R] \cdot (1/T) + [(\Delta\Delta S/R) + \ln (\Phi_{2,sm}/\Phi_{1,sm})] \quad (7)$$

In equation (7) the two phase ratio values are equal, simplifying the equation to:

$$\ln [\alpha(T)] = [-\Delta\Delta H/R] \cdot (1/T) + \Delta\Delta S/R \quad (8)$$

A van't Hoff plot of equation (8) would yield values of $\Delta\Delta H$ and $\Delta\Delta S$ and the free energy difference is simply derived as:

$$\Delta\Delta G = \Delta\Delta H - T \cdot \Delta\Delta S \quad (9)$$

Or, by multiplying equation (8) by RT, $\Delta\Delta G$ can be determined by the relation:

$$\Delta\Delta G = -RT \cdot \ln \alpha \quad (10)$$

Equation (10) is the expression most commonly used by separation scientists to derive differential free energies of binding for the enantiomeric pair. In most publications chromatographers provide retention orders, i.e., R elutes before S or *vice versa*, capacity factors, k'_R and k'_S , and, separation factors, α . It is rare to find published temperature studies where differential enthalpies, $\Delta\Delta H$, and differential entropies, $\Delta\Delta S$, are presented. Hence, for the computational chemist, the experimental data one typically has to work with are retention orders and differential free energies. Thus a computational methodology giving free energy differences rather than just enthalpy differences is required. Unfortunately most studies to date have been enthalpic type calculations but several groups doing most of the modeling in this field have computed the differential free energies for direct comparison with chromatographic separation factors. Accordingly, much of this review will focus on their work.

3. MOLECULAR MODELING

A model is a likeness, a semblance or a representation of reality. Interestingly, molecular modeling means different things to different people. On the one hand are scientists who have no need for detailed atomic information. They are interested in coarse grained features of the system as found in, say, kinetic models where diffusion, transport, reactant depletion and product formation are of concern. This type of modeling involves *macroscopic* features of the system under study. In contrast is the fine grained modeling represented by most chapters in this book series. Here we look, in painstaking detail, at precisely how each atom interacts with all the other atoms in a molecule. This view is a *microscopic* one.

Because we wish to know how each individual atom or group of atoms contribute to an observed phenomenon this kind of modeling is referred to as "atomistic modeling" or, more commonly, "molecular modeling." Atomistic modeling may be divided into "fitting" methods as in, say, QSAR/QSPR or other regression models and "applied theory" as represented by molecular mechanics. Both approaches make use of atomic level detail, however. The regression models require fitting to a dataset of known information and their use is therefore restricted only to that class of compounds to which they were fit. In contrast, "applied theory" models like quantum mechanics should be applicable to all systems. Both atomistic modeling approaches have their advantages and disadvantages and each should be used judiciously. Both of these atomistic modeling approaches have been used in chiral chromatography and will be described below.

4. CHIRAL STATIONARY PHASE SYSTEMS

The large number of CSP's developed, tested and marketed present somewhat of a problem for how best to categorize them. Wainer has suggested a classification scheme for HPLC CSP's based on the mode of formation of the solute-CSP complex [16]. There are five categories, labeled Type I-V, and molecular modeling has been done on most of these. The categories and modes of association are:

- Type I.** Where solute-CSP complexes are formed by attractive interactions like hydrogen bonding, pi-pi interactions and dipole stacking as represented by Pirkle-like CSPs.
- Type II.** Where the solute-CSP complexes are formed by attractive interactions and through the inclusion into a chiral cavity or ravine as represented by some cellulose based CSPs.
- Type III.** Where the primary mechanism involves the formation of inclusion complexes as represented by cyclodextrins.

The major difference between these three categories, irrespective of the type of intermolecular attractions, is the *extent of inclusion*. Type I has no inclusion complexation, Type II has partial inclusion and Type III uses inclusion complexation as the "primary mechanism." For this review I shall create a greater line of demarcation than Wainer between Type II and Type III phases. Here Type III shall be considered to be exclusively guest-host complexes as found in crown ethers, cyclodextrins and related systems, whereas Type II uses only partial guest-host complexation.

Type IV. Where the solute is part of a diastereomeric metal complex as in chiral ligand exchange chromatography.

Type V. Where the CSP is a protein and the analyte-CSP complexes are based on combinations of hydrophobic and polar interactions.

5. MODELING ENANTIOSELECTIVE BINDING

Modeling the enantioselective binding of analytes to chiral stationary phases or to chiral mobile phase additives is perceived by many as an easy task. Unfortunately nonspecialists who have access to modern molecular modeling software make many mistakes and the results from such studies are suspect. Most of the papers published in this field are derived from experimentalists who are not well versed in computational chemistry. These chemists rightfully view computational chemistry as an aid for interpreting their observations but they are unaware of the many pitfalls to avoid when doing such calculations [17]. For example, the majority of papers report interaction energies, typically from molecular mechanics, obtained by docking one molecule in only its most stable conformation with another in its most stable conformation and in some arbitrary orientation and position. Moreover it is assumed that the molecular mechanics energy difference between the two diastereomeric complexes has some relevance to the observed free energy differences. Fortunately there are a significant number of papers where more elaborate approaches have been taken and the work from these papers will be highlighted here.

6. TYPE I CSPS

As analytes migrate through the column they encounter solvated CSP, displace solvent and form the corresponding solvated diastereomeric complexes. There are two competing equilibria to consider. In equation 11 the CSP has the R configuration and the analyte, designated by A, also has the R configuration while in equilibrium 12 the analyte is of the S configuration. The bound states are weakly bound diastereomeric complexes that are transient in nature.



Rather than compute ΔG for eqn. 11 and ΔG for eqn. 12 to obtain $\Delta\Delta G$, recognize that the left hand sides of both equilibria are identical. This arises from an enantiomeric relationship where $\text{A}^{\text{R}} = \text{A}^{\text{S}}$ in an unbound state (recall from above that enantiomers have identical properties in an achiral environment which, in this case, is the unbound state). Consequently one need only compute the energies of the two diastereomeric complexes to determine which analyte is more tightly bound to the CSP and, accordingly, has the longer retention time on the column.

There are many assumptions typically made in these calculations. These include: assuming the rate of complex formation is the same for R as for S analyte and that only the relative stabilities of the complexes are important; complete neglect of mobile phase additives, ions or solvent, although we know that diastereomers have differential solvation free energies and experimentally we can sometimes find reversal in retention orders depending on solvent; elimination or truncation of the

spacer chain connecting the CSP to the silica surface even though it is known that the length and type of attachment to the packing material is important; neglecting the packing altogether (usually silica gel). Hence, most modeling done to date (published that is) is in the "gas phase" and often using CSP *analogs* rather than the actual CSP itself.

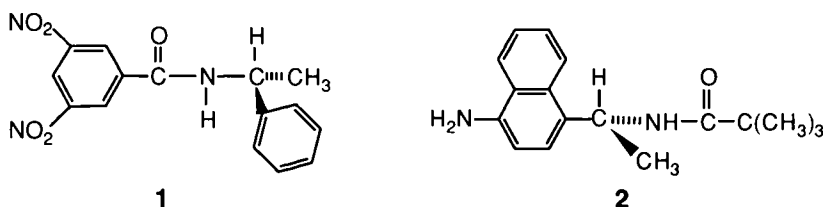
To model the interaction between CSP and analyte one must account for 1) the shapes of the two molecules in the binary complex, 2) the relative position of the two molecules, i.e. the analyte should be at its proper binding site on or around the CSP, and 3) the orientation of the two molecules with respect to each other. This is just a simple way of saying that some sort of ensemble average is needed wherein a molecular dynamics protocol must ensure adequate sampling of phase space, or, if using a Monte Carlo strategy, a sufficient number of important configurations must be sampled.

With regard to the first point above it should be noted that these CSP molecules are not rigid, lattice-like molecules, but rather are flexible organic compounds tethered to silica gel by aliphatic spacer chains (*vide infra*). Hence using only the most stable conformation of the selector and of the selectand is not adequate. This can be understood by considering an imaginary CSP whose lowest energy conformation contains an intramolecular hydrogen bond. This particular conformer may have little influence on the binding process between CSP and analyte even though it is the most heavily populated isomer. As analyte approaches the CSP, both CSP and analyte molecules may need to adopt higher energy conformations suitable for binding. Interestingly, then, the most stable conformer of a CSP may not be responsible for either binding or for chiral recognition.

With regard to the second point above, one also needs a suitable sampling strategy that accounts for the position and orientation of both molecules in the complex. Two general approaches exist. The first is to embed some *a priori* knowledge about the binary complex into the search strategy. For example, if the analyte is cationic one would search around an anionic receptor site. This type of search strategy is called "motif-based" searching because there are well defined intermolecular binding motifs existing in nature that one can take advantage of. Below we provide examples of such search strategies where the *a priori* knowledge comes from solution phase NMR or IR spectral investigations of soluble CSP-analyte analogs while other motifs come from "chemical intuition." The second type of search strategy is to let the computer do the search fully, without any preconceived notions and without any operator intervention.

6.1. Motif Based Searches

An excellent example of motif based search strategies is found in the work of Däppen, Karfunkel and Leusen [18]. These scientists were interested in understanding how chiral separations take place and then to use that knowledge to design enhanced stationary phases. The authors first determined experimentally that the R enantiomer of analyte¹ is bound more tightly to 2, a chiral stationary phase selector that is tethered to silica by the amino group.



Because these authors had determined experimental $\Delta\Delta H$ binding energies they focused their efforts on computing enthalpies rather than free energies. The following steps were used to do this.

1. Carry out a complete conformational analysis of compounds 1 and 2 using molecular mechanics and semiempirical quantum mechanics.
2. Use these conformers to construct the initial binary complexes of the two diastereomeric complexes.
3. Use an *a priori* classification of the possible interactions between parts of the two molecules to reduce the complexity of the problem (there are 6 trans/rotational degrees of freedom for rigid body dockings and more when torsions are considered for flexible docking). Every low energy conformer of 1 is then combined with every low energy conformer of 2, in an orientation such that one or more of the binding motifs is realized.

In their problem three basic binding motifs were considered relevant: hydrogen bonds between selector and selectand, dipole stacking of the two amide moieties, and pi-pi stacking (charge transfer complexation) between the dinitro pi-acid portion of 1 with the pi-basic aminonaphthyl ring in 2. The motif code they used has the general form **C_nA_{sm}T_j** where :

C_n represents the CSP in the n^{th} minimum energy conformation.

A_{sm} designates the absolute configuration (R or S) of the analyte's m^{th} energy minimum.

T_j indicates the motif interaction type defined by the authors as:

T1: A hydrogen bond from the amide NH of CSP to the CO of analyte

T2: A hydrogen bond from analyte NH to CSP CO

T3: Dipole stacking

etc.

Hence C2AR1T3 means the initially formed complex has the CSP in its second most stable conformation, the R isomer of analyte in its most stable conformation is bound, and the molecules are oriented using a dipole stacking motif. In their study only three stable conformations of CSP and two stable conformers of analyte were located, giving rise to a relatively small number of initial binary complexes to consider. But, considerably more complex chromatographic systems can be envisaged, making the problem far more difficult. Their nomenclature system is a convenient way to keep track of which molecules, conformations and orientations are being used to create the initial complexes in motif-based search strategies.

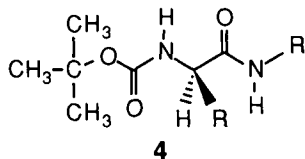
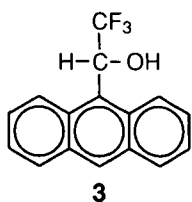
From this, a small number of possible structures were generated. Unreasonable structures were removed by visual inspection using molecular graphics. The authors

decided to use only 49 such structures because they intended to energy minimize each complex with AMPAC (49 for the R enantiomer and 49 for the S enantiomer!) and somehow needed to reduce the number of complexes to be considered.

Eventually their Boltzmann weighted average for the RR complex was found to be -7.35 kcal/mol at 300K while that of the corresponding RS complex was -5.35 kcal/mol. The 2 kcal/mol energy difference corresponds well with the observed value of 1.22 kcal/mol. Based on this success the authors examined various structures to rationalize why one diastereomer is more stable than the other as well as to gain insights about what is needed to improve the stereoselectivity of binding. The authors then delineate an extensive design protocol for the creation of improved CSPs. The crux of this paper, and something noted by the authors as being extraordinary, is the small number of initial orientations needed for geometry optimization. They conclude that the binding motif approach is suitable for routine investigations when computational costs must be taken into account. However, this is the only example from that research group and using a motif based search strategy may not be reliable all the time; more testing is warranted.

These authors also considered a grid search strategy (*vide infra*). In the grid search method they used the same conformations as above but then let the computer rather than the chemist generate a large number of possible starting orientations. They minimized a subset of these with an empirical force field to create a collection of 30 low energy structures that were subjected to full geometry optimization using both quantum and molecular mechanics. The result of their systematic grid search was that the lowest energy RR complex (experimentally RR is more stable than RS) that was located by their manual, motif-based docking strategy could not be found by the grid search. It should be pointed out that in their grid search they used only the lowest energy conformation of 2 and only the lowest energy conformation of 1 and, apparently, their initial strategy for filtering structures to create their subset for full geometry optimization rejected those structures that could have led to the low energy motif-based structure found above. These authors thus abandoned further grid searching attempts that would have used other conformations of 1 and 2. Later we illustrate that fully implemented grid searches, albeit time consuming, provide an excellent way of computing differential free energies.

Another group concerned with the question of how best to sample configurations for statistical averaging is from Rogers' laboratory. Rogers' group had synthesized chiral stationary phases by bonding *tert*-butyloxy carbonyl (BOC) derivatives of amino acids to a butyl spacer on silica and then examined their ability to discriminate between R and S 2,2,2-trifluoro-1-(9-anthryl)ethanol, 3. The modeling involved CSP analog 4 where R = CH₃ (alanine), R = isopropyl (valine), and R' = different length n-alkyl chains.



Using the MM2 force field Still and Rogers [19] assessed the distribution of conformers for the analyte and the CSP analogs. Four docking strategies were

employed in this study but only the most stable structures from their conformational analysis were used except in the case of BOC-D-valine-N'-n-propylamide where two conformers of similar energy were used. The docking strategies were based on previously determined NMR chemical shifts, and so, this too is an example of motif-based docking.

The first strategy involved maximizing the interactions between the carbonyl oxygen of the BOC group and the hydroxyl hydrogen of the analyte in addition to maximizing the interaction of the protected amine's hydrogen with the analyte's anthryl ring. Different orientations did not have large interaction energies for either assumed points of interaction so three other motif-based docking maneuvers were explored until low energy structures were found.

For analyte binding to the BOC-D-alanine-N'-n-propylamide CSP the S enantiomer was computed by the MM2 force field to be favored by 0.05 kcal/mol. This does not agree with experimental retention orders but is consistent with the small energy difference observed experimentally ($\alpha = 1.02$). When analyte binds to the BOC-D-valine-N'-n-propylamide CSP the R enantiomer is computed to be favored by 0.18-0.52 kcal/mol, depending on which conformer of CSP was used in the docking, and this does agree with experiment. The authors concluded that the valine phase would be more effective than the alanine phase and that the R analyte would be eluted later than S on the valine phase. Both predictions agree with experiment. Notice here that very small energy differences are being computed. The question to ask is: can one legitimately use a method like molecular mechanics to compute such small differences, especially in light of the fact that the mean errors in computed heats of formation of well parameterized force fields are at best around 0.5 kcal/mol? The answer is yes, and an explanation about why this is so is given later in the chapter.

Still and Rogers then began assessing the origins of enantioselectivity. They first examined energy differences between the molecular mechanics' component energies contributing to the total energy (e.g. E_S , E_B , E_{NB} etc.). They also examined the energy of interaction between parts of the analyte with parts of the CSP. The largest difference, and thus the most discriminating fragment, is the anthryl ring and not the oxygen of the analyte even though this atom is contributing heavily to the formation of the complex. A similar treatment allowed them to determine the most discriminating parts of the valine CSP. Again, there were found some atoms and groups of atoms that help stabilize the complex but which do not discriminate. This is a wonderful application of molecular modeling; Still and Rogers are extracting important information that is difficult or impossible to derive otherwise.

In spite of this the authors found some inconsistencies with other computational and experimental results. For example, calculation of 4 (with $R' = \text{Me}$ and Et) differed significantly with those values when a propyl spacer was used. If anything, a trend in this homologous series is expected. It appears that the authors' results are a consequence of insufficient sampling and/or not using Boltzmann averaged interaction energies (they used only the global minima of their binary complexes to determine what is essentially an enthalpy difference rather than a free energy difference). Eventually these authors abandoned the motif based search strategy for a fully automated search described below.

6.2. Automated Search Strategies

Rather than to bias the results using preconceived notions about binding, Lipkowitz and Darden decided to fully automate the search [20]. Their approach was to evaluate all conformations of CSP and of analyte, and then combine all conformations of CSP with all conformations of analyte as in the work of Däppen *et al* [18]. Hence if there exist M conformational states of CSP and N conformational states of analyte there will be $M \times N$ binary complexes to consider for each diastereomeric complex. Lipkowitz and Darden decided to treat the individual components as rigid bodies and sample all Euler angles of the two molecules with respect to one another using a grid search. The position of analyte with respect to CSP is represented in a spherical coordinate system (r, Θ, ϕ) . The authors select an origin and three orthogonal axes on the CSP. They then select an origin and a set of axes on the analyte in a way that allows them to systematically sample all orientations of the analyte with respect to the CSP. The distance, r , between origins, and the latitude, Θ , and the longitude, ϕ , between origins, precisely defines where the analyte is with respect to the CSP (Figure 2). It was expected, and found by the authors, that the results are independent of where the origins are located; the origins could be at the centers of mass of each molecule or, for convenience, at the stereogenic centers as the authors use.

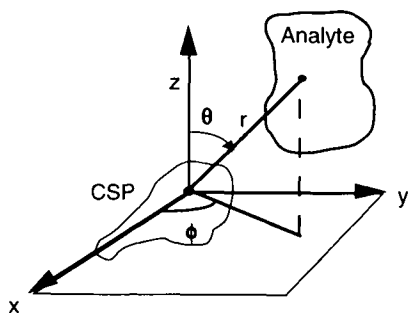


Figure 2. Position of analyte with respect to chiral stationary phase is given in spherical coordinates, (r, Θ, ϕ) . r is the distance between arbitrarily selected origins and Θ and ϕ describe the latitude and longitude of analyte with respect to CSP. The relative orientations of the two molecules at each latitude and longitude are defined by their Euler angles.

Lipkowitz and Darden then consider an imaginary rod emanating from the origin of the CSP at a fixed latitude Θ and longitude ϕ . They allow the analyte to slide down the rod until the van der Waals surfaces of both molecules in the binary complex just touch (actually slightly interpenetrate). They then compute an intermolecular energy, reorient the analyte, slide it back down the rod, and recompute the energy at that Θ and ϕ . After all orientations of the analyte have been sampled at that Θ and ϕ , they move the rod to a new latitude and longitude and repeat the aforementioned procedure. Eventually all values of Θ and ϕ are sampled. In essence the authors are rolling the analyte molecule (as a rigid body probe) over the van der Waals surface of the CSP, sampling configurations for statistical averaging while simultaneously looking for the lowest energy binding region as well as the most stable orientation of the two molecules with respect to each other.

In that work the question about how best to sample the Euler angles was considered. Two options exist: stochastic and deterministic sampling methodologies. Using a Metropolis Monte Carlo or a "smart" Monte Carlo method with some type of importance sampling algorithm might suffice. Indeed, the authors' code allows one to carry out MC searching but that approach was never implemented. The reason for this is one can never be certain that all important microstates have been sampled for

the statistical averaging (see below). This is a critical issue because the experimental free energy differences are so small; typically less than kT !! The concern is that MC searches may sample one region more heavily in one of the diastereomeric complexes than in the other, leading to abhorrent results derived from computational artifacts. Instead, the authors adopted a "brute force," grid search methodology. For each of the $M \times N$ rigid body binary complexes, the same sampling motion used for the RR diastereomer is used for the RS diastereomer, but they are just mirror images rolling motions of one another. This is extremely CPU intensive but it does ensure all minimum energy structures are accounted for.

How, then, did the authors ensure this? First they selected a coarse grained search and determined the number of minima on each complex's intermolecular potential energy surface. They then repeated this rolling motion using finer and finer grids until the number of minima found remained constant. A satisfactory trade-off between computer time and grid coarseness was such that for each $M \times N$ complex, approximately 155,000 orientations equally spaced around the CSP are computed and stored for processing. So, for CSPs with three conformations and analytes with four conformations, this method requires at least $3 \times 4 \times 155,000 = 1.4 \times 10^6$ configurations for the R analyte and an equivalent number for the S analyte binding to the CSP. Certainly as the number of conformations grows this approach becomes unwieldy and a stochastic approach becomes warranted.

Rather than compare only the lowest energy structures of the competing diastereomeric complexes, Lipkowitz computes a statistically averaged interaction energy, \bar{E} , as in eqn. 13.

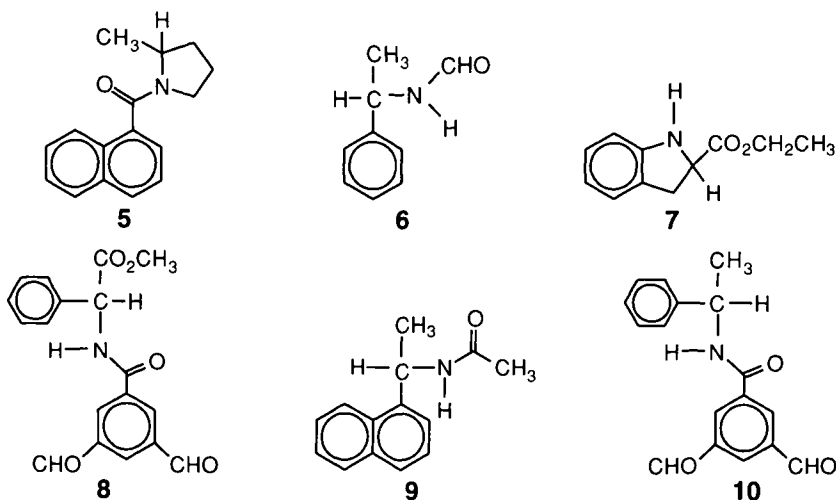
$$\bar{E} = \sum_{h=1}^1 \sum_{i=1}^m \left(\frac{e^{-E_{CSP,h}/kT}}{\sum_{h'=1}^1 e^{-E_{CSP,h'}/kT}} \right) \left(\frac{e^{-E_{A,i}/kT}}{\sum_{i'=1}^m e^{-E_{A,i'}/kT}} \right) \sum_{j=1}^n g_{hij} \left(\frac{e^{-\epsilon_{hij}/kT}}{\sum_{j'=1}^n e^{-\epsilon_{hij'}/kT}} \right) \quad (13)$$

The terms within the parentheses are simply probabilities. The first term is the probability of finding the CSP in a given conformational state, the second term is the probability that the analyte is in a particular conformation and the last term is the probability that the two molecules are positioned and oriented in a particular way with respect to each other. Note that because the authors locate all the minima on the complex's intermolecular potential energy surfaces they can derive the entropy of the system as well. Therefore \bar{E} is actually a good representation of the macroscopic free energy of interaction, which in this case corresponds to a Gibbs free energy.

Because of the large number of configurations sampled the authors restricted themselves to empirical force fields for computing the intermolecular energies. This is not only faster than a quantum-based approach but it is better because empirical force fields reproduce more accurately the dispersion forces between molecules than do quantum methods at a Hartree-Fock level of theory. Because empirical force fields are not particularly accurate or precise, the authors decided to treat the R and S analytes in an identical manner as explained above. This way, if the force field underestimates, say, hydrogen bonding and overestimates electrostatics, the errors should be nearly the same for both R and S analytes. This cancellation of errors

should result in small but meaningful energy differences between diastereomeric complexes.

Using this deterministic approach, Lipkowitz' group tested their software and sampling protocol on a broad set of analytes, including 5-10, binding to various CSP analogs [21,22]. Their reason for selecting molecules like 5-10 was that these compounds had been resolved experimentally and retention orders together with separation factors were documented. Parenthetically, most chromatographers do not divulge the absolute stereochemistry of the molecules eluting from the column (R or S), but rather, give only the sign of the optical rotation (+ or -), and accordingly, comparison between theoretical and experimental results is not usually possible. The second reason for selecting such analytes is that they represent a wide range of functional groups. Lipkowitz and Darden wanted to ensure that their protocol was robust and could handle a diverse class of analytes as well as CSPs.

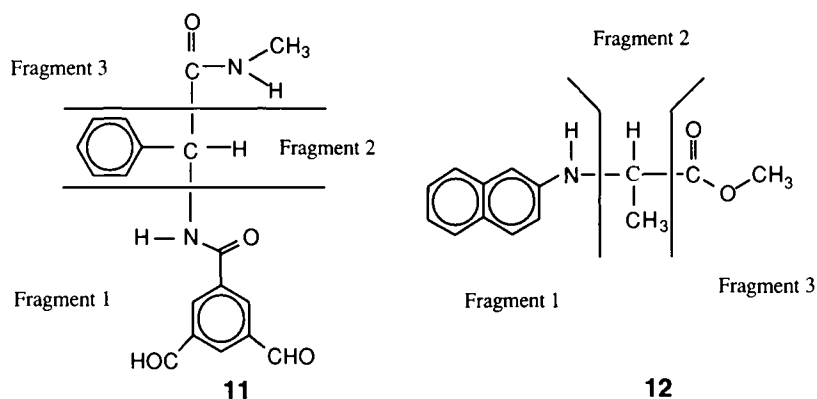


Because they were computing the differential free energies of binding, these authors were also able to compute the corresponding separation factors, α . Good agreement between theory and experiment was found. Hence, for Type I stationary phases these authors always predict the correct retention orders and, with some degree of precision, they can determine α . Having demonstrated the modeling protocol to be reliable they began extracting information from the simulations not amenable to experimentation.

First those authors considered the binding site on the CSP. This is simply where the analytes spend most of their time around the CSP. The main question was: do both analytes bind to the same place on the CSP or to different places? *A priori* there is no way of knowing this, but by examining the intermolecular potential energy surfaces they were able to conclude, for the analytes studied, that the binding sites are the same for both enantiomers indicating that it is not *where* the analyte binds that is important but rather *how* it binds that is important.

Next they considered the stereodifferentiation process itself. An energy partitioning scheme was developed allowing them to divide the total binding enthalpy into molecular fragments constituting the CSP and/or the analyte. The idea to examine intermolecular energies attributable to parts of one molecule with another came from

the many elegant DNA-drug intercalation studies by Kollman who divided the binding energies into contributions from parts of the DNA and parts of the intercalated drug [23]. The only difference between Lipkowitz' and Kollman's approach is that the latter was used for a single configuration (the most stable intercalated structure) whereas the former is ensemble-averaged. A fragment may be any atom or collection of atoms. For convenience Lipkowitz divided CSP analogs 11 and 12 into three fragments each, as depicted below.

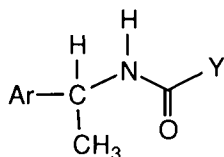
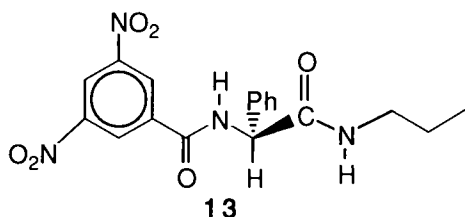


This division is subjective and completely arbitrary; different divisions would give different results. Nonetheless they found that for both CSPs, fragments 1 and 3 are primarily responsible for holding the complexes together. In other words, irrespective of the chirality of the probe molecule, fragments 1 and 3 are doing most of the work holding the complex together. More important, though, is the difference each fragment feels because this difference is an index of chiral discrimination. Certainly, if one finds an atom or a cluster of adjacent atoms experiencing a large difference in interaction energies between the two mirror image probe molecules, one would say that atom or group of atoms is discriminatory. But, if they feel little or no difference they are not enantiodiscriminating. Generally the fragments doing most of the work holding the complex together were also found to be most discriminatory. Bear in mind this need not be true for all selector-selectand pairs and one may find chiral discrimination arising from fragments not primarily responsible for binding. Interestingly, fragment 2 in both CSPs is usually least cognizant of differences between enantiomeric analytes. This finding is counterintuitive at first but consistent with the fact that the entire CSP is chiral, not just the stereogenic carbons of fragment 2. Again, molecular simulations have achieved their goal by uncovering something that would not have been observed otherwise.

Other groups have also pursued systematic searches. In the work by Däppen *et al.* described above [18], the authors carried out a grid search as follows. Two low energy conformers are aligned as rigid structures on a six-dimensional grid corresponding to the degrees of freedom of the system. The translational increments were 1 Å and the rotational increments were 24 degrees. The authors implemented CHEM-X to exclude unfavorable lattice regions (presumably those corresponding to van der Waals overlap) and generated hundreds of thousands of starting structures. The best 300 structures were selected and then minimized with the CHEM-X force

field. The 30 best structures were subjected to AMPAC minimization with all torsions and some bond angles allowed to relax. The authors did not find the same low energy diastereomeric complex derived from their motif based search and decided not to pursue this methodology further, citing the inordinate amount of CPU time required to do the search completely.

In contrast, Still and Rogers abandoned their motif-based search strategy and began developing improved grid searching methods [24]. The system Still and Rogers focused on was the R-phenylglycine phase, 13.



14 Ar = 1-naphthyl, Y = CH₃ α = 1.86

15 Ar = 1-naphthyl, Y = OCH₃ α = 1.52

16 Ar = phenyl, Y = CH₃ α = 1.15

Three aminoethanes, 14-16, whose retention orders and separation factors had been determined experimentally, were examined. In all cases the S enantiomer is retained longer on the R CSP. First a conformational analysis using the MM2 force field was carried out. Then, using the most stable structures, the CSP and the analyte were docked using an automated search algorithm. The chiral carbon on the CSP was selected as the origin. The analyte molecule was then set at a specific starting distance and orientation. The six degrees of freedom describing the position and relative orientations of the two molecules were randomly selected with the constraint that the molecules avoid interpenetrating each other. A range of positions and orientations was selected to ensure full sampling over most of the CSP. A simplex optimization

procedure was then used to minimize the nonbonded contact energies of the rigid bodies in the complex. A large number of starting orientations were obtained and a screening process was devised to remove equivalent or near-equivalent structures. Finally, full geometry optimizations were carried out on the docked starting structures that were within 4 kcal/mol of the lowest one.

Three variations of this minimization procedure were compared, the first two of which simply involved tightening the convergence criteria of the optimization. The last method involved minimizing the complex, using a simplex routine to dock those new structures, reminimizing the complex and cycling through this docking/minimizing procedure until the energy of the system can no longer be lowered. Once the docked pairs' energies are determined, a Boltzmann weighted average enthalpy and an entropy were calculated much like Lipkowitz *et al.* The advantage over Lipkowitz' method is that induced fit changes are allowed to occur in this procedure. Several computational methodologies were compared and contrasted in this paper and the third method not only predicted the correct elution order for each enantiomeric pair, but also gave reasonable values. Most interesting was the fact that this docking/minimization strategy used only the most stable conformers of CSP and analyte.

In an ensuing paper [25] they extended the computational study to consider how the dielectric of the medium affects the conformer populations, discussed modeling of

different size spacer linkages, and, provided far more detail of the structures of the docked species. The authors also demonstrated that relying only on the weighted average enthalpy terms did not always agree with those based on free energies nor with experimental data; as pointed out earlier by Lipkowitz, entropy must be considered.

Lipkowitz and his group also modeled this system [26]. The enantioselective binding of the same analyte to Rogers' BOC-D-Val chiral stationary phase was carried out using a grid searching strategy. It was correctly predicted that the enantiomer with the R configuration is longer retained but the separation factor, α , was slightly overestimated. The authors divulge that both enantiomers bind to the same general region around the CSP but that the intermolecular potential energy surfaces are much flatter than in other Type I systems. Also, the BOC group was determined not to be most responsible for chiral recognition as proposed by Rogers. Rather, the amide group on the spacer is most enantiodifferentiating. Finally, in an attempt to understand why the separations are insensitive to solvent polarity (the k' values decrease but α is invariant to polar modifiers), an analysis of diastereomer solvation was undertaken. Fully 3/4 of the BOC-D-Val CSP's surface was found to be hydrocarbon in spite of the CSP having two amides and an ester functionality. These polar functional groups are hidden under an umbrella of aliphatic hydrocarbon atoms preventing polar solvents from interacting with CSP.

Eventually the concern of neglecting solvent when modeling enantioselective binding precipitated a full study of the differential solvation energies of weakly bound, nonionic diastereomers as found in chiral chromatography [27]. The results of those simulations convey the following picture. As two molecules in their bundled-up, minimum-energy conformations encounter one another they tend to unravel somewhat to maximize the attractive dispersion forces with their partners. Unraveling from a low-energy conformation to a less stable form is offset by the gain in energy from complexation. For the weak complexes studied the more stable of the two diastereomers has both the CSP and the analyte most unraveled and furthest extended. This enhances substrate binding and it results in the more stable diastereomeric complex having the larger solvent-accessible surface areas. It is found that for weakly bound diastereomeric complexes, the differential solvation energies are within an order of magnitude of the differential free energies of analyte binding and that solvent conditions should play a major role in analyte resolution.

At this juncture it is worth emphasizing that the sampling protocol of Lipkowitz and Darden, albeit successful, uses rigid bodies for both selector and selectand. Hence while they do account for the different conformations of the two molecules making up the binary complex, they do not account for induced fit changes that take place upon binding as one would derive from, say, MD simulations. These authors actually addressed that issue; they energy minimized a large number of low energy configurations obtained from their rigid body search. As expected, the energy of each configuration dropped significantly but the computed retention orders and separation factors using those new energies did not change appreciably. Effectively, then, the same results were obtained allowing for induced fit changes as with the rigid body sampling.

The reason Lipkowitz and Darden were able to successfully use rigid bodies in their calculations is because the complexes studied are very weakly bound; they are held together by dispersion forces, charge transfer complexation, dipole stacking and limited hydrogen bonding. Indeed, these authors found that root mean squared (RMS) deviations between structures compared before minimization to those after full

relaxation were very small, for both selector and selectand. However, one would imagine that large structural changes could take place for molecules having multiple intermolecular hydrogen bonding, and certainly for tightly bound complexes with, say, a charged group on one molecule interacting with a charged group on the other molecule. In these cases large structural changes will be induced and rigid body docking strategies become inappropriate.

This was pointed out by Aerts who developed an alternative modeling method for prediction of enantioselectivity [28]. His approach for sampling configurations is to use high temperature molecular dynamics trajectories to generate a large number of conformations and orientations of the two molecules in the complex followed by energy minimization (sometimes called quenched dynamics). Using 1500K trajectories it is presumed that one can overcome torsional barriers so that new conformations can be generated while simultaneously creating random orientations of the two molecules with respect to one another. A hard wall constraint was imposed to prevent the two molecules from flying apart during this process. It is assumed that both molecules take all possible conformations this way and the results become independent of starting conformation, a big advantage in search strategies.

Sampling at regular intervals the author collects a set of configurations, typically around 5000, that are each energy minimized with an empirical force field. These structures are sorted by energy and redundant structures are deleted from the list (those structures whose energies are within 10^{-4} kcal/mol of one another and have similar dihedral angles and intermolecular atom-atom distances are deemed redundant). The resulting dataset is only accepted if the lowest energy configurations (those within 1.5 kcal/mol from the lowest) each occur at least 5 times in the dataset. Otherwise, additional datasets are generated until this criterion is met. The probability of finding a D-complex is:

$$p_D = \frac{\sum_{j=1}^{N_D} e^{-E_{D_j}/kT}}{\sum_{i=1}^N e^{-E_i/kT}} \quad (14)$$

where N_D is the number of D complexes. The enantioselectivity is given as the ratio of probabilities:

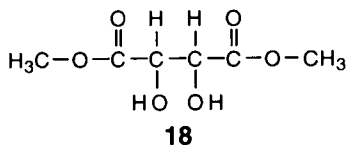
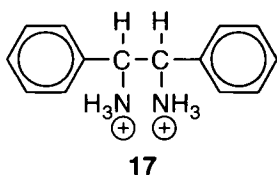
$$\begin{aligned} \alpha &= [D]_o / [L]_o = N_D / N_L = p_D / p_L \\ &= \frac{\sum_i e^{-E_{D_i}/kT}}{\sum_n e^{-E_{L_n}/kT}} \end{aligned} \quad (15)$$

For the summations the author used only those unique complexes within 2.0 kcal/mol of the lowest-energy complex. Setting the threshold to higher energy values did not much affect the results, as anticipated from such a Boltzmann weighting.

The chromatographic system studied by Aerts is exactly the same as that evaluated by Däppen, Karfunkel and Leusen [18]. Using 10,000 configurations, Aerts was able to correctly predict the retention order and, depending upon the method used to compute separation factors, obtain fair to excellent results. The conclusions derived from this study refute the hydrogen bonding scheme proposed by Däppen *et*

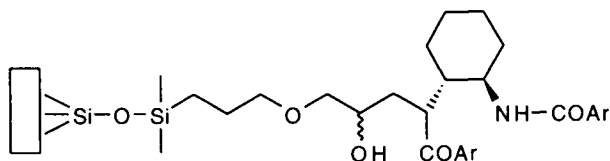
al. as being important and suggest rather that complexation arises from hydrophobic and dipole stacking forces.

The application of Aerts' strategy to bis-protonated 1,2-diphenyl-1,2-diaminoethane, **17**, binding to R,R-tartrate, **18**, illustrates the applicability of this strategy to tightly bound complexes.



The bis-ammonium ion has three conformations with relative energies of 0.0, 4.04 and 5.40 kcal/mol as determined by Discover. The lowest energy conformer has the $\text{NH}_3^+\text{-NH}_3^+$ aligned *trans* because of unfavorable electrostatic interactions as expected. However, upon complexation the *gauche* conformers form the more stable complexes with the tartrate. Another application of R,R-tartrates binding to protonated norephedrine is presented, highlighting the applicability of this search strategy to tightly bound complexes, but, since these examples are outside the scope of this chapter, they are not described further. Parenthetically, Aerts was not the first to use this kind of search strategy for chromatography. Topiol had earlier been studying the chiral recognition of methyl N-(2-naphthyl)alanate with N-(3,5-dinitrobenzoyl)leucine n-propylamide, a Type I CSP analog [29-31]. Sabio and Topiol used room temperature MD trajectories to collect configurations, from preselected positions and orientations obtained by considering the binding motifs available to that complex, for further energy minimizations with quantum and molecular mechanics [32]. Their work along with that of others has been reviewed by us earlier [33].

We now highlight the sampling strategies of Gasparrini and Misiti. These scientists have had an ongoing collaboration concerning the creation and use of Type I CSPs [34]. In particular they have developed a variety of phases based on chiral, *trans*-1,2-diaminocyclohexane (DACH), **19**.



- 19a** Ar = 3,5(NO₂)₂C₆H₃
19b Ar = 3,5(Cl)₂C₆H₃
19c Ar = C₆F₅

Extensive analyses of selector-selectand interactions based on NMR studies (of soluble analogs), X-ray crystallography and chromatographic resolutions were used to complement their molecular modeling studies. Gasparrini's approach is called the Global Molecular Interaction Evaluation (Glob-Moline) whose flow chart is outlined on the following page [35]. Like Lipkowitz and Rogers, Gasparrini considers all

conformations of both molecules and computes true free energies from this search strategy. It is to be noted that Gasparrini's simplex optimization of rigid guest with rigid host positions is similar to the work of Rogers and is comparable to Lipkowitz' method using a very fine grid. Eventually, though, the structures located by Gasparrini are fully geometry optimized accounting for induced fit changes of structure. This methodology provides meaningful results when compared to experiment and is currently the most robust method available for computing analyte interaction with Type I CSPs.

Finally, we consider the work of Bartle and his collaborators who proposed a method accounting for the matrix to which these brush-like stationary phases are attached [36]. As we pointed out earlier, all studies to date neglect the stationary phase matrix and treat the CSP as if it were freely available to the analyte from all directions. Bartle contends that regions of the CSP near the matrix to which the CSP is grafted are less accessible to the analyte than are other regions. These authors do not explicitly model the matrix, but rather include a penalty function for analyte approach.

A schematic of their sampling strategy is presented in Figure 3a. The analyte molecule is moved stepwise around the CSP in a grid-like fashion and this is done in an automated and systematic way. In the figure the dotted lines are the loci of points around the CSP. At each point several analyte orientations are considered. Those initial geometries of the binary complex are allowed to relax by minimizing the complex's energy with Hyperchem's MM+ force field (MM1 with additional force field parameters). While the authors indicate they are locating the nearest local minimum on the potential energy surface being explored, they do not say if all degrees of freedom are being relaxed or if their minimization involves a rigid analyte being translated and rotated over a rigid CSP. The authors carry out several thousand minimizations using this sampling strategy and then compute a Boltzmann weighted average. Comparison of the weighted averages allows for predicting elution orders and times.

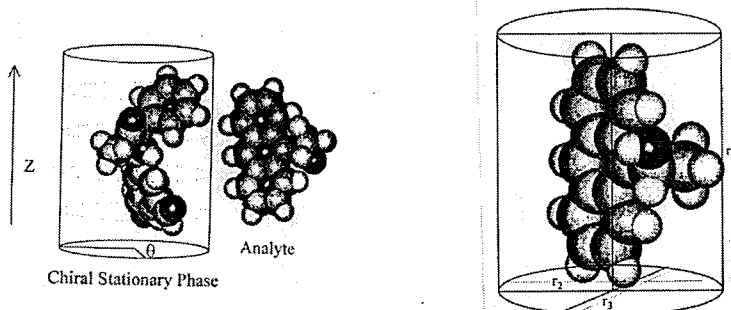
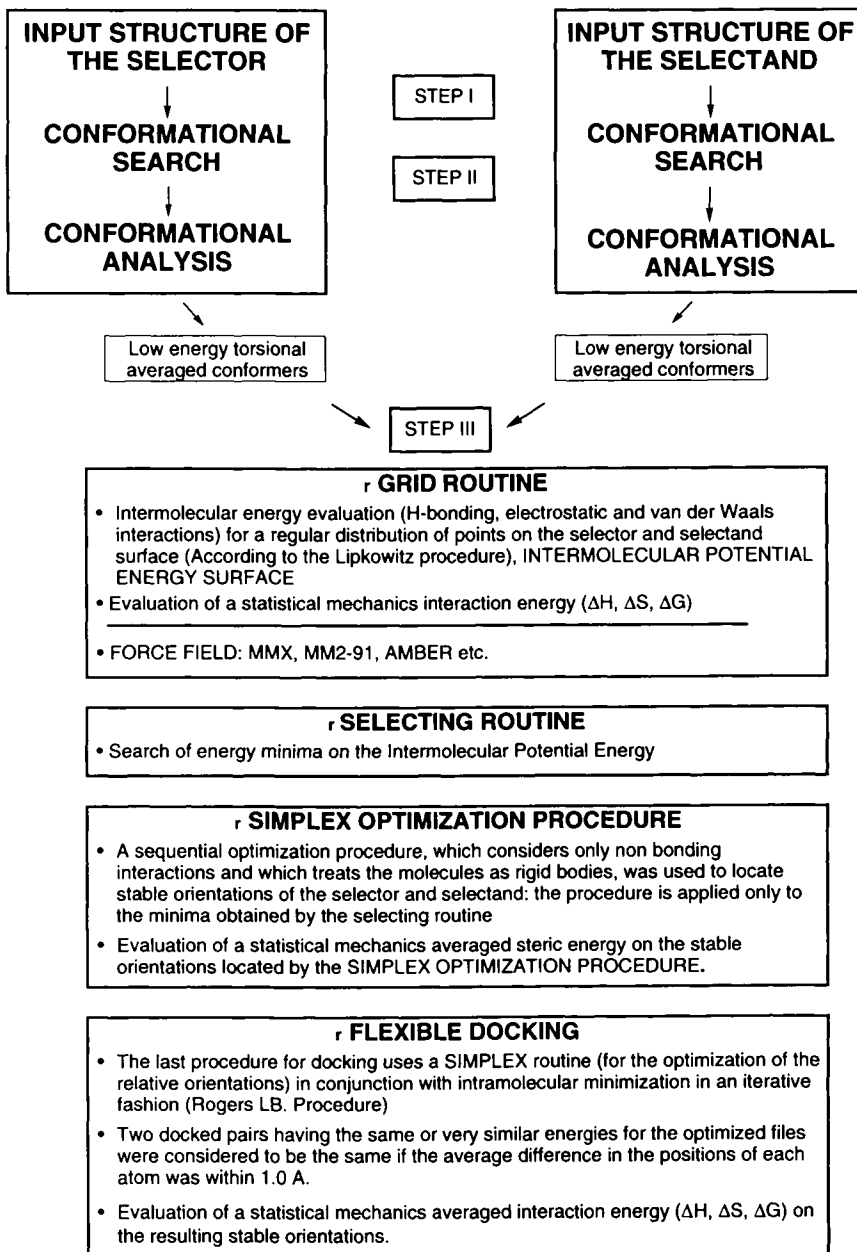


Figure 3. Left: Analyte is moved stepwise around the CSP with a grid that is determined by cylindrical coordinates. Right: Dimensions of an imaginary elliptical cylinder surrounding the analyte molecule. Figures provided by K.D. Bartle.

As an analyte molecule approaches the matrix the steric hindrance increases and the interaction energies should also increase. A function, $f(r)$, governs this steric hindrance and a penalty function is incorporated as in equation 16 where $p(E_i) = e^{-E_i/RT}$.

-AUTOMATIC DOCKING PROCESSING-

FLOW CHART FOR "GLOB-MOLINE"



$$\Delta\bar{G} = \frac{P(E'_i) \cdot e^{f(r)K}}{\sum_{i=1}^I E'_i} \quad (16)$$

The nomenclature adopted has E_i as the energy of interaction between the CSP and the analyte molecule, $p(E'_i)$ the probability of the i^{th} complex, E'_i the energy of the i^{th} complex, T and R are the temperature and gas constants, respectively, and I is the number of discrete samples taken. Note that the authors use the term $\Delta\bar{G}$ but it is not clear if entropy is explicitly included in this averaged energy. The probabilities are weighted by the penalty function $f(r)$ where K is an assigned constant. Three distinct regions are presumed to exist around the CSP:

$$f(r) = 0 \quad r > r_1$$

$$f(r) = \left(1 - \frac{r - r_3}{r_1 - r_3} \right) \quad r_1 \leq r \leq r_3$$

$$f(r) = 1 \quad r < r_3$$

Figure 3b depicts some of the terms used. r_1 and r_3 are the maximum and minimum dimensions of an imaginary elliptical cylinder surrounding the analyte molecule (not the CSP) and r is the height of the analyte above the matrix. These conditions mean that for distances greater than r_1 above the matrix there exist no steric effects, for distances less than r_3 there is a maximum steric effect, and in between these limits an exponential increase in the steric effect is assumed.

Two examples of these calculations were given. One was the substituted trifluoroethanol, 3, binding with CSP analog 11 previously studied by Lipkowitz, and the second example was the association of an *N*-(3,5-dinitrobenzoyl)- α -methylbenzylamine on a phenyl urea CSP. Their results for the first example are better than those obtained by Lipkowitz and for the latter example are very close to experiment. The authors also addressed how the separation factor, α , is affected by the steric hindrance, K , and found that resolution becomes increasingly difficult as the steric effect increases. Their interpretation of this is that fewer binding sites become available to discriminate between the enantiomers but other interpretations can be envisaged. While the authors implemented only a rudimentary penalty function, their results are good and tend to support their methodology. Further testing is warranted and we highlight here that this is the first attempt at treating the matrix to which the CSP is tethered in the literature to date.

Up to this juncture we have considered atomistic molecular modeling approaches where "atomistic" means detailed atomic level information is being used in the calculations. Certainly application of quantum mechanics or molecular mechanics for geometry optimizations can be construed as atomistic modeling, as can the implementation of empirical force fields for molecular dynamics and Monte Carlo simulations. However, as mentioned in the beginning of this chapter another type of modeling exists that also uses detailed atomic level information, and it too has been used for explaining and predicting retention orders in chromatography. This category of atomistic molecular modeling invokes regression models as best exemplified,

perhaps, by Quantitative Structure-Activity Relationships (QSAR) [37]. These regression models truly are at the atomic level, taking into account atomic charges, three dimensional shape (topography), and atom connectivity (topology) in addition to other atomic or molecular descriptors that can be derived experimentally or from computation. This approach, being as viable as any other computational method, has had less use for analysis of chiral separations. In these cases the term Quantitative Structure-Enantioselective Retention Relationship (QSERR) has been coined [38].

An example of a QSERR study is by Carotti [39]. It involves the resolution of sulfoxides, 20, on a π -acid CSP containing (S,S)-N,N'-(3,5-dinitrobenzoyl)-*trans*-1,2-diaminocyclohexane (DACH-DNB), 19a. After establishing capacity factors, k'_S and k'_R for the first and second eluted enantiomers, respectively, as well as determining the separation factors, α , the authors began their computational studies. They used a variety of electronic descriptors for the analytes including electrophilic superdelocalizabilities (S^{homo}) and nucleophilic superdelocalizabilities (S^{lumo}) of various key analyte atoms in addition to S_{Ph}^{homo} , the sum of superdelocalizabilities of analyte ring carbons. These descriptors were deemed important because the analyte is known to be a π -donor while the CSP is a π -acceptor giving rise to charge transfer complexes. Steric descriptors like the Verloop sterimol parameters L , B_1 and B_5 , molar refractivity, MR , as well as Charton's steric parameter, (ν), were also used. Other descriptors like experimental stretching frequencies of the sulfoxide, ν_{SO} , were used but partition coefficients, $\log P$, were determined computationally.

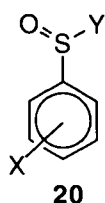
A set of regression equations were derived with a partial least squares (PLS) statistical analysis that allowed the authors to establish structural features most relevant to analyte affinity (capacity factors, k') as well as for enantioselectivity (separation factors, α). The best models are summarized in Table 1 below.

Table 1
Best regression equations for sulfoxides, 20, binding to CSP 19a

Equation number ^a	Regression equation	$(r_{cv}^2)^b$
3	$\log k_S = 0.902 S_{Ph}^{HOMO} + 0.942$	0.415
4	$\log k_S = 0.823 S_{Ph}^{HOMO} - 9.62 q_o - 6.01$	0.456
5	$\log k_S = 0.856 S_{Ph}^{HOMO} - 0.046 \log P + 1.05$	0.541
6	$\log k_S = 0.771 S_{Ph}^{HOMO} - 0.048 \log p - 10.2 q_o - 6.32$	0.609
7	$\log k_R = 0.960 S_{Ph}^{HOMO} + 1.04$	0.484
8	$\log k_R = 0.859 S_{Ph}^{HOMO} - 12.4 q_o - 7.89$	0.569
10	$\log k_R = 0.925 S_{Ph}^{HOMO} - 0.036 \log P + 1.117$	0.554
11	$\log k_R = 0.818 S_{Ph}^{HOMO} - 0.037 \log P - 12.8 q_o - 8.13$	0.650
12	$\log \alpha = 0.166 S_S^{LUMO} + 0.054$	0.507
13	$\log \alpha = 0.173 S_S^{LUMO} + 1.60 \nu_Y - 1.25 \nu_Y^2 - 0.444$	0.539
14	$\log \alpha = -3.94 q_o + 2.54 \nu_Y - 1.72 \nu_Y^2 - 3.63$	0.518

^aFrom Table 3 of reference 39.

^bCross-validated squared correlation coefficient.



For retention, the statistically best equations are 6 and 11, indicating retention is governed by the π -basic character of the analytes (expressed as S_{PH}^{homo}) as well as the net charge on the sulfoxide's oxygen (q_O) implicating analyte's ability to form hydrogen bonds to the CSP.

While these equations are useful for interpreting experimental results, the authors point out they have limited predictive ability (modest cross-validated squared correlation coefficients) and they only account for ca. 75% of the variance in $\log k'$, indicating that additional factors not yet recognized are at play in retention. Moreover, the regression coefficients of the independent variables in these two equations are not much different and cannot provide information about structural factors responsible for enantioselection. Equations 12 and 13 were derived, further implicating the electrophilic nature of the sulfoxide (S_S^{lumo}) as being important. Eventually equation 14 was obtained. Equation 14 suggests that steric factors together with charges on analyte oxygen (reflecting hydrogen bonding to analyte), along with dipole stacking are working simultaneously, but equation 14 does not allow one to assess the dominant driving force for molecular association leading to chiral recognition. Moreover, these equations, like the others above, have limited predictive abilities. This led the authors to use Comparative Molecular Field Analysis (CoMFA) of enantioselectivity.

The CoMFA method has been reviewed elsewhere [40]. The idea is to develop regression models wherein the molecular fields of a series of related molecules, which presumably bind to the same recognition site, can be compared. In this case the molecular fields being compared are the electrostatic and steric fields. The problem with this type of analysis is the decision of how best to align the molecules being compared. In this study the authors used benzyl phenyl sulfoxide as the template for superpositioning the other compounds because it displays the highest separation factor, α . Two conformational states of this template are the "folded" and "extended" shapes with the latter being 1 kcal/mol more stable. The nine heavy atoms of the phenyl-SO-C moiety were used for the superpositions for both conformers creating two alignments that were subjected to a PLS field analysis. The final 3D-QSERR models accounted for more than 94% of the variance in the $\log \alpha$ values with the extended conformational alignment being slightly better than that of the folded form.

From their statistics it is found that the electrostatic field contribution is about the same as the steric field contribution in the chiral recognition for both conformations. Using graphical representations the authors were able to select steric and electrostatic regions that would lead to enhanced chiral recognition as well as regions leading to decreased enantioseparation. In contrast to traditional QSERR, CoMFA is able to find π -acid - π -base interactions as a driving force for enantioseparation and it provides statistically better regression models. Moreover it has the advantage of providing detailed descriptions of the physicochemical interactions between this class of analyte with that particular Type I CSP. Their conclusions are that retention on the column is dictated by π - π interactions and hydrogen bonding, but that enantioseparation is governed primarily by steric factors modulated somewhat by polar and electrostatic properties of the SO group. The authors thus have applied, explained and extended earlier "3-point" attachment models used to rationalize chiral recognition in chromatography this way.

Because of this success, Carotti's group then carried out a QSERR and CoMFA analysis of α -alkyl- α -aryloxyacetic acid methyl esters, on the same CSP and then compared their results with those from the sulfoxide binding study described above

[41]. From their QSERR they find solute lipophilicity and steric properties as being responsible for analyte retention (k') while enantioseparation (α) varied mainly with electronic and steric properties. The main difference between the analytes is that the enantioseparation of the esters is correlated with steric parameters that scale linearly with $\log \alpha$ while the sulfoxides scale nonlinearly (parabolic), but this may be due to a computational artifact. The 3D-QSERR derived from field analysis revealed that while superpositioning of field maps for both analytes are not exactly the same, a similar balance of physicochemical forces involved in the chiral recognition process are at play for both sets of analytes. This type of atomistic molecular modeling, then, is a powerful adjunct to the type of modeling described earlier in this chapter and will, no doubt, be used more frequently in future studies.

7. TYPE II CSPS

Weinstein, Leiserowitz and Gil-Av were the first scientists to apply atomistic molecular modeling in chiral chromatography. Gil-Av had earlier discovered that chiral secondary amides were suitable as chiral stationary phases for gas-liquid chromatography [42]. To rationalize how these amides, in the form of a melt, act as a CSP, Gil-Av's chromatography group began working with Leiserowitz' crystallography group to determine the packing modes of mono-N-substituted primary amides [43]. Most of those amides contain H-bonded stacks with an interstrand spacing of 5Å, forming the basis for an intercalative model capable of accounting for enantiomer discrimination. The model, described below, assumes that bound analyte intercalates within the H-bonded array of the CSP matrix without disrupting that 5Å, H-bonded motif.

Conformational analysis and packing energies were computed with the QCFF/PI empirical force field. The $\phi\psi$ conformation energy maps gave minimum energy structures in fair agreement with their x-ray data. The authors then considered three molecules, related by the 5Å translation, and with amides fixed as in the crystal structure. Maintaining translational symmetry, a conformational analysis provided $\phi\psi$ maps agreeing even better with experiment. These authors described their GLC chiral recognition mechanism in a second paper [44]. The recognition model is based on the stacking which fulfills the requirements for linear hydrogen bonds, close packing of aromatic residues and allows for favorable contacts of long R groups. Both enantiomeric analyte molecules can intercalate and maintain the original H-bonding motif but the enantiomer having the same configuration as the CSP fits better than does its antipode (see Figure 4). This simple intercalation model explains the resolution of enantiomers on the N-lauroyl- α -(1-naphthyl)ethylamine CSP and it also explains the reversal of elution orders for α -phenylalkanoic acid amides.

An analysis of the intercalation energy was carried out using the same EFF as in their earlier paper. Calculations were done on ensembles of N-acetyl- α -(1-naphthyl)ethylamines where it was assumed that the intercalation of analyte would not perturb the host CSP structure. Three molecules were used to model the interactions: two rigid flanking host molecules and one flexible guest, all related by the 5Å stacking motif. The diastereomeric trimers are designated RRR and RSR as in Figure 4. The amide groups were held fixed in all calculations and the $\phi\psi$ map was generated by moving torsion angles ϕ and ψ in 10° increments.

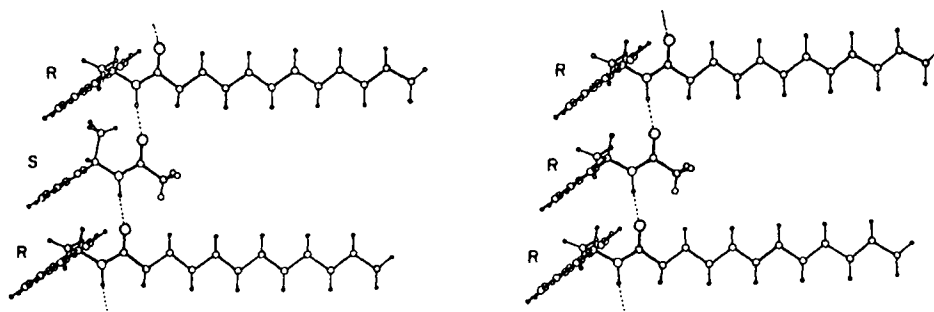
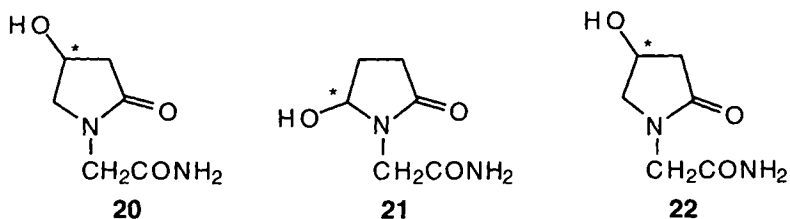


Figure 4. *N*-Trifluoroacetyl- α -phenylethylamine intercalated into *R*-*N*-lauroyl- α -1-naphthyl-ethylamine. Right: homochiral stack (favored), left: heterochiral stack (disfavored).

The conformation of the intercalated *R* analyte is approximately the same as *S*, given the assumptions made in the force field and the assumption of using a rigid host model. It is found that the faster eluting *S* enantiomer does not contain overlap of aromatic rings as does the more stable *R* enantiomer. Rather, the analyte's methyl group is sandwiched between two adjacent naphthyl rings while its naphthyl ring is nestled between adjacent methyls. The *RRR* minimum energy structure was found to be more stable than the *RSR* minimum by 9 kcal/mol. Such dramatic overestimation of preferential *R* binding (*R* is longer retained experimentally) was attributed to poor parameters used in the EFF and that the trimer is too rigid. In the melt the CSP is thought to be less rigid than the model but a geometry relaxation of the model complex was not attempted by the authors.

Another example of Type II CSPs are those derived from sugar-based polymers like amylose and cellulose. Derivatized celluloses form an especially versatile class of materials that have gained popularity because of their ability to resolve a wide range of drug sized molecules containing many different functional groups. Molecular modeling studies of enantiorecognition by the cellulose triphenylcarbamate CSP has been reported by Camilleri, Murphy, Saunders and Thorp [45]. They carried out separations of oxiracetam, 20, and two related molecules 21 and 22 on this CSP to assess retention times and separation factors.



For modeling purposes they generated a trisaccharide of β -1,4-linked D-glucoses, end-capped with methyls. Energy minimization with quantum mechanics gave a linear structure with intramolecular H-bonds. The hydroxyls were replaced with phenylcarbamate residues and the resulting trisaccharide was energy minimized again. The phenyl carbamate groups induce a helical twist to the polymer due to steric repulsions of the phenyls but the helicity of the CSP was dismissed as being

responsible for chiral recognition. After manual docking of analyte with CSP, energy minimization with molecular mechanics was performed. It was found that only for the R isomer of 20 could a viable association exist that would explain why R is longer retained than S.

Another example of computing interactions between analyte and CSP comes from Goya working collaboratively with Roussel [46]. Here the molecular modeling consisted of correlating the molecular mechanics binding energies of substituted benzenes, phenols and naphthalene with retention orders on cellulose triacetate (CTA). Eventually a linear relationship between $\log k'$ and the interaction energies was obtained. This study does not explain chiral interactions, but, it does introduce a picture of what that important CSP looks like.

Another example of modeling the structure of this type of CSP is presented by Francotte and Wolf [47]. They prepared benzoylcellulose beads, in a pure polymeric form as a sorbent, for the chromatographic resolution of racemic compounds like benzylic alcohols and acetates of aliphatic alcohols and diols. Their experimental results implicated multiple interaction sites to be involved in the complexation. Rationalizing the interaction mechanism required a more systematic investigation of the factors influencing separations and, to address the structural features of the cellulose tribenzoate, they carried out molecular modeling with molecular mechanics. The key question being addressed is: to what extent is the polysaccharide backbone exposed to small molecules when sterically encumbered benzoates are attached?

Representative decameric chain segments were generated by excising the third unit of an energy minimized hexamer which then served as their "monomer." This monomer was polymerized, computationally, using the glycosidic bridge angles between the two middle segments to create the polymer backbone. This way the computational artifact of having terminal, end-groups are eliminated. Color-coded molecular graphics displays of the decameric strands in two low energy conformations revealed the sugar residues are able to interact, at least partially, with small molecules so that the chiral discrimination does not come solely from the benzoyl groups.

Perhaps the most detailed and most comprehensive computational study of chiral analytes binding to Type II CSPs comes from Okamoto's laboratories. This group has taken the lead role in developing sugar-based stationary phases and in analyzing chiral discrimination on polysaccharide derivatives [48]. Their molecular modeling involved evaluation of binding energies of \pm *trans*-stilbene oxide, a diphenyl epoxide, and \pm *trans*-1,2-diphenylcyclopropane (the same system lacking a hydrogen bond binding site) complexing to a cellulose triscarbamate polymer (CTPC) [49]. The reason this particular polymer was selected for study computationally is because it resolves the aforementioned enantiomers with amazing efficiency, but more importantly because this polymer is the only one to date amenable to NMR spectral analysis using polar solvents (other such polymers lose their resolving ability in polar solvents) and comparison with NMR results can be made.

The polymer was constructed by fully optimizing (CHARMM) a tris-carbamate monomer containing methoxy groups at the 1 and 4 positions. The resulting structure was polymerized, computationally, by linking them to form an octamer with a left-handed, three-fold (3/2) helix similar to a CTPC structure previously proposed in the literature based on X-ray fiber diffraction studies. This octamer was fully geometry optimized and molecular dynamics carried out. From the trajectory file was obtained a set of new structures that were energy minimized but no lower energy structures could be found.

It was felt (from previous experimental work) that the most important adsorbing site for the stilbene oxide involves the N-H hydrogens of the carbamate groups. Accordingly, the authors set up sampling boxes centered around the carbamate's N-H hydrogen. A schematic of both the octamer and the sampling grid are depicted in Figure 5. Note that a sampling box was set up for all three amides on each monomer, and that only monomers 3-6 were sampled to avoid the influence of the end groups (experimentally these polymers have a degree of polymerization ~ 100). Also note that different grid sizes were used ($r = 3\text{\AA}-6\text{\AA}$) but the grid mesh was initially fixed at 1\AA .

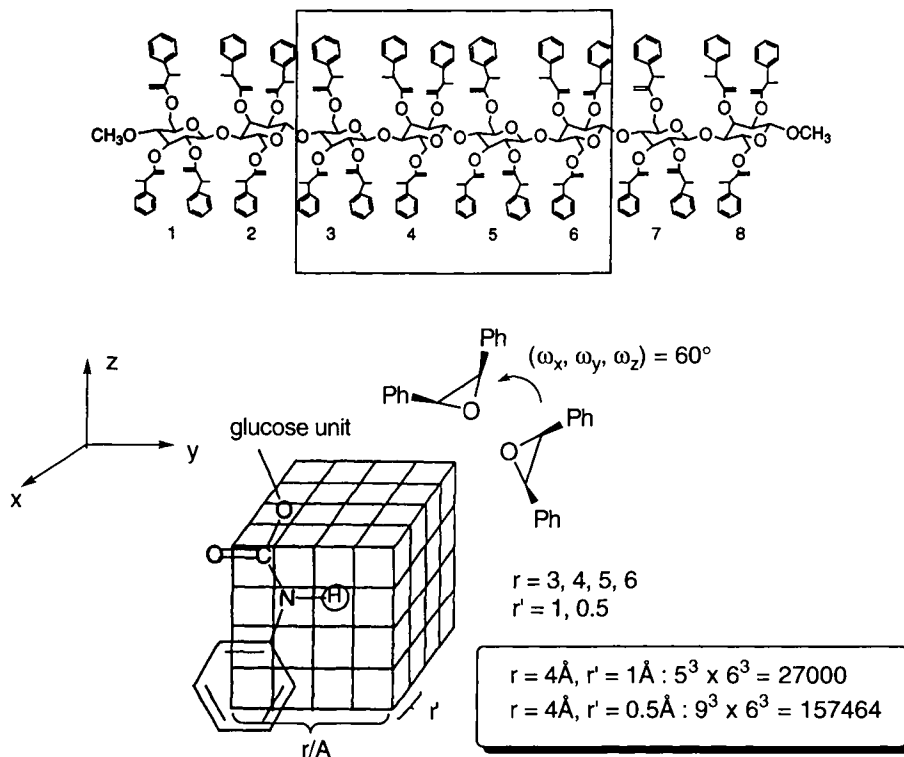


Figure 5. Top: Schematic of cellulose trisphenylcarbamate octamer. Sampling was done within the enclosed box. Bottom: Cubic sampling box of dimension r and mesh size r' in \AA . Analyte is rotated in 60° increments about x, y and z axes at each grid point. Reproduced with permission from ref. 49.

At each grid point the (R,R)-(+)-epoxide or (S,S)-(-)-epoxide was rotated in 60° degree intervals around the x, y , and z axes, individually. The calculations involved a rigid analyte interacting with a rigid CSP. Using this sampling strategy the authors were able to tabulate the minimum interaction energy between analyte and each carbamate moiety on each monomer. Those energies varied substantially, depending on the glucose as well as the position of the carbamate within a particular monomer. Next, a finer grid mesh of 0.5\AA was used and the lowest interaction energies

determined. Both grid meshes provided results in agreement with retention orders (the R,R enantiomer elutes before the S,S enantiomer) but the computed differential interaction energies were substantially overestimated. These energies are not averaged in any way and do not correspond to free energies. Rather, they are presumed to be the global minima of the diastereomeric complexes. Nonetheless the authors were able to find the most probable analyte binding site, which, for the epoxide, is in a chiral groove existing along the main chain of the polymer backbone. Moreover, the authors were able to discern the most important types of interactions including hydrogen bonding to the oxirane ring and π -stacking of carbamates with the analyte phenyl groups. For the diphenylcyclopropane where the hydrogen bonding is absent these researchers found no enantioselective preference from their calculations, and, none was observed experimentally.

Okamoto's group extended this work to a related CSP; cellulose tris(5-fluoro-2-methylphenylcarbamate). In a recent paper they revealed, from a multifaceted study including chromatography, detailed 2-D NMR analyses and molecular modeling, that this type II CSP is capable of resolving enantiomers like 1,1'-bi-2-naphthol and 2,2'-dihydroxy-6,6'-dimethylbiphenyl with large separation factors ($\alpha > 3$) and that both ^1H and ^{13}C NMR have large chemical shift changes for the enantiomers. Hence they focused their attention on these kinds of axially dissymmetric analytes to address where and how chiral selection takes place [50].

From their HPLC and NMR studies they were able to demonstrate that the S-1,1'-2-binaphthol, which is more tightly bound to the CSP and accordingly has the longer retention time, binds to a chiral groove on the CSP and is directed toward the glucose H_2 proton. This enantiomer displays several intermolecular NOEs whereas the R antipode has none. To further explore the mode of binding Okamoto's group again constructed a left-handed, three-fold ($3/2$) helix but this time used the Dreiding force field in lieu of CHARMM (no reason provided). The R and S analyte structures were derived from a published crystallographic analysis of the racemate, and they too were fully optimized with the Dreiding force field. The S enantiomer was then manually docked into the groove of the CSP in an orientation satisfying all the NMR data and the entire complex was then energy minimized. The helical chiral cavity along the polymer backbone contains polar carbamate groups inside the groove and hydrophobic aromatic groups outside the groove. From their calculations the authors find that the S enantiomer can simultaneously form two hydrogen bonds to suitably placed polar groups along the chiral groove whereas the R enantiomer can only form a single hydrogen bond. Additionally, the authors were able to rationalize why more bulky systems like 10-10'-dihydroxy-9,9'-biphenanthracene are not resolvable on this particular CSP. While their work entailed mainly molecular mechanics energy minimizations rather than full molecular simulations, the results were particularly useful for discerning where and how enantioselection takes place when axially dissymmetric ligands bind to helical polysaccharides.

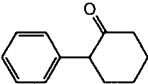
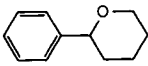
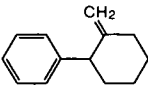
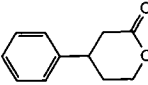
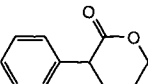
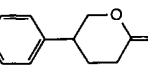
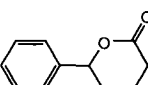
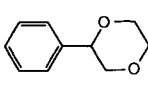
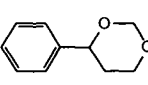
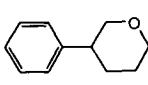
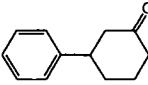
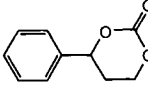
Most of the computational studies of Type II CSPs do not consider the CSP directly. Instead, regression models are constructed to explain how a set of probe molecules interact with the CSP. We now present selected examples from the literature illustrating the diversity of such computational methodology used by chemists to address how these CSPs work.

The first example is from Isaksson, Wennerstrom and Wennerstrom [51] who considered analyte binding to cellulose triacetate (CTA). They used statistics to assess the relationships between chiral recognition and analyte symmetry. Rather than attempt to compute the actual binding constants they addressed how symmetry

influences differential binding constants for chemically similar compounds. Their conclusions are that a better resolution for symmetrical analytes containing a proper axis of symmetry exists than for asymmetrical enantiomer pairs, and, that the higher the order and number of symmetry axes in those analytes, the better the separation is expected to be.

Another example, also directed towards understanding how CTA works, was carried out by Wolf's group at Ciba-Geigy [52]. They investigated the influence of the chemical structure for a series of related racemates. The molecules, their capacity factors, k' , and separation factors, α , are presented in Table 2. It is to be noted that capacity factors for the first eluted enantiomer are all about the same while those of the second enantiomer span a large range. On CTA, then, a large k'_2 value almost always leads to a high α . Thus these authors decided to use k'_2 as the relevant experimental parameter to be correlated with molecular descriptors, which in turn were computed. They attempted first to define quantitative molecular descriptors for these 12 analytes and then to correlate those descriptors with their chromatographic data. These molecular properties, it must be emphasized, are independent of the configuration of the analyte so no stereochemical arguments can be made as done in the previous section of this chapter.

Table 2.
Chromatographic results: capacity factors k'_1 and k'_2 and separation factor α

Racemate	k'_1	k'_2	α	Racemate	k'_1	k'_2	α
	1.2	2.6	2.2		1.0	3.3	3.3
	0.8	0.8	1.0		1.7	1.7	1.0
	1.3	4.2	3.3		1.5	1.5	1.0
	1.3	29.7	23.3		2.2	12.0	5.4
	2.4	11.1	4.6		0.8	0.8	1.0
	1.2	1.6	1.4		1.4	9.5	6.8

Two criteria seemed important for chiral separations. First, as had been noted by many authors, the shape of the analyte seems critical and second, all compounds containing an oxygen adjacent to the stereogenic center give rise to large k'_2 values

indicating negative charge adjacent to that center enhances resolution. Conformational studies (to determine a shape descriptor) were carried out with molecular mechanics and atomic charges derived from quantum mechanics were used to create molecular electrostatic potential maps (to help define a suitable electronic descriptor). It is presumed by chromatographers that flat molecules generally fit better into the chiral cavities or crevices of CTA than do nonplanar molecules. The authors examined the shape and rigidity of the analytes by Boltzmann weighting their conformational energies. They found that the twelve analytes could be partitioned

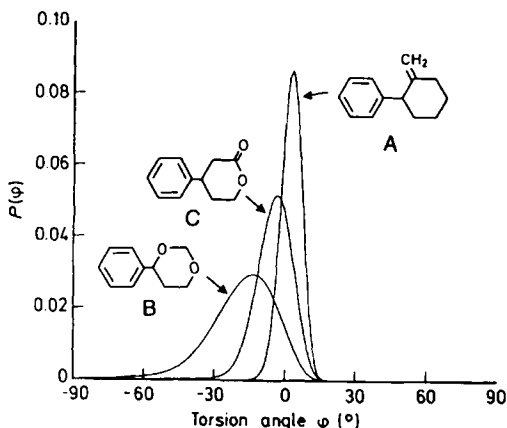


Figure 6. Boltzmann distribution of the conformational states of the three classes of compounds displaying very restricted, intermediate and high conformational freedom.

into three categories, depicted in Figure 6. One grouping, represented by A in Figure 6, has a very narrow probability distribution and may be considered "rigid". In Figure 6 a dihedral angle of zero means the phenyl ring is orthogonal to the saturated ring and these molecules will never become planar and thence most likely will bind poorly to the CSP. The second category, represented by molecule B, has a broad distribution of conformational states and can adopt near-planar shapes. The third category is between the "stiff" and the "flexible" analytes.

It turns out that the flatness of the analyte alone is not the only factor needed for good complexation with CTA. To probe the electronics of these interactions the authors computed molecular electrostatic maps as in Figure 7 and extracted from them a normalized electrostatic interaction energy.

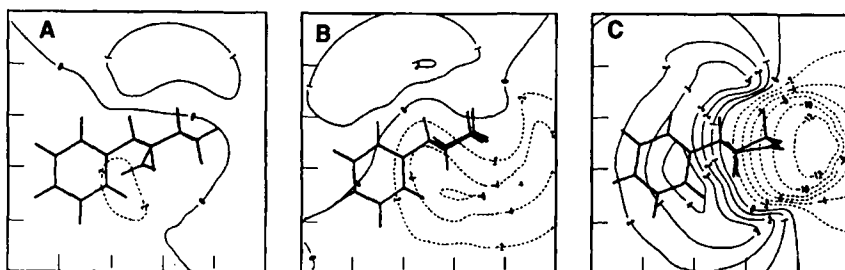


Figure 7. Molecular electrostatic contour maps for molecules A, B and C in Fig. 6. Contour lines are in units of kcal mol⁻¹, full lines represent positive regions and broken lines represent negative regions.

The two molecular descriptors determined as being most important were $[\Omega]$, a measure of flatness and E_{00} , a measure of negative charge distribution around the

stereogenic carbon. Using these two simple descriptors, a regression model was developed (eqn. 17)

$$\ln k_2^{(i)} = AE_{00}^{(i)} + B[\Omega]^{(i)} + C \quad (i = 1-12) \quad (17)$$

where $A = -0.122$, $B = 0.0211$ and $C = 0.497$. Using this model the authors predicted k_2' values for comparison to experiment and found good agreement (correlation coefficient = 0.96).

Other attempts to quantify the effect of structural parameters on the separation of enantiomers by CTA have been reported. An example comes from the work of Roussel and his group. They use full factorial design to quantify the effect of structural parameters (factors) on either the separations or capacity factors (responses) of atropisomers on CTA [53]. This way one can quantitate the variables responsible for retention and separation. The three structural modifications are shown in Figure 8.

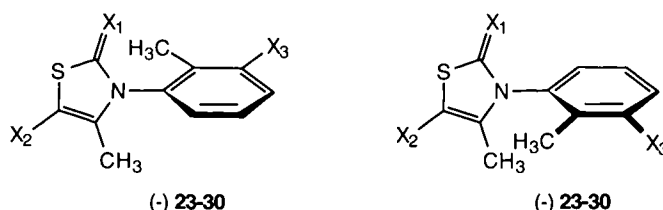


Figure 8. Structure of the N-arylthiazoline-2-thiones and N-arylthiazolin-2-ones used in these studies. X_1 = oxygen or sulfur, X_2 = hydrogen or methyl, X_3 = hydrogen or methyl.

Because there are three changes being made there must be $2^3 = 8$ molecules (for three changes at two levels) to fully explore the recognition process. X_1 can be oxygen (level -1) or sulfur (level +1); X_2 can be hydrogen (level -1) or methyl (level +1); X_3 can be hydrogen (level -1) or methyl (level +1). These eight test probes were synthesized and resolved on CTA. Assuming that the influence of substituent patterns for each enantiomer can be linearized, by difference, one can then assess the enantioselectivity. This linearization is expressed by equation (18).

$$Y = c_0 + c_1X_1 + c_2X_2 + c_3X_3 + c_{12}X_1X_2 + c_{13}X_1X_3 + c_{23}X_2X_3 + c_{123}X_1X_2X_3 \quad (18)$$

Here Y is the response, X_1 , X_2 , X_3 are the primary effects and X_1X_2 , X_1X_3 , X_2X_3 , $X_1X_2X_3$ are the bilinear and trilinear cross terms. The c 's are the coefficients for each term determined from solving the eight equations, from eight experiments, by replacing the X_i with +1 or -1 according to the experiment. A positive coefficient means that upon going from the low level to the high level of that factor, the response is increased. A negative coefficient, in turn, means the response is decreased. Table 3 lists the molecules studied, their design levels and chromatographic results.

Table 3
Compounds, design levels and responses

Compound	Design level			Response		
	X_1	X_2	X_3	$k(+)$	$k(-)$	k_2/k_1
23	-	-	-	2.05	1.66	1.24
24	+	-	-	1.52	2.30	1.51
25	-	+	-	0.66	0.55	1.21
26	+	+	-	0.87	0.78	1.12
27	-	-	+	0.81	1.92	2.36
28	+	-	+	0.96	3.09	3.20
29	-	+	+	0.35	0.35	1.00
30	+	+	+	0.64	0.61	1.04

Equation (19) is derived for the (+) enantiomer and equation (20) for the (-) enantiomer.

$$k(+)=0.98+0.015 X_1-0.35 X_2-0.29 X_3+0.11 X_1 X_2+0.09 X_1 X_3+0.15 X_2 X_3-0.075 X_1 X_2 X_3. \quad (19)$$

$$k(-)=1.40+0.29 X_1-0.83 X_2+0.08 X_3-0.16 X_1 X_2+0.07 X_1 X_3-0.17 X_2 X_3-0.06 X_1 X_2 X_3. \quad (20)$$

For the dextrorotatory isomer the most dominant primary interactions are X_2 and X_3 . The sign and the magnitude of these coefficients indicate that upon going from a low level (H) to a high level (Me) at either C_5 on the thiazoline or C_3 on the aryl ring, steric effects between analyte and CSP decrease the capacity factor. The relatively small coefficient for X_1 means replacing the ketone with a thione has a minor influence on the dextrorotatory enantiomer's retention.

For the levorotatory isomer the most important primary effects are X_1 and X_2 . Hence for this enantiomer replacement of the carbonyl with a thiocarbonyl has a substantive impact on retention whereas replacing the H with Me on the aryl's C_3 has almost no effect on retention. The dominant primary effect is still due to X_2 .

Roussel and Popescu [54] extended this work by developing a lipophilicity parameter, $\log k'_w$. The authors were able to explain the relationship between chiral retention of the enantiomers and their lipophilic interactions with the CSPs. Quantification of the influence of structural parameters X_1 , X_2 and X_3 was also possible. The relationship between lipophilicity and chiral chromatographic behavior was explained for compounds 23-30 and an extension to other alkyl substituted atropisomers was made. A related study concerning the resolution of 23-30 on various *p*-methylbenzoyl cellulose beads has also been published [55] but will not be described here because it employs the same methodology as above.

Using this kind of analysis one can address the sensitivity of functional group response to a particular CSP. Focused information concerning functionality along with the physicochemical causes of enantioselection can be derived. This type of computational approach is complementary to that described by Isaksson [51] and by

Wolf [52], and, it has great utility for unraveling how the complex intermolecular forces responsible for chiral separations in chromatography.

Finally, we mention here the work by Ning [56] on salting effects in reversed mobile phases for chiral separation of both *cis* and *trans* benzonaphthazepine enantiomers on a cellulose tris(3,5-dimethylphenylcarbamate) CSP. The salting-in effect of sodium perchlorate was noted to make the analytes more soluble in the mobile phase so that the CSP can selectively retain the four stereoisomers. The salting-out effect of sodium chloride induces hydrophobic self association and, accordingly, the author proposed that sodium chloride works differently than conventional ion-pair reagents used in nonchiral reverse-phased chromatography.

Experimentally, the retention order is (-) *trans*, (-) *cis*, (+) *trans*, (+) *cis*. This order is quite unusual and was rationalized by atomistic molecular modeling. Using molecular mechanics the lowest energy conformers for both *cis* and *trans* analyte were computed. Ning then discovered that the *cis* isomer can be superimposed almost perfectly on top of the *trans* diastereomer. In contrast, two enantiomers do not possess such perfect matching. Hence, when a *cis* diastereomer encounters a *trans* diastereomer they are forced, by hydrophobic forces, to form a "dimer" and they move into and out of the mobile phase until they become trapped in a suitable cavity of the CSP. There they are further discriminated. By comparison, the enantiomers do not match well when forced together and are proposed to be moving about individually, randomly associating with the CSP and being well-resolved. This proposal rationalizes why the corresponding diastereomer rather than the enantiomer follows in the separation. Finally, because the *trans* isomer has a smaller pucker angle by 6 degrees than does the *cis* isomer, it has less steric bulk and should elute less quickly than the *cis* isomer. This is consonant with experiment.

Most of the molecular modeling studies involving Type II CSPs, as illustrated above, do not directly involve computations of the analyte with the CSP to discern where and how chiral recognition takes place. The reason for this is, clearly, the lack of structural information about these polymeric CSPs. This is in contrast to modeling studies of Type I CSPs described in an earlier section of this chapter and to those computational studies of Type III CSPs discussed below.

8. TYPE III CSPS

Type III CSPs work by forming inclusion complexes. Two main categories of Type III CSPs have been studied computationally: cyclodextrins (CDs) and their derivatives, and, crown ethers and related non-natural guest-host cavities. In this chapter we address mainly the cyclodextrins but provide an example of a completely synthetic receptor that is used as a type III CSP. Cyclodextrins are cyclic oligomers of α -D-glucose that exist in varying sizes depending on the number of monomers forming the macrocycle. The hexamer is called α -CD, the heptamer is β -CD, the octamer is γ -CD. Other cyclodextrins exist but these three are most commonly used for chiral separations.

Armstrong was one of the first scientists to discover, implement and market covalently linked CDs as viable CSPs for chiral chromatography. He was also the first to use computational chemistry to explain how these chiral host molecules discriminate chiral guests in chromatography. In a series of papers, Armstrong used molecular graphics to represent how β -CD separates diastereomers [57] and enantiomers [58, 59]. These graphic images were presented in color to highlight similarities and differences of analyte binding, but these studies did not involve complete energy

minimizations. Using rigid cyclodextrin, the authors allowed only the important torsion angles and the location of the analyte in the macrocycle to change. The modeling highlighted the importance of the secondary and tertiary hydroxyl groups along the rim of the macrocycle as being responsible for the resolution of enantiomers, and this allowed the authors to rationally design derivatives of CDs to optimize a particular separation.

Berthod, Chang and Armstrong [60] later devised a scheme useful for attributing individual substituent contributions to chiral recognition. They carried out their studies with a derivatized β -CD CSP called R-NEC- β -CD as well as with the corresponding S phase, S-NEC- β -CD, both of which had been designed and constructed in Armstrong's laboratory. Their objective was to be able to predict whether or not an analyte containing four different substituents connected to a single stereogenic center would be resolvable on these CSPs. In their approach the separation factor, α , is due to the chiral interactions for enantiomer 1 and enantiomer 2 by $\alpha = \exp \Sigma [(\Delta G_{c1} - \Delta G_{c2}) / RT]$. The differential free energy due to the chiral interaction of the two enantiomers, $\Delta\Delta G_c$, was divided into four terms, each representing one of the four groups attached to the stereogenic center as in equation 21.

$$\Delta\Delta G_c = (\Delta G_{c11} - \Delta G_{c12}) + (\Delta G_{c21} - \Delta G_{c22}) + (\Delta G_{c31} - \Delta G_{c32}) + (\Delta G_{c41} - \Delta G_{c42}) \quad (21)$$

It is assumed that the chiral interaction contributions are independent of one another, are additive and, accordingly, predictive. A total of 126 compounds were analyzed on these two CSPs. However, only 81 unique functional groups exist in the dataset due to redundancy. The hydrogen substituent was arbitrarily set = 0 cal/mol for a chiral free energy contribution. The authors generated one equation (eqn. 21) for each compound studied, and, setting $\Delta G_H = 0$ cal/mol, solved the simultaneous equations while the error function $E = \Sigma | \alpha_{calc} - \alpha_{obs} |$ was minimized. If the substituent has a positive value it means an enhanced chiral recognition by the NEC- β -CD CSP exists compared to H, and, a negative energy value means the opposite. While this method does not allow prediction of an elution order, the substituent constants (for those particular CSP's only) can be used to make an estimate about enantioselectivity of an unknown analyte simply by adding the energy contributions of the four substituents connected its stereogenic carbon.

A slightly different approach by Roussel and Favrou [61] was taken toward understanding chiral separations by cyclodextrins and quantifying the effect of substituents. Using compounds 23-30 described above, the authors carried out chiral separations using β - and γ -cyclodextrins as a chiral mobile phase additive. The full factorial design methodology was applied to k'_o , (retention without cyclodextrin), $k'(+)$, $k'(-)$ and α for two different achiral stationary phases. In the presence of γ -CD on a nonendcapped phase, equations 22-25 were derived.

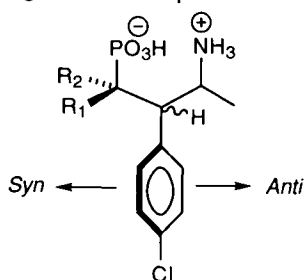
$$k'_o = 86.43 - 2.77 X_1 + 33.39 X_2 + 31.70 X_3 - 6.19 X_1 X_2 - 2.74 X_1 X_3 + 12.73 X_2 X_3 - 3.43 X_1 X_2 X_3 \quad (22)$$

$$k'(+)=20.92-6.15 X_1+7.31 X_2+8.17 X_3-2.65 X_1 X_2-2.81 X_1 X_3+3.00 X_2 X_3-1.27 X_1 X_2 X_3 \quad (23)$$

$$k'(-)=20.65-6.25 X_1+7.14 X_2+8.44 X_3-2.65 X_1 X_2-2.71 X_1 X_3+3.17 X_2 X_3-1.27 X_1 X_2 X_3 \quad (24)$$

$$\alpha=1.024+0.016 X_1+0.008 X_2-0.024 X_3+0.008 X_1 X_2-0.016 X_1 X_3-0.008 X_2 X_3-0.008 X_1 X_2 X_3 \quad (25)$$

From eqn. 22 one finds that without γ -CD, two structural features, X_2 and X_3 , have a predominant influence on retention. A change in electrostatics by converting the carbonyl to thiocarbonyl does not much influence the capacity factor. In contrast, when γ -CD is present, equations 23 and 24 show X_1 , X_2 and X_3 to affect retention similarly for each atropisomer, and, these influences are weak. In the presence of the chiral modifier γ -CD, then, the change from carbonyl to thiocarbonyl becomes important with the latter being more retained. Finally, from eqn. 25, it is seen X_1 and X_3 are most important.



- 31** $R_1 = R_2 = H$
32 $R_1 = R_2 = F$
33 $R_1 = H, R_2 = F$
34 $R_1 = F, R_2 = H$

Next we consider the chromatographic behavior of phaclofen, 31, and its difluoro derivative, 32. These molecules along with the diastereomeric monofluoro species 33-34 have been chromatographed on acetylated β -CD. In methanol with aqueous triethylamine-acetic acid buffer, 31 is not resolved but 32 is with $\alpha = 1.11$. Furthermore, 34 is resolved but not 33.

A theoretical model for this behavior, in terms of the asymmetry in the π -facial molecular electrostatic potentials of the phenyl ring, was proposed by Camilleri and Rzepa [62]. Using semiempirical molecular orbital theory the electrostatic potentials on the *syn* and *anti* faces of the aryl ring were determined. Those analytes with larger electrostatic differences between *syn* and *anti* faces appear to correlate with separations but precisely how this electrostatic asymmetry works was not explained.

Most of the aforementioned studies represent quantitative structure-retention relationship studies where a series of analytes are used as probes of enantiodiscrimination. There are, however, a number of atomistic molecular modeling studies where the interactions of chiral guests (analytes) with chiral hosts (CSPs) are explicitly determined. Here guest and host are considered as transient diastereomeric complexes and both liquid and gas chromatographic separations have been modeled.

Lipkowitz [63] used molecular dynamics simulations to answer the following questions: a) What are the intermolecular forces responsible for analyte binding to the CSP? b) Where on or in the host does the analyte bind? c) What are the differential interactions giving rise to chiral discrimination? d) What differences do R and S-enantiomers experience in the CD cavity? e) Are existing chiral recognition mechanisms valid? His computational work was based on experimental separations

carried out by Armstrong [64] who had previously determined that R-tryptophan is more tightly bound to an α -CD CSP than is S-tryptophan (separation factor in aqueous solution is $\alpha = 1.20$).

Using the CHARMM force field for molecular dynamics, Lipkowitz' simulations reproduced the correct retention order and separation factor from the chromatography experiments. Their simulation results also agreed with both intermolecular and intramolecular NOE observations from NMR experiments they carried out.

Because intermolecular hydrogen bonding was considered to be important, the authors evaluated the number and kinds of intermolecular H-bonds between guest and host. They found that not only does the more retained R enantiomer form a greater number of H-bonds than does S, but that these H-bonds are usually simultaneous, multiple-contact H-bonds between guest and host. A cartoon summarizing their results is depicted in Figure 9.

Three key features emerge from this cartoon. First, both complexes are highly localized on the interior of the CD with R binding to one side of the macrocycle and S to the other. Second, the R enantiomer forms almost twice as many intermolecular H-bonds (2662) as does S (1307) and, as pointed out above, they are of the multiple-contact type. Third, the hydrogen bonding occurs primarily from tryptophan's carboxylate and indole N-H but not the ammonium group. Based on this the authors confirmed an earlier recognition model but pointed out that a high degree of localization is tantamount to a tight fit in the CD cavity.

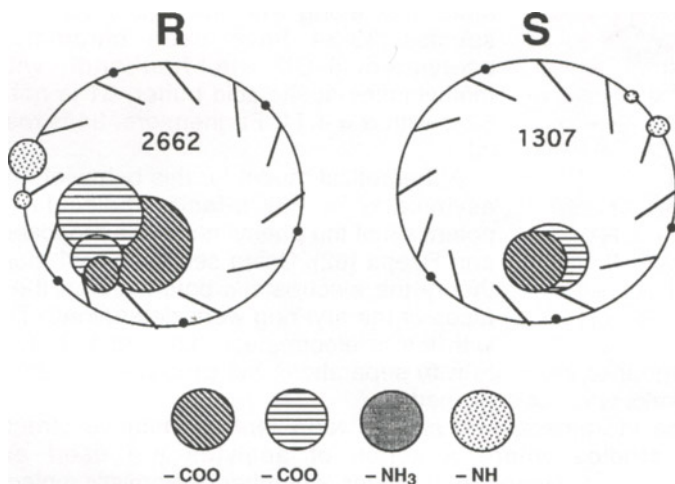


Figure 9. Graphical representation of the intermolecular hydrogen bonds of (R)- and (S)-tryptophan with α -cyclodextrin. The large circle represents the macrocyclic host. The small black dots on that circle are the acetal linker oxygens, and the lines attached to the circle represent the unidirectional C2 and C3 hydroxyl groups. The cross-hatching indicates the atoms or groups of atoms on the tryptophan that are forming the intermolecular hydrogen bonds. The size of the cross-hatched circle corresponds to the number of hydrogen bonds formed during the simulation. The centers of the cross-hatched circles are placed at the mean positions of the hydrogen bond contacts.

The concept of chiral recognition on the exterior of the CD has been raised by Lipkowitz and others and is especially relevant to gas chromatography. One of the predominant forces for guest-host complexation in aqueous phase liquid chromatography is the hydrophobic force. This is absent in the gas phase and consequently it is not clear what forces induce an analyte to bind as an inclusion complex in CD stationary phases in gas chromatography. Indeed, from an extrathermodynamic analysis of analyte binding to derivatized cyclodextrins in the gas phase, Armstrong suggests both exterior and interior binding modes are possible [65]. Exterior binding and/or partial inclusion complexation seems consistent with molecular mechanics and dynamics calculations showing these otherwise toroidal macrocycles to collapse upon themselves in the gas phase [66]. However, all publications to date assume binding takes place in the interior.

The pioneering paper on using molecular dynamics simulations to understand chiral gas chromatographic results was done by König's group [67]. Experimentally they found the S enantiomer of methyl-2-chloropropionate to be more retained on Lipodex D ([heptakis(3-O-acetyl-2,6-di-O-pentyl)- β -CD] coated on a capillary column) at 333 K. A large separation factor, $\alpha = 2.02$, corresponding to a $\Delta\Delta G = 2 \text{ kJ/mol}^{-1}$ was observed and an attempt to discern the structural features of the transient complexes was made.

Using a neutron diffraction structure of β -CD, the authors homogeneously fixed the C2 and C3 substituents upward and the C6 pentyl groups downward thus forming an initial conical-shaped CD (most CDs have this shape). The analytes were then placed into the interior of the resulting cavity in two orientations, "up" or "down." The authors found the "down" orientation led to immediate expulsion of guest from the host cavity upon warming and equilibration of their guest-host complex. The R-"up" orientation migrated outside of the cavity but remained near the hydrophobic sidechains and the S-"up" complex was found to be most stable. These results are consistent with the GC results and also with NOE intensities from NMR studies that further showed the methyl ester to be near the C3 groups of the CSP.

In a follow-up paper the authors carried out longer simulations [68]. The computed complexation energy difference, favoring the S enantiomer, is 5.75 kcal/mol at 300K (experimental $\Delta\Delta G = 0.65 \text{ kcal/mol}$), and, at 333K it is favored by 1.12 kcal/mol (experimental $\Delta\Delta G = 0.47 \text{ kcal/mol}$). In this paper the authors considered the shape of the host, showing "self-inclusion" takes place when no guest is present. They also addressed the shape of the analyte in both the free and complexed states. From their simulations they were able to deduce the time-averaged orientation of the analyte in the CSP host cavity, evaluate detailed structural features of the guest-host complex, and, to describe important intermolecular distances between the enantiomeric analytes and chiral cavity.

Koen de Vries, Coussens and Meier likewise found that a combined molecular mechanics and molecular dynamics approach is a valuable tool for rationalizing qualitative gas chromatographic trends [69]. Experimentally they evaluated the thermodynamic parameters (ΔG , $\Delta\Delta G$, ΔH , $\Delta\Delta H$, etc.) for guest-host complexation of six analytes on a variety of derivatized CD columns. Their interpretation of the computational results is that one enantiomer fits the CD cavity better than the other resulting in a larger interaction energy and greater loss of mobility.

Kobor, Angermund and Schomburg [70] also used molecular modeling to examine how polar and nonpolar analytes bind to derivatized CD's used as selectors in gas chromatography. Their goals were to systematically explore potential GC-compatible chiral selectors that might be more universal with regard to their application as

stationary phases. They prepared 2,3-di-O-methyl-6-O-*tert*-butyldimethylsilyl- β -CD (TBCD) for comparison to the more common permethyl- β -CD (PMCD) CSP. Their idea was to narrow the secondary opening of the CD cavity and to block the opening on the primary side of the CD giving rise to a structurally unique CSP. The TBCD was dissolved in a lipophilic polysiloxane and coated on treated and nonpretreated silica capillary surfaces. The test compounds studied were limonene and 1-phenylethanol. S-limonene is eluted before R and $\alpha = 1.078$ on TBCD and $\alpha = 1.029$ on PMCD. The R enantiomer emerges before S for the polar alcohol with $\alpha = 1.044$ on TBCD and 1.048 on PMCD.

Their molecular modeling involved converting the x-ray structure of permethyl- β CD to TBCD. The structure was energy minimized and subjected to molecular dynamics at high temperatures to probe its conformations. Likewise the analyte conformations were probed and these conformers were placed in the interior of the CD cavity. A Monte Carlo docking algorithm was used to generate guest-host complexes which were then geometry optimized by molecular mechanics. A distance dependent dielectric constant was used to simulate the presence of the polysiloxane matrix. This way a range of energies for each diastereomeric complex was generated as were average energy differences that were in agreement with their experiment. Moreover, from their simulations, they were able to determine the number of inclusion complexes formed between each guest and host, with the more stable diastereomer always having the larger number of inclusion complexes. Based on their modeling studies and experimental observations the authors made several conclusions about how the analytes fit into the TBCD cavity as well as the influence of CSP rigidity on chiral discrimination.

Black, Parker, Zimmerman and Lee [71] have used molecular modeling tools to see if they could correctly predict the retention order of polar and nonpolar analytes that are resolvable on cyclodextrin stationary phases in gas chromatography. They also wanted to address the subtle but important issue of induced-fit binding when these analytes associate with their macrocyclic receptors. The cyclodextrins considered include permethylated β -cyclodextrin and native α -cyclodextrin. The analytes studied include α -pinene, a nonpolar unfunctionalized hydrocarbon, and three cyclohexanetriol derivatives that are polar and capable of forming a variety of hydrogen bonds to the CSP. All analyte molecules are rigid having only a single conformation, thereby simplifying the analysis.

The authors generated the α -CD *de novo*, in the building module of their modeling program. They carried out MD for 60 ps and obtained an average structure from the last 50 ps that was subsequently energy minimized with the Biosym CVFF force field. The permethyl β -CD was retrieved from a published inclusion complex wherein the guest molecule was removed and the remaining macrocycle energy minimized with the same force field. The analyte molecules were built and energy minimized and then a series of rigid body grid searches carried out. Boltzmann weighted averages were obtained from these calculations but the results were uniformly poor; four of the five examples had incorrectly predicted retention orders.

The authors then extracted the low energy structures from their grid searches and fully optimized the entire complex of each. A large number of degenerate structures resulted that were subsequently deleted, based on an RMS fitting criterion of similarity. The Boltzmann weighted energies from this gave good results; four of the five examples were correctly predicted and in some instances the separation factors had the same trends as the experiments. The structural changes of the cyclodextrin cavities were small, but, apparently large enough to better accommodate the guest

molecules and to change the statistically weighted energies to agree with experiment. The authors conclude that induced-fit behavior of cyclodextrin complexation is important.

The computational studies described above are representative examples meant to illustrate the diversity of computational techniques used to assess chiral recognition by cyclodextrins in chiral chromatography. Other published examples include the use of molecular mechanics to describe shapes of aminoalkylphosphonic acids binding to a covalently linked acetylated cyclodextrin [72], the computation of free energies of atenolol binding to a perphenylcarbamate- β -cyclodextrin likewise covalently bound to silica gel [73], and studies of cyclodextrins used as chiral mobile phase additives in reverse-phase HPLC [74] and in capillary electrophoresis [75].

We now consider a completely synthetic receptor based on the design of Still [76] and used as a chiral stationary phase by Gasparrini [77]. Figure 10 illustrates an example of a C_3 symmetric macrocycle derived from L-tyrosine. For the purposes of modeling, the tyrosyl side chains have been replaced with methyl groups and the spacer chain connecting the macrocycle to the silica gel has been omitted. The analyte studied is N-methoxycarbonyl-D,L-Ala- \dagger butyl ester, hereafter called D,L-analyte.

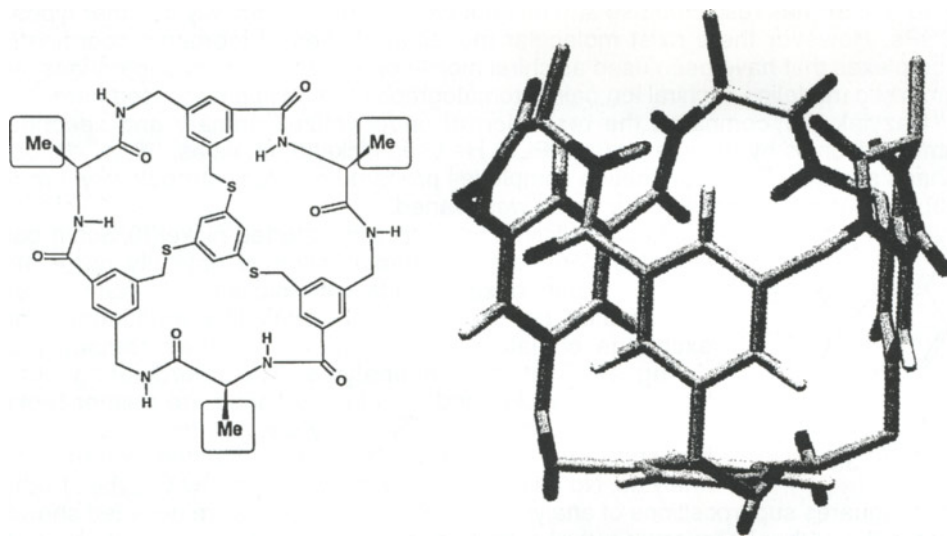


Figure 10. Left: Schematic of C_3 symmetric L-tyrosyl macrocyclic CSP with tyrosyl side chains replaced by methyl groups. Right: Lowest energy conformer of host molecule illustrating the depth of its chiral cavity.

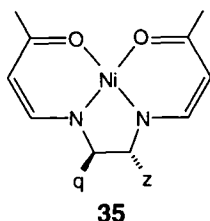
The computational goals were to understand the origins of enantioselection in these synthetic receptors and the authors used Glob-Moline described earlier in this chapter. First they carried out an exhaustive conformational analysis of guest and host molecules using several conformer search strategies in MacroModel. The united AMBER force field was used and the conformers were energy minimized with a continuum solvation model (GB/SA CH_3Cl) turned on. The lowest energy conformer of host was found to be populated in excess of 99%. The D,L-analyte was likewise

evaluated and six unique conformers located. Although the first two of these accounted for nearly 98% of the population, the authors considered the binding of all six conformers for D and L-analyte isomers binding to the C_3 symmetric host. Following the Glob-Moline flow chart, the quasi-flexible fitting of guest with host resulted in a free energy difference of -1.08 kcal/mol favoring the L enantiomer. This corresponds to $\alpha = 6.19$ comparing favorably with the experimental value of 6.30. Because true free energies are being computed with Gasparrini's protocol, both $\Delta\Delta H$ and $\Delta\Delta S$ contributions to $\Delta\Delta G$ are computed. These values were compared with experimental values derived from vant Hoff plots using equations 6 or 10 and good agreement is found. From this the authors are able to extract information concerning both the mode of binding and enantioselection with much confidence. Extensions to other systems by these authors is in progress [78].

9. TYPE IV CSPS

These kinds of stationary phases involve solute association with a metal complex as in chiral ligand exchange chromatography (CLEC) [79]. There are no reported atomistic computations in this area that we are aware of, most likely because this type of CSP has restricted use and has not been studied as heavily as other types of CSPs. However there exist molecular modeling studies of inorganic coordination complexes that have been used as chiral mobile phase additives for separations, and atomistic modeling in chiral ion pair chromatography that will be discussed here.

Bazylak [80] compared the resolution of underivatized primary and secondary amino alcohols by reverse phase HPLC. He used nickel(II) chelates, 35, as a mobile phase additive. The coordination complexes prepared are depicted below where the substituents at stereocenters z and q were varied.



These additives are helically distorted nickel(II) Schiff base chelates derived from condensation of optically pure tetradentate Schiff base ligands with nickel(II) acetate. These coordination complexes do not work like traditional ligand exchange chelates because their structures remain intact during association with analytes. The coordinating bonds between the nickel and Schiff-base ligand are neither broken nor formed, and the coordinatively unsaturated nickel (II) only has specific steric and electrostatic interactions with the two analytes. No energy calculations were done by Bazylak. Rather,

least-squares superpositions of analyte with chelating reagent were depicted showing where the author believes negatively charged amine groups associate with the metal and how other interactions like π -stacking with the chelate rings induce chiral discrimination.

Another report on ion-pair chromatography, where computed molecular descriptors from molecular modeling were nicely correlated with experimental separation factors, was published by Karlsson, Luthman, Pettersson and Hackzell [81]. They examined factors responsible for separation of aminotetralins on achiral stationary phases in the presence of the chiral additive N-benzyloxycarbonylglycyl-L-proline (L-ZGP), a protected peptide derivative.

Using the MMX force field these authors first determine the distribution of conformational states accessible to the analytes and then evaluated the preferred conformations of L-ZGP in the neutral and ionic forms. Then the authors brought these components together to form the various diastereomeric complexes. Their

strategy was to implement a flexible docking scheme because it was felt that both molecules of the complex may change their conformations during association.

The authors found typically 40 conformations within 3 kcal/mol of the global minimum for their analytes and 243 low energy conformations for L-ZPG, requiring ca. 40 x 243 docking combinations, each having a large number of orientations and positions with respect to one another. Thus a computer-aided docking protocol was used in addition to manual docking. Following the methodologies of Lipkowitz, Rogers and Gasparrini described in an earlier section, the authors decided not to use only the lowest energy structures, but rather, all their structures to derive averaged energies and averaged properties for comparison to experiment. The best molecular descriptor they found correlating theory with experiment is the averaged nonpolar unsaturated surface area of the complex. Another paper extended their studies [82] where similar conclusions were derived.

These computational studies are comparable to those described in the section covering Type I CSPs. Experimentally the only difference between these separations and those above is that here the selectors are not stationary phases but rather are co-additives that form the diastereomeric complexes. Because no computational studies on type IV CSPs exist, molecular modeling of inorganic coordination complexes directed towards rationalizing enantioselective binding and chiral recognition presents itself as a ripe area for exploration.

10. TYPE V CSPS

Type V CSPs are protein phases. Because of the well established chemo- and stereospecificity of enzymes, a large number of experimentalists have adapted proteins in one form or another as stationary phases for chiral separations. The intermolecular forces responsible for analyte binding to these biopolymers are the same as for most other CSPs but the size and complexity of proteins makes them difficult to study computationally. One would think that with approximately 400 entries in the Brookhaven Protein Databank to select from, separation scientists would have used one of these proteins as a chiral selector and then use those atomic coordinates to carry out molecular modeling studies. Only one example has appeared in the literature where information from the PDB has been used to serve as a beginning point for molecular modeling of a protein CSP. In all other examples the CSP is viewed as having an unknown structure and Quantitative Structure-Enantioselective Retention Relationships (QSERRs) have been carried out.

A multidisciplinary study by Pinkerton et al. [83] addressed the chiral discrimination of stationary phases made from intact and fragmented turkey ovomucoid (a commercially available CSP exists). Avian ovomucoids generally contain three tandem, homologous domains and they can be subjected to controlled proteolysis. Upon isolation, purification and cleavage of the protein, the authors isolated several of these domains that were then covalently attached to silica gel and used to examine enantioselectivity for a large number of test racemates. Turkey ovomucoid was selected for this study because previous NMR assignments of its third domain had been published. The authors generated chiral columns containing the following materials: whole turkey ovomucoid (OMTKY), a combination of first and second domains (OMTKY[1+2]), the second domain (OMTKY2), the glycosylated third domain (OMTKY3S) and the nonglycosylated third domain (OMTKY3). As expected, a rich and diverse set of chromatographic results were obtained, especially with the

third domain (OMTKY3) which displayed the best chiral recognition for benzodiazepines and profens.

Extensive NMR titration studies together with NOESY results provided analyte association constants and information about the binding sites on the third domain. Their molecular modeling began by extracting the atomic coordinates of silver pheasant ovomucoid third domain from the PDB and converting Met 18 to Leu, thus generating OMTKY3. Two analytes were considered: pranoprofen and a related heterocycle designated as U-80,413. These analytes were geometry optimized and then bound to the protein using an in-house program called Autodock. This docking program allowed each enantiomer to search the surface of the OMTKY3 for low energy binding orientations while keeping both the selector and selectand rigid. It uses a modified docking algorithm developed by Kuntz [84]. First an extended radius dot surface is generated over the protein. Dots are removed leaving those that allow the analyte to interact with more than one protein atom. The retained dots are then used for rigid overlaying of analyte on the CSP. In essence the retained dots are potential target sites for the analyte. Each docking is scored by summing the van der Waals and electrostatic interactions of the selector with the selectand.

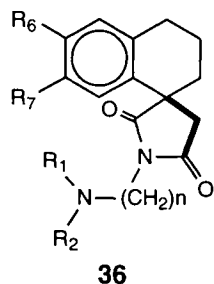
Typically about 6500 orientations are generated (for each enantiomer) and the best 200 are energy minimized, allowing the analyte to fully relax but keeping the protein rigid. Only 100 of the best structures (lowest energy) are kept as a working set. The authors located potential binding sites on the protein this way and found their results to be consistent with independent NMR results. Then, using the NMR results, they selected the six best docking orientations for each enantiomer in each set.

From their computations they found 1) the R enantiomer is bound closer to the protein than is S; 2) the R analytes typically have lower binding energies; 3) two major binding domains exist, one on each side of the protein, such that analytes can not bind simultaneously to both sites. Of these two sites one was found to be comprised of mainly hydrophobic amino acids. Computed binding energies at this region are generally of higher energy than at the other site. Moreover, nonspecific binding orientations lead to lack of chiral discrimination in this site. For the lower energy binding domain the authors enumerated the similarities and the differences of enantiomer binding. All of their computed results were consonant with experimental facts. While graphical representations of R vs S analyte binding were presented, and some details about key interactions between each enantiomer with specific amino acids on the protein were described, the authors indicated that chiral recognition mechanisms could not yet be described without ambiguity. In spite of this, the work described here is the only example of atomistic modeling for type V CSPs where selector-selectand interactions are explicitly determined. It is also an experimental tour de force illustrating the synergy of experiment and theory.

Because so few type V CSPs have well established molecular structures, most scientists are forced to use a series of probe molecules and some sort of regression analysis to divulge pertinent information about the mechanism of chiral recognition. For example, Norinder and Hermansson [85] separated thirty-five N-aminoalkylsuccinimides, 36, on an α_1 -acid glycoprotein (AGP) column.

To explore the relationship between molecular structure and enantioselectivity a principal component analysis with partial least squares projection techniques allowed the authors to determine which of 50 physicochemical descriptors correlated with separation factors. A partial list of variables used in their models is given in Table 4.

Similar descriptors along with indicator variables for the other R groups on 36 provide the other variables used.

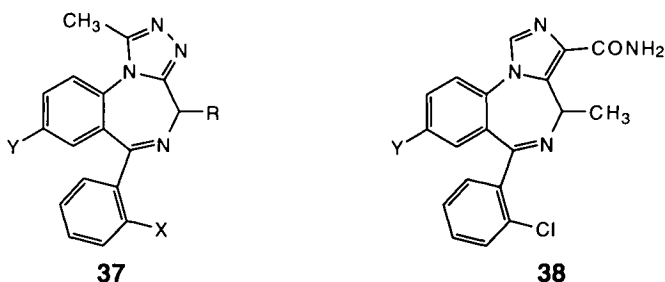


Four significant principal components describing 85% of the variance were found. The most important variables were associated with positions 6 and 7 on the aryl ring. Lipophilic groups containing aromatic character are especially important for enantioselectivity and the length of the aliphatic side chain was also found to be important.

Another QSERR was carried out by Kaliszan, Noctor and Wainer [86] who suggest many of the abstract and arbitrary indicator variables used by Norinder and Hermansson obscures the physical interpretation of the correlations. These authors measured retentions of 21 chiral and achiral 1,4-benzodiazepine derivatives, like 37 and 38 on an immobilized human serum albumin CSP. The variables used to determine the QSERR consisted of molecular descriptors derived from computational chemistry.

Table 4
Partial list of variables used in the Norinder-Hermansson models

No.	Variable	Position	Explanation
1	n		Chain length, number of CH ₂ groups
2	n^2		Chain length, number of CH ₂ groups
3	MR	R ₁	Molecular refractivity
4	MR^2		Molecular refractivity
5	L	R ₁	Verloop Sterimol parameter
6	L^2	R ₁	Verloop Sterimol parameter
7	B_1	R ₁	Verloop Sterimol parameter
8	B^2	R ₁	Verloop Sterimol parameter
9	B_5	R ₁	Verloop Sterimol parameter
10	B^2	R ₁	Verloop Sterimol parameter
11	f	R ₁	Rekkers aliphatic fragmental constant
12	f^2	R ₁	Rekkers aliphatic fragmental constant
13	MR	R ₂	Molecular refractivity
14	MR^2	R ₂	Molecular refractivity
15	L	R ₂	Verloop Sterimol parameter
16	L^2	R ₂	Verloop Sterimol parameter
17	B_1	R ₂	Verloop Sterimol parameter
18	B^2	R ₂	Verloop Sterimol parameter
19	B_5	R ₂	Verloop Sterimol parameter
20	B_5	R ₂	Verloop Sterimol parameter
21	f	R ₂	Rekkers aliphatic fragmental constant
22	f^2	R ₂	Rekkers aliphatic fragmental constant



In this study $\log k'_1$ and $\log k'_2$ were considered as two sets of mutually independent variables. The resulting models were:

$$\log k'_1 = -1.75 + 0.39 \log f_Y - 1.84 C_3 - 0.16 W + 0.04 \beta_{CCN} + 0.17 f_X \quad (26)$$

$$\log k'_2 = 1.99 + 0.89 P_{SM} + 0.48 f_Y - 4.15 C_3 - 0.12 W + 0.13 f_X \quad (27)$$

Here f_Y is the hydrophobicity of substituents on position 7; C_3 is the atomic charge on carbon 3 derived by semiempirical molecular orbital calculations; W is the width of the analyte; β_{CCN} is the diazepin C_2 - C_3 - N_4 angle; f_X is the hydrophobicity of substituents at carbon 2'; and P_{SM} is the substructure dipole which is the charge difference between the hydrogen at C_3 and the most negatively charged atom multiplied by the distance (D in Figure 11) between them. Equations 26 and 27 allow

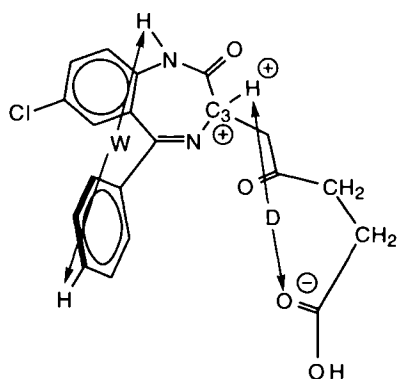


Figure 11. Several of the molecular descriptors used in a QSERR study of 1,4-benzodiazepine analogs on a human serum albumin-based HPLC chiral stationary phase.

one to gain insight into the retention mechanism. Equation 26 indicates that the binding site responsible for the first eluted isomer contains structural and spatial constraints and that the hydrophobicity of group Y at C_7 is most important for anchoring the analyte to the CSP. Counteracting this is the excess charge at C_3 and the width of the binding site which appears to be restricted. Equation 27 indicates that the second eluted isomer is influenced by the same structural features as the first eluted enantiomer, i.e. hydrophobicity, f_Y , excess charge at C_3 and width. However, the most significant retention descriptor for $\log k'_2$ is the local dipole, P_{SM} . This indicates that the greatest difference between binding sites for the enantiomers involves the charge sites of the cationic area on the exterior of the CSP. The significant conclusion derived from this computational approach is that there appear to be two distinct binding sites on the protein CSP for chiral benzodiazepines.

Summary

In this chapter we examined how atomistic molecular modeling is used to address questions concerning enantiodiscrimination in chiral chromatography. For Type I CSPs it is revealed that a variety of strategies are commonly used for sampling microstates accessible to the transient, diastereomeric complexes. One extreme is to rely primarily on chemical intuition and/or knowledge obtained from experiment. These strategies are referred to as "motif-based" search strategies and they can be effective when used judiciously. Moreover they have the benefit of reducing CPU time that can become problematic for large and flexible CSPs. The other extreme is to let the computer do all the sampling without user intervention, and, a variety of stochastic and deterministic searching techniques have been successfully employed. Examples of all these strategies were presented in this chapter for the sake of comparison.

In contrast to Type I stationary phases where molecular modelers explicitly treat the intermolecular interactions between selector and selectand, one finds more use of regression models for Type II-V CSPs. The reason for this is that the shape of these CSPs is, with the exception of cyclodextrin and several synthetic hosts, not well defined or not known at all. Thus all one can do is rely on regression models to divulge information concerning the mechanism of retention and enantioselection for a series of related analytes. These models, albeit lacking a detailed atom-by-atom account of the interactions taking place as analytes percolate through a chromatographic column, nonetheless provide important information concerning where and how chiral recognition takes place. Moreover, these models are capable of making predictions. That is, once the model has been constructed and validated, one can use those same kinds of molecular descriptors to predict what the separation will be for an as yet unknown analyte.

The computational tools needed for simulating analyte separation under a variety of chromatographic conditions with various stationary phases, chiral and achiral, gas or liquid, currently exist. However we point out that while these computational tools are powerful when used properly, it is still advantageous to use one's own experience when selecting a CSP for a chiral separation. In this regard, then, we point out the enormous research effort by Roussel [87] and Koppenhoefer [88] who created and maintain CHIRBASE, a graphical molecular database on the separation of enantiomers by gas, liquid and supercritical fluid chromatographies. A more recent and potentially very useful database is CHIRULE, a column selection system, designed by Stauffer and Dessy [89]. Databases like these together with the computational methodologies described above allow one to make a better selection of the chromatographic tools needed for a resolution and provide insights concerning the mechanism of chiral discrimination.

Finally, most of the published computational studies directed toward chiral chromatography have been carried out by chromatographers rather than by computational chemists. Most of these scientists look at computational chemistry as an adjunct to their experimental work, but understand the information content derived from molecular simulations can provide valuable information not otherwise available. In that sense they are right. However, most chromatographers are not well versed in computational chemistry and make too many serious errors for their results to be of benefit. So, on the one hand there is a need for computational chemistry but on the other hand too many pitfalls exist for the non-expert to step into. The conclusion one draws from this is that chromatographers should work collaboratively with

computational chemists to help them solve their problems. In this regard, then, the future of molecular modeling in the separation sciences looks bright.

Acknowledgments

Some of the work described herein was carried out under the auspices of grants from the National Science Foundation.

REFERENCES

1. J.D. Bolcer and R.B. Hermann, in *Reviews in Computational Chemistry*, Vol. 5, K.B. Lipkowitz and D.B. Boyd (eds.), VCH Publishers, New York, NY, 1994, Chapter 1, pp. 1-60.
2. Approximately 50% of the full papers in the *Journal of the American Chemical Society* used computational chemistry in 1995. *Reviews in Computational Chemistry*, Vol. 8, K.B. Lipkowitz and D.B. Boyd (eds.), VCH Publishers, New York, NY, 1996, preface.
3. *Computer-Aided Molecular Design. Applications in Agrochemicals, Materials and Pharmaceuticals*, C.H. Reynolds, M.K. Holloway and H. Cox (eds.), ACS Symposium Series 589, American Chemical Society, Washington, DC, 1995.
4. D.B. Boyd, in *Reviews in Computational Chemistry*, Vol. 1, K.B. Lipkowitz and D.B. Boyd (eds.), VCH Publishers, New York, NY, 1990, Chapter 10, pp. 355-371.
5. R.W. Souter, *Chromatographic Separations of Stereoisomers*, CRC Press, Boca Raton, FL, 1985.
6. *Chromatographic Chiral Separations*, M. Zeif and L. Crane (eds.), (Chromatographic Science Series, Vol. 40), Marcel Dekker, New York, NY, 1987.
7. W.A. König, *The Practice of Enantiomer Separation by Capillary Gas Chromatography*, Hüthig, Heidelberg, 1987.
8. S.G. Allenmark, *Chromatographic Enantioseparation. Methods and Application*, Ellis Horwood, Chichester, 1988.
9. *Chiral Separations*, D. Stevenson and I.D. Wilson (eds.), Plenum Press, New York, NY, 1988.
10. W.J. Lough (ed.), *Chiral Liquid Chromatography*, Blackie, London, 1989.
11. *Recent Advances in Chiral Separations*, D. Stevenson and I.D. Wilson (eds.), Plenum Press, New York, NY, 1990.
12. *Chiral Separations by Liquid Chromatography*, S. Ahuja (ed.), ACS Symposium Series, 471, American Chemical Society, Washington, DC, 1991.
13. *Chiral Separations by Liquid Chromatography*, G. Subramanian (ed.) VCH, Weinheim, 1994.
14. D. Casarini, L. Lunazzi, F. Pasquali, F. Gasparri and C. Villani, *J. Am. Chem. Soc.*, 114 (1992) 6521.
15. W.H. DeCamp, *Chirality*, 1 (1989) 2.
16. I.W. Wainer, *Trends in Analytical Chemistry*, 6 (1987) 125.
17. K.B. Lipkowitz, *J. Chem. Educ.*, 72 (1996) 1070.
18. R. Däppen, H.R. Karfunkel and F.J.J. Leusen, *J. Comput. Chem.*, 11 (1990) 181.
19. M.G. Still and L.B. Rogers, *Talanta* 36 (1989) 35.
20. K.B. Lipkowitz, D.A. Demeter, R. Zegarra, R. Larter and T. Darden, *J. Am. Chem. Soc.*, 110 (1988) 3446.

21. K.B. Lipkowitz, B. Baker and R. Zegarra, *J. Comput. Chem.*, 10 (1989) 718.
22. K.B. Lipkowitz and B. Baker, *Anal. Chem.*, 62 (1990) 770.
23. J.M. Blaney, P.K. Weiner, A. Dearing, P.A. Kollman, E.C. Jorgensen, S.J. Oatley, J.M. Burrige and C.F.C. Blake, *J. Am. Chem. Soc.*, 104 (1982) 6424 and earlier work.
24. M.G. Still and L.B. Rogers, *J. Comput. Chem.*, 11 (1990) 242.
25. M.G. Still and L.B. Rogers, *Talanta*, 37 (1990) 599.
26. K.B. Lipkowitz, S. Antell and B. Baker, *J. Org. Chem.*, 54 (1990) 5449.
27. K.B. Lipkowitz, B. Baker and R. Larter, *J. Am. Chem. Soc.*, 111 (1989), 7750.
28. J. Aerts, *J. Comput. Chem.*, 16 (1995) 914.
29. S. Topiol, M. Sabio, J. Moroz, and W. B. Caldwell, *J. Am. Chem. Soc.*, 110 (1988) 8367.
30. S. Topiol and M. Sabio, *J. Chromatogr.*, 461 (1989) 129.
31. M. Sabio and S. Topiol, *Int. J. Quantum Chem.*, 36 (1989) 313.
32. M. Sabio and S. Topiol, *Chirality*, 3 (1991) 56.
33. K.B. Lipkowitz, *J. Chromatogr. A*, 666 (1994) 493.
34. F. Gasparrini, L. Lunazzi, D. Misiti and C. Villani, *Acc. Chem. Res.*, 28 (1995) 163.
35. F. Gasparrini, M. Pierini, S. Alcaro, S. Mecucci and O. Incani, personal communication.
36. A. M. Edge, D. M. Heaton, K. D. Bartle, A.A. Clifford, and P. Myers, *Chromatographia*, 41 (1995) 161.
37. C. Hansch and A. Leo, *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*, ACS Professional Reference Book, Washington, DC, 1995.
38. R. Kaliszan, *Quantitative Structure-Chromatographic Retention Relationships*, John Wiley and Sons, New York, NY, 1987.
39. C. Altomare, A. Carotti, S. Cellamare, F. Fanelli, F. Gasparrini, C. Villani, P.-A. Carrupt and B. Testa, *Chirality*, 5 (1993) 527.
40. H. Kubinyi, *QSAR: Hansch Analysis and Related Approaches in Methods and Principles in Medicinal Chemistry*, Vol. 1, R. Mannhold, P. Krogsgaard-Larsen and H. Timmerman (eds.), VCH Publishers, Weinheim, 1993, Chapter 9.3.
41. A Carotti, C. Altomare, S. Cellamare, A. Monforte, G. Bettoni, F. Loiodice, N. Tangari and V. Tortorella, *J. Computer-Aided Mol. Design*, 9 (1995) 131.
42. S. Weinstein, B. Feibush and E. Gil-Av, *J. Chromatogr.* 126 (1975) 97.
43. S. Weinstein and L. Leiserowitz, *Acta Crystallogr.*, Sect. B, 36 (1980) 1406.
44. S. Weinstein, L. Leiserowitz and E. Gil-Av, *J. Am. Chem. Soc.*, 102 (1980) 2768.
45. P. Camilleri, J.A. Murphy, M.R. Saunders and C.J. Thorpe, *J. Computer-Aided Mol. Design*, 5 (1991) 277.
46. I. Alkorta, J. Elguero, P. Goya and C. Roussel, *Chromatographia*, 27 (1989) 77.
47. E. Francotte and R.M. Wolf, *Chirality*, 3 (1991) 43.
48. E. Yashima and Y. Okamoto, *Bull. Soc. Chem. Jpn.*, 68 (1995) 3289.
49. E. Yashima, M. Yamada, Y. Kaida and Y. Okamoto, *J. Chromatogr. A*, 694 (1995) 347.
50. E. Yashima, C. Yamamoto and Y. Okamoto, *J. Am. Chem. Soc.*, 118 (1996) 4036.
51. R. Isaksson, H. Wennerstrom and O. Wennerstrom, *Tetrahedron*, 44 (1988) 1697.

52. R.M. Wolf, E. Francotte and D. Lohmann, *J. Chem. Soc., Perkin Trans. 2*, (1988) 893.
53. C. Roussel, J.-L. Stein, M. Sergent and R. Phan Tan Luu, in D. Stevenson and I.D. Wilson (eds.), *Recent Advances in Chiral Separations*, Plenum Press, New York, NY, 1990, p. 105.
54. C. Roussel and C. Popescu, *Chirality*, 6 (1994) 251.
55. C. Roussel, S. Lehuédé, C. Popescu and J.-L. Stein, *Chirality*, 5 (1993) 207.
56. J.G. Ning, *J. Chromatogr. A*, 659 (1994) 299.
57. R.D. Armstrong, T.J. Ward, N. Pattabiraman, C. Benz and D.W. Armstrong, *J. Chromatogr.*, 414 (1987) 192.
58. R.D. Armstrong, in W.L. Hinze and D.W. Armstrong (eds.), *Ordered Media in Chemical Separations ACS Symposium Series, 342*, American Chemical Society, Washington, DC, 1987, Ch. 16.
59. D.W. Armstrong, T.J. Ward, R.D. Armstrong and T.E. Beesley, *Science*, 232 (1986) 1132.
60. A. Berthod, S.-C. Chang and D.W. Armstrong, *Anal. Chem.*, 64 (1992) 395.
61. C. Roussel and A. Favrou, *Chirality*, 5 (1993) 471.
62. P. Camilleri, A.J. Edwards, H.S. Rzepa and S.M. Green, *J. Chem. Soc., Chem. Commun.*, (1992) 1122.
63. K.B. Lipkowitz, S. Raghothama and J.-A. Yang, *J. Am Chem. Soc.*, 114 (1992) 1554.
64. D.W. Armstrong, X. Yang, S.M. Han and R.A. Menges, *Anal. Chem.*, 59 (1987) 2594.
65. A. Berthod, W. Li and D.W. Armstrong, *Anal. Chem.*, 64 (1992) 873.
66. K.B. Lipkowitz, *J. Org. Chem.*, 56 (1991) 6357.
67. J.E.H. Köhler, M. Hohla, M. Richters and W.A. König, *Angew Chem.*, 104 (1992) 362; *Angew. Chem., Int. Ed. Engl.*, 31 (1992) 319.
68. J.E.H. Köhler, M. Hohla, M. Richters and W.A. König, *Chem. Ber.*, 127 (1994) 119.
69. N. Koen de Vries, B. Coussens and R.J. Meier, *J. High Resolut. Chromatogr.*, 15 (1992) 499.
70. F. Kobor, K. Angermund and G. Schomburg, *J. High Resolut. Chromatogr.*, 16 (1993) 299.
71. D. R. Black, C. G. Parker, S. S. Zimmerman and M. L. Lee, *J. Comput. Chem.*, 17 (1996) 931.
72. P. Camilleri, C.A. Reid and D.T. Manallack, *Chromatographia*, 38 (1994) 771.
73. Y. Kuroda, Y. Suzuki, J. He, T. Kawabata, A. Shibukawa, H. Wada, H. Fujima, Y. Go-oh, E. Imai and T. Nakagawa, *J. Chem. Soc. Perkin Trans. 2*, (1995) 1749.
74. D.G. Durham and H. Liang, *Chirality*, 6 (1994) 239.
75. C.L. Copper, J.B. Davis, R.O. Cole and M.J. Sepaniak, *Electrophoresis*, 15 (1994) 785.
76. S.D. Erikson, J. Simon and W.C. Still, *J. Org. Chem.*, 58 (1993) 1305.
77. F. Gasparrini, D. Misiti, C. Villani, A. Borchardt, M.T. Burger and W.C. Still, *J. Org. Chem.*, 60 (1995) 4314.
78. F. Gasparrini, C. Villani, M. Pierini, S. Alcaro, S. Mecucci and D. Misiti, presented at the 8th International Symposium on Chiral Discrimination, Edinburgh, Scotland, July 2, 1996.
79. V.A. Davankov, J.D. Navratil, and H.F. Walton, *Ligand Exchange Chromatography*, CRC Press, Boca Raton, FL, 1988.

80. G. Bazylak, *J. Chromatogr. A*, 665 (1994) 75 and 668 (1994) 519.
81. A. Karlsson, K. Luthman, C. Pettersson and U. Hacksell, *Acta Chem. Scand*, 47 (1993) 469.
82. K. Luthman, A.V. Jensen, U. Hacksell, A. Karlsson and C. Pettersson, *J. Chromatogr. A*, 666 (1994) 527.
83. T.C. Pinkerton, W.J. Howe, E.L. Ulrich, J.P. Comisky, J. Haginaka, T. Murashima, W.F. Walkenhorst, W.M. Westler and J.L. Markley, *Anal. Chem.*, 67 (1995) 2354.
84. R.L. DesJarlais, R.P. Sheridan, J.S. Dixon, I.D. Kunz and R. Venkataraghavan, *J. Med. Chem.*, 29 (1986) 2149.
85. U. Norinder and J. Hermansson, *Chirality*, 3 (1991) 422.
86. R. Kaliszan, T.A.G. Noctor and I.W. Wainer, *Chromatographia*, 33 (1992) 546.
87. C. Roussel and P. Piras, *Pure Appl. Chem.*, 65 (1993) 235.
88. B. Koppenhoefer, A. Nothdurft, J. Pierrot-Sanders, P. Piras, C. Popescu, C. Roussel, M. Stiebler and U. Trettin, *Chirality*, 5 (1993) 213.
89. S.T. Stauffer and R.E. Dessy, *J. Chromatogr. Sci.*, 32 (1994) 228.

This Page Intentionally Left Blank

Theoretical Investigation of Carbon Nets and Molecules

Alexandru T. Balaban

Department of Organic Chemistry, Polytechnic University, Bucharest, Roumania

1. INTRODUCTION

We shall discuss a few systems which are formed exclusively or partially from carbon atoms, stressing contributions brought to this field by the Roumanian research group during the last 30 years. In 1968 [1], alternative carbon allotropes different from graphite and diamond were examined for the first time ; a review entitled "Carbon and Its Nets" was published in 1989 [2]. Two joint papers were published with Professor Roald Hoffmann from the Cornell University. During the last few years after 1990, when travel abroad was no longer restricted, investigations were continued in cooperation with Professors Douglas J. Klein, Thomas G. Schmalz and William A. Seitz from the Texas A & M University in Galveston.

As a criterion for a systematic discussion, we shall adopt as first classification the dichotomy infinite/finite systems, i. e. macromolecule vs. molecule ; as a second criterion, we shall discuss systems in terms of hybridization, using the well-known types sp , sp^2 , sp^3 although these are only first approximations for dicoordinated, tricoordinated and tetracoordinated atoms, because the exact hybridization is determined by the valence angles.

2. INFINITE PLANAR NETS OF sp^2 -HYBRIDIZED CARBON ATOMS

2.1. Graphite : two-dimensional infinite sheets

The best known net, with the lowest ground-state energy under normal conditions of pressure and temperature, is graphite. Its aromaticity (i. e. electronic delocalization) is associated with high electrical and thermal conductivity within the graphene plane (single crystals have conductivities about 200 times higher within the molecular planes than across them). The strong anisotropy, due to covalent bonding within the honeycomb lattice and to Van der Waals forces in the orthogonal direction, leads to linear compressibilities $10^4 - 10^5$ times larger in the latter direction. The opacity and black color of graphite are due to the large aromatic chromophore.

Interatomic C—C distances within the graphene sheet are 141.5 pm, whereas molecular planes are 335 pm apart, leading to a low density ($2,270 \text{ kg}\cdot\text{m}^{-3}$ in the ideal case, practically less). The facile gliding of graphene sheets atop one another explains why polycrystalline graphite is soft ; its use for lubrication or for writing in pencils is thus understandable (interestingly, black colors used for prehistoric painting in caves or in ancient pottery are also due to graphite in soot). The word graphite is derived from the Greek verb to write ($\gamma\rho\alpha\phi\epsilon\iota\nu$).

The first nuclear reactors were moderated with pure graphite because the neutron capture cross-sections, $\sigma(n, \gamma)$, of ^{12}C (3.4 millibarn, abbreviated as mb) and ^{13}C (0.9 mb) are much lower than for protons (332 mb). Nevertheless, the large mass of these nuclei relative to that of the proton leads to long distances for "cooling" the fission neutrons produced by ^{235}U in its interaction with a thermal neutron in the nuclear chain reaction. Nowadays water-moderated or heavy-water-moderated nuclear reactors have displaced in most cases the graphite-moderated reactors. Water contains also ^2H (0.52 mb), ^{16}O (0.18 mb), ^{17}O in low amount (240 mb) and ^{18}O (0.18 mb), therefore H_2O -moderated reactors must function with ^{235}U -enriched uranium as fuel ; however, D_2O -moderated reactors may function with natural uranium. Other unfavorable features of graphite-moderated nuclear reactors are their positive void coefficient, allowing a positive feed-back when the chain reaction escapes control (remember the Chernobyl catastrophe), and the Wigner effect (sudden release of energy, i. e. heating, produced by strain due to neutron-caused accumulation of defects in the lattice).

The major use for graphite nowadays is for carbon fibers (to reinforce composites) produced either from pitch or from polymers such as polyacrylonitrile. Graphitization of these materials is performed stepwise by oxidation at 250° for losing much of the hydrogen, followed by carbonization in the absence of oxygen at 800° , and then by graphitization at 1400° - 2500° .

According to the stacking of graphene sheets, two forms of graphite are known, namely hexagonal (ABAB... type) and rhombohedral (ABCABC... type).

Since graphite is the thermodynamically favored allotropic form of elemental carbon (by 0.2 kcal/mol relative to diamond), one can gradually convert diamond into graphite on heating at 1000° at normal pressure in the absence of air (cf. also item 4.2 below).

2.2. Other planar infinite lattices with sp^2 -hybridized carbon

Local defects in the molecular plane of graphite may involve fracture zones with 4- and 8-membered rings, as seen in Fig. 1, causing angle strain and lowering the aromaticity ; they may become annealed on heating (see above the Wigner effect). Other defects were enumerated by Dias [3].

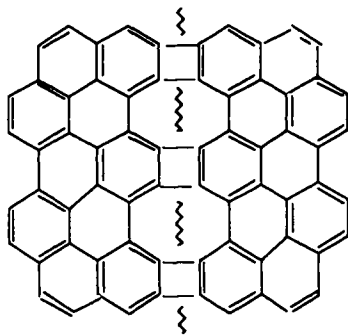


Fig. 1. A fracture zone in the graphite lattice leading to the formation of 4- and 8-membered rings.

The first systematic enumeration of semiregular lattices with sp^2 -hybridized carbon was published in 1968 [1]. We may denote the graphite lattice by $\{6^3\}$ indicating that three regular hexagons meet at each vertex ; in this allotrope there is no angle strain because the 120° angle corresponds exactly to the valence angle of $C(sp^2)$, and there is no loss of aromaticity because there are no 4- or 8-membered rings (there exist 12-membered rings but, according to the conjugated circuits model [4-6], the corresponding negative contribution Q_3 for such rings is quite small, cf. Table 1).

Table 1. Coefficients (in eV) for various models of conjugated circuits : R_k for $(4k+2)$ -membered rings, and Q_k for $4k$ -membered rings

Model	R_1	R_2	R_3	R_4	Q_1	Q_2	Q_3	Q_4
Randic [4]	0.869	0.246	0.100	0.041	-1.600	-0.450	-0.150	-0.060
Trinajstic [5]	0.869	0.247	0.100	-	-0.781	-0.222	-0.090	-
Herndon [6]	0.841	0.336	-	-	-0.650	-0.260	-	-

However, in the $\{4, 8^2\}$ semiregular planar lattice (Fig. 2) both the angle strain and the antiaromaticity contribute to destabilization ; in the conjugated circuits model there are 10- and 14-membered circuits with small positive contributions R_2, R_3 as well as 4-, 8- and 12-membered rings with appreciable negative coefficients Q_1, Q_2, Q_3 .

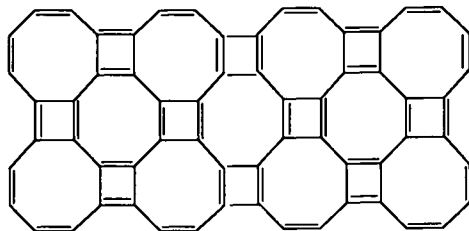


Fig. 2. The $\{4, 8^2\}$ semiregular planar lattice of carbon atoms.

Following the procedure of Barriol and Metzger [7], a first MO calculation [1] carried out in 1968 yielded for the $\{4, 8^2\}$ lattice a relative energy of 11.6 kcal/mole above that of graphite. A more recent (1994) and elaborate calculation [8] using the conjugated circuits method afforded for graphite a resonance energy per electron (RE/e) of 0.168, and for the $\{4, 8^2\}$ lattice -0.0990 eV.

Two other semiregular nets examined in 1968 [1] and 1994 [8] were $\{3, 9^2\}$ and $\{4, 6, 12\}$ with resonance energies per electron -0.007 and -0.0744 eV, respectively. Other 3-connected planar nets, but without regular polygons, were analyzed. From

their corresponding RE/e values, it was seen that the only ones with positive values are those with equal numbers of 5- and 7-membered rings, but these values are appreciably lower than for graphite. However, taking into account the high inertia for rearrangements of carbon nets (i. e. high activation barriers, as seen by the stability of diamond), such metastable systems may be able of existence, at least as local defect structures in graphite.

A different approach, also using using the same computational method [8], investigated infinite lattices with regularly alternating strips of $\{6^3\}$ and $\{5, 7^2\}$ assemblies. Finally, similar calculations with islands of $\{5, 7^2\}$ assemblies within the graphene sheets yielded RE/e values intermediate between those of graphite and those of $\{5, 7^2\}$ nets. As a conclusion, the higher the ratio of hexagons relative to 5- and 7-membered rings, the higher the resonance energy.

Burdett et al. [9,10] reported calculations by means of the moments method on the nets $\{3, 9^2\}$, $\{4, 8^2\}$ and $\{5^2, 7^2\}$. Hoffmann, Eisenstein and Balaban [11] discussed a hypothetical strain-free oligoradical with 8-membered rings whose carbon atoms have sp^2 -hybridization.

Both graphite and diamond in practice are not infinite, but finite nets and therefore have dangling bonds at the peripheries of single crystals ; normally, these bonds have hydrogen atoms attached to them. Consequently one may consider graphite to be an "honorary polycyclic peri-condensed benzenoid hydrocarbon", and diamond to be an "honorary adamantanoid or diamondoid hydrocarbon"

2.3. Tridimensional infinite lattices with sp^2 -hybridized carbon atoms

Hoffmann and coworkers [12] described a hypothetical carbon net (a three-dimensional metallic allotrope) with infinite polyene chains running along two mutually orthogonal dimensions, and with no conjugation along a third dimension. Although there is no angle strain, the π -clouds of neighboring polyenes, being too close together, lead to " π -strain". The calculated density is $2,970 \text{ kg}\cdot\text{m}^{-3}$, intermediate between that of diamond ($3,510 \text{ kg}\cdot\text{m}^{-3}$) and graphite ($2,270 \text{ kg}\cdot\text{m}^{-3}$). The smallest rings have 10 and 12 carbons, and the space group is $I4_1/amd$ with eight carbons in the unit cell. The band structure, computed by the extended Hückel theory, and the calculated density of states, were reported.

Cohen and coworkers [13] used the pseudopotential localized-orbital approach and found that this allotropic carbon structure is a low-compressibility metal ; because of the nearly perfect lattice match with the diamond (100) surface, they proposed that it may be possible to grow this structure epitaxially on the diamond (100) surface.

2.4. Graphitic cones with sp^2 -hybridized carbon atoms

On dividing a planar graphene sheet into six sectors by three infinite straight lines meeting at the center of a hexagon, each sector has at the center a 60° angle (Fig. 3). On removing between one and five of these sectors and by folding the remainder of the sheet into a cone, one can obtain five types of graphitic cones [14]. Actually, their number is larger because, as seen from Fig. 4, the connection between the marginal dangling bonds may be formed in a variety of skew manners, leading to different types of apex structures.

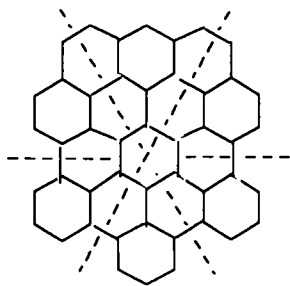


Fig. 3. Division of a graphene sheet into six sectors

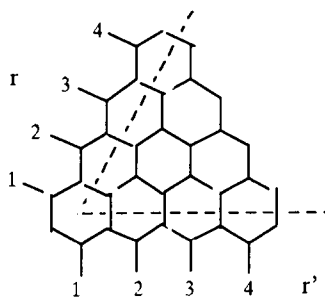


Fig. 4. Ranks r and r' of dangling bonds for the 60° sector ($n = 1$)

In order to avoid connections leading to free radicalic structures, allowed connections impose restrictions on the values of the number n of 60° sectors, and the ranks r and r' of dangling bonds: when $r + r'$ is an odd number, $n = 2$ is allowed; when $r + r'$ is even, any n value is allowed.

Interestingly, two years after these ideas on graphitic cones were presented at the 7th International Symposium on Novel Aromatic Compounds (Victoria, BC, Canada, July 18-24, 1992), and a few months after the apparition of the corresponding paper [14], Ge and Satler from the University of Hawaii [15] reported the experimental observation of cones with the smallest solid angle, formed by folding a single sector with a planar angle of 60° as in Fig. 4.

3. INFINITE NETS OF sp^3 -HYBRIDIZED CARBON ATOMS

3.1. Diamond: three-dimensional infinite network

The normal (cubic) diamond lattice, similar to sphalerite, has no eclipsed bonds, i. e. all 6-membered rings are in chair form. The C—C bond distance is 154 pm as in alkanes. One may observe in this lattice approximately planar arrangements of condensed 6-membered rings (their equatorial plane), about 140 pm apart from each other, i. e. at much smaller distances than for the graphene sheets; this arrangement accounts for the high density of diamond relatively to other forms of carbon (ideally $3,510 \text{ kg}\cdot\text{m}^{-3}$). The fact that all carbon atoms are connected via strong covalent σ -bonds accounts for the fact that diamond, like graphite, does not melt on heating at normal pressure in the absence of oxygen, and is completely insoluble. Its high refractive index, transparency, hardness and scarcity in nature make diamond a coveted gem with a high stock value which is carefully watched and adjusted by South African, Dutch, American, and Russian cartels. The lack of electrical conductivity for pure (non-doped) diamond is paradoxically associated with a high thermal conductivity; diamond films are deposited epitaxially for electronic circuits to help in dissipating heat and to confer mechanical protection. As with silicon and germanium at present, but less easily, diamond may be doped with 3rd and 5th Main Group elements to become a highly valued electronic material in the future.

A second type of diamond (hexagonal lattice or isodiamond) having a crystal structure similar to wurtzite is less common and slightly more energy-rich owing to its eclipsed bonds (boat-shaped 6-membered rings) as in iceane.

The synthesis of artificial diamond from graphite requires elevated pressure (about 54 kbar), high temperature (1500°) and catalysts (transition metals) as reviewed earlier [2, 16-20] ; cf. also item 4.1 below.

3.2. Other systems with sp^3 -hybridized carbon atoms

In 1968 it was shown [1] that the tridimensional space may be filled with semiregular polyhedra called truncated octahedra. The angle strain, i. e. the difference from the normal tetrahedral angle of 109.5° , would account for at least 12 kcal/ mole, because of the planar 4- and 6-membered rings. However, the spiranic strings of 4-membered rings in perpendicular directions (Fig. 5, a stereo-view like the following figures) introduce an angle strain of a special type, due to the fact that two opposite valencies at 90° angles meet at one and the same carbon atom.

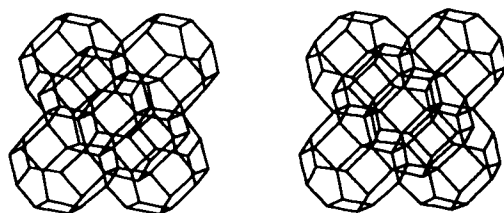


Fig. 5. Stereo-view of the tridimensional carbon net formed by truncated octahedra.

A hypothetical planar square lattice (an infinite fenestrane sheet) would have an enormous strain because each carbon atom would be forced to have two opposite valence angles reduced from 109.5° to 90° , and then these two pairs of bonds would have to be twisted to become coplanar.

3.3. Holes bordered by heteroatoms within the diamond lattice

If a carbon atom in the diamond lattice is deleted and if the four carbon atoms to which it had been connected are replaced by nitrogen heteroatoms, a tetrahedral "hole" results [21]. In principle, a small metal cation or a proton may become trapped in the hole (Fig. 6). When a similar process is performed for a pair of adjacent carbon atoms, the result is a quasi-octahedral hole bordered by six nitrogen heteroatoms ; it too may include a metal cation ; in Fig. 7 a portion of the diamond lattice with several such holes is presented.

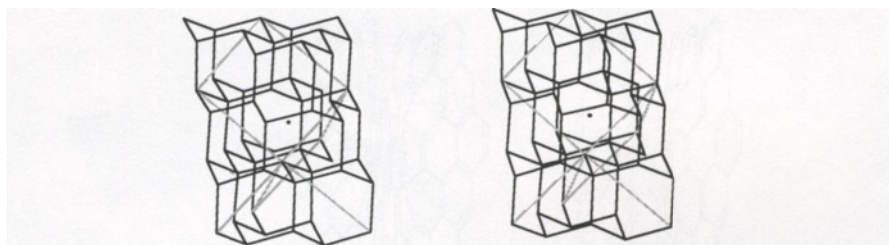


Fig. 6. Stereo-view of a tetrahedral hole in diamond with a metal cation inside.

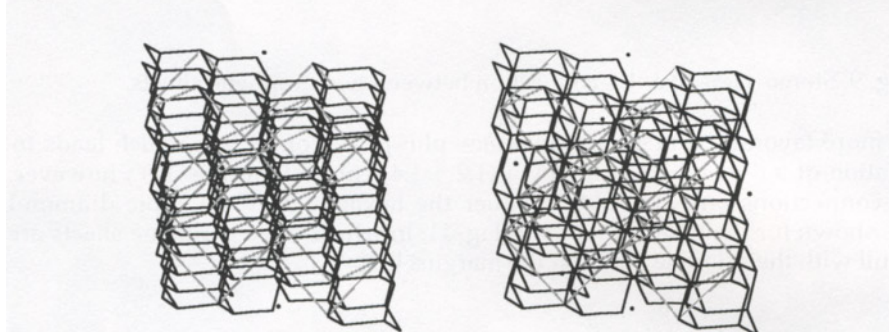


Fig. 7. Stereo-view of quasi-octahedral holes in diamond with metal cations inside.

4. INFINITE NETS WITH BOTH sp^2 - and sp^3 -HYBRIDIZED CARBON ATOMS

4.1. Local defects in the graphite lattice

For converting graphite into diamond, one has to connect carbon atoms belonging to different graphene sheets. Theoretical data were reported by Hoffmann et al. [22], and practical data are summarized by Bundy [18] and De Vries [19]. If two connections involve *ortho*-situated carbon atoms (we call them 1,2-connections) one would obtain 4-membered rings (Fig. 8) ; if they involve *para*-situated carbon atoms (1,4-connections) one would obtain 8-membered rings (Fig. 9) [23].

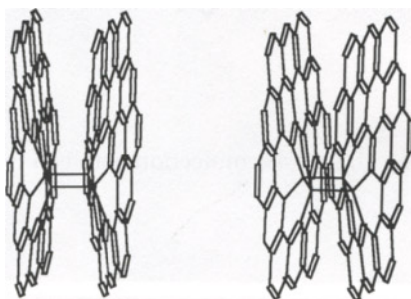


Fig. 8. Stereo-view of a 1,2-connection between two graphene sheets.

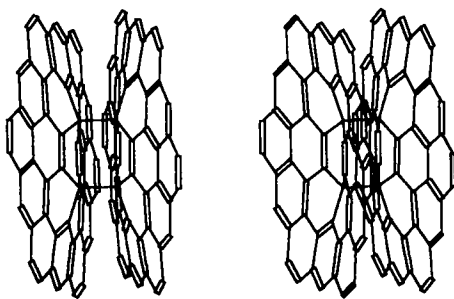


Fig. 9. Stereo-view of a 1,4-connection between two graphene sheets.

A more favorable situation is an *ortho-* plus *para-* connection which leads to the formation of a 6-membered ring by a (1,2 + 1,4)-connection (Fig. 10) ; however, with 1,3-connections one may obtain either the hexagonal or the cubic diamond lattice, as shown for the latter situation in Fig. 11, in which three graphene sheets are shown, still with their double bonds at the margins [23].

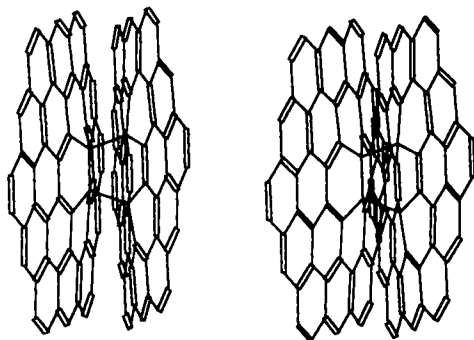


Fig. 10. Stereo-view of a (1,2 + 1,4)-connection between two graphene sheets.

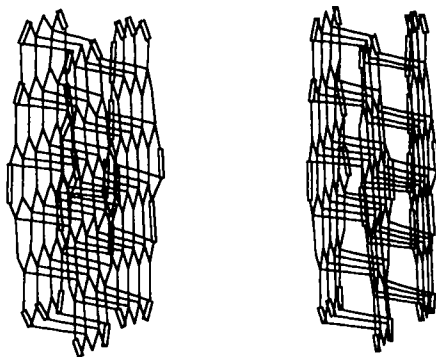


Fig. 11. Stereo-view of many 1,3-connections between three graphene sheets.

4.2. Local defects in the diamond lattice

The graphitization of diamond may involve thermally-allowed six π -electron concerted processes : an electrocyclic reaction converts a six-membered ring in the diamond lattice into three isolated double bonds as shown in Fig. 12. Subsequently, each of the three cyclohexene rings can undergo a retro-Diels-Alder reaction, multiplying the number of double bonds in a triple cascade reaction [23].

Alternatively, one can consider in the diamond lattice three layers of condensed hexagons in chair conformation ; by converting two pairs of neighbor carbon atoms in the central layer into sp^2 -hybridized atoms connected by double bonds one obtains two three-membered rings in each of the layers flanking the central layer. Two possibilities arise, denoted by 1,2,3,4-bond breaking (Fig. 13) and 1,2,4,5-bond breaking. The former process is slightly favored energetically. One can assume that the steric strain in the 3-membered rings leads to changes in hybridization and to propagation of the graphitization in a "domino" process which involves gradual changes passing from 6-membered rings in diamond to cyclopropane-containing systems and finally to olefinic/ aromatic delocalization [23].

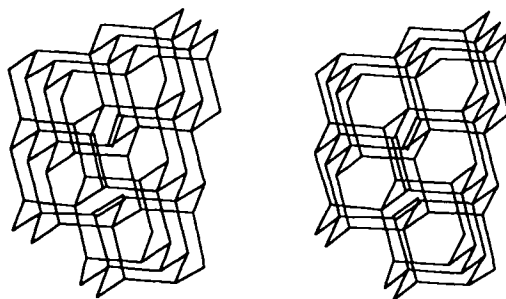


Fig. 12. Diamond net with three double bonds resulted by electrocyclic ring opening.

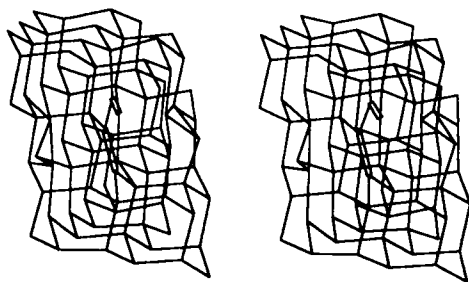


Fig. 13. Stereo-view of 1,2,3,4-bond breaking in diamond.

A few theoretical and experimental references relevant to the present discussion on transitions between diamond and graphite (or the B,N-analogs) and vice-versa follow [24-26].

4.3. Block-copolymers of graphite and diamond (diamond-graphite hybrids)

Since the inter-plane distance between graphene sheets is 335 pm, one may look for at the diamond lattice for such an interatomic separation, and connect correspondingly graphite and diamond fragments. For cubic diamond, there are two arrangements, one in which the graphene sheets are parallel in all blocks, and another one in which they are mutually orthogonal. For hexagonal diamond, the arrangement in which the graphene sheets are parallel in all blocks leads to matching. Of course, in all cases the magnitude of blocks (fragments) can be varied, leading to various types of block-copolymers that have regular or irregular fragment arrangements [27].

4.4. Systems with regularly alternating sp^2/sp^3 -hybridized carbon atoms

Merz, Hoffmann and Balaban [28] described a large variety of infinite three-dimensional carbon nets in which $C(sp^2)$ and $C(sp^3)$ atoms alternate in regular arrangements. Their band energies and density of states have been calculated.

Elguero et al. [29] described another possible carbon form which incorporates in a similar fashion sp^2 - and sp -hybridized carbon atoms ; the unit cell contains 12 carbon atoms, 8 of which are sp^2 and 4 are sp ; it has a calculated density of $2,720 \text{ kg}\cdot\text{m}^{-3}$ and may have metallic properties.

Karfunkel and Dressler [30] described an interesting type of infinite net based on the following idea : triptycyl moieties are connected by benzene rings in two dimensions so that 2D hexagonal nets (with regular or irregular hexagons) emerge, and then one connects identical nets by single bonds ; the resulting 3D net possess a translational symmetry in one dimension. For these allotropes, the energy obtained by modified neglect of diatomic overlap solid-state calculations is comparable to that of diamond.

O'Keefe et al. [31] described "polybenzene" which may be viewed as a 3D truncated octahedron lattice in which edges have been replaced by benzene rings. Only one kind of carbon atom forms the entire net. The six- and eight-membered

rings occur in the ratio 2:3. The unit cell has 24 atoms. One of these polybenzenes has a substantially lower energy per carbon atom than buckminsterfullerene.

Strelnitskii et al. [32] reported a superdense carbon allotrope ($4,100 \text{ kg}\cdot\text{m}^{-3}$) obtained as carbon films formed by radio-frequency condensation of carbon plasmas on cooled substrates; the crystalline phase, obtained along with amorphous phase, was studied by electron diffraction and revealed a primitive rhombohedral unit cell with 8 carbon atoms, hence this phase was called C_8 , and its structure, as proposed by Stankevich et al. [33] and Biswas et al. [34], involved cubes connected by single bonds (supercubane). Burdett and Lee [9] found the supercubane structure to be less stable than diamond if the constituting atoms have 4 or less electrons per atom, but more stable for electron-rich systems (i. e. >4 electrons per atom). Johnston and Hoffmann [35], observing discrepancies in the crystallographic analysis and the unusual bond length distribution, found that a likely alternative structure for C_8 is the body-centered BC-8 structure adopted by the high-pressure γ -Si allotrope.

5. INFINITE CHAINS OF sp - HYBRIDIZED CARBON ATOMS

5.1. Chains of sp -hybridized carbon atoms : one-dimensional system

The reports of Sladkov [36] on a carbon allotrope (carbyne) which would add to the two known carbon hybridization states a third one, namely sp -hybridization in macromolecular chains, were followed by data collected by Whittaker et al. [37] who preferred the name chaoite.

5.2. Heteroatom substitution inside polyacetylenic chains

If carbon atoms inside polyacetylenic chains are substituted by nitrogen atoms, positively-charged nitrilium or isonitrilium centers result; boron substitution leads to negatively-charged centers. No experimental or theoretical data are yet available on such systems, which (if charges would be correctly placed) would benefit from electrostatic interactions, in addition to Van der Waals forces.

6. MOLECULES WITH sp^2 -HYBRIDIZED CARBON ATOMS

6.1. Fullerenes

The 1996 Nobel Prize for Chemistry was awarded to Harold W. Kroto, Richard E. Smalley and Robert F. Curl for having discovered in 1985, through the corresponding intense mass spectrometric peaks, the molecules C_{60} (buckminsterfullerene) and C_{70} and having postulated cage-like structures formed from 12 pentagons and any number of hexagons different from 1 for such molecules [38]. They went on to argue that these two molecules had no abutting pentagons on the basis that for these two polyhedra only one structure existed fulfilling this condition [39]. Five years later, a team of physicists (Krätschmer, Huffman and Fostiropoulos [40]) described a method for obtaining fullerenes in macroscopic amounts and for purifying them by extraction and chromatography. Spectroscopic characterization (especially vibrational and ^{13}C -NMR spectroscopy) fully confirmed the structures proposed earlier by Kroto, Smalley and Curl [38], and adumbrated by several theoretical predictions connected with reviews on carbon allotropes [18-20]. The literature on fullerenes increases exponentially. A bibliography was published by

Braun et al. [41], several books [42-50] and reviews on this topics have appeared [39, 51-54].

Proper fullerenes consist of 12 five-membered rings, and any number of six-membered rings different from 1. Euler's theorem for polyhedra also allows other sizes of faces, and even the presence of vertex degrees different from 3 [55], but such systems should not be considered to be proper fullerenes. Herndon and coworkers [56] calculated that some carbon cages with 4-membered rings might be as stable as proper fullerenes.

Proper fullerenes with more than 26 carbon atoms have more than one isomer, but so far only isolated-pentagon fullerenes, which are the most stable among their isomers, have been obtained experimentally; isolated-pentagon fullerenes with more than 64 carbon atoms also have more than one isomer. Among the 1812 isomers of C_{60} there are seven unique ones for any given pair of p and q values [57]; p is the number of carbon atoms common to two pentagons ($0 < p < 20$), and q is the number of carbon atoms common to three pentagons ($0 < q < 10$). An equivalent parametrization can be made [58] in terms of A and B , which denote, respectively, the numbers of carbon atoms common to three pentagons ($A = q$) and the number of carbon atoms common to two pentagons and one hexagon ($B = 2p - 3q$; $0 < B < 24$). Buckminsterfullerene C_{60} (Fig. 14) and C_{70} (Fig. 15) are the only isomers that are isolated-pentagon fullerenes with $p = q = A = B = 0$; one of the unique isomers of C_{60} is the least stable one ("sausage" with $p = 20$, $q = A = 10$ and $B = 24$) which has two half-dodecahedrane caps separated by a belt of 20 hexagons, shown in stereo-view in ref. [59]; another unique C_{60} isomer is the "pillow" with two coronene moieties surrounded by six pairs of abutting pentagons, having $p = 12$, $q = A = 0$ and $B = 24$ (Fig. 16).

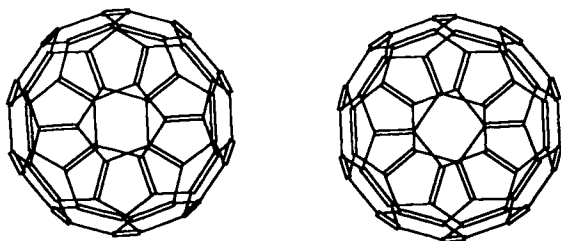


Fig. 14. Stereo-view of buckminsterfullerene C_{60} .

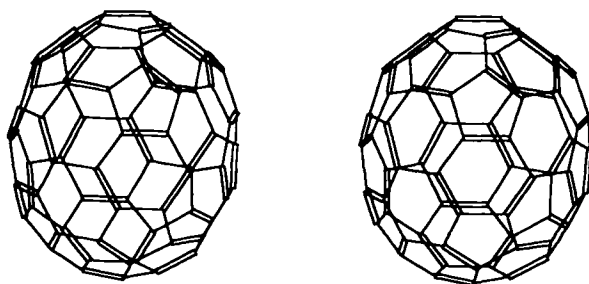


Fig. 15. Stereo-view of the C_{70} isomer with $p = q = 0$ which was obtained experimentally.

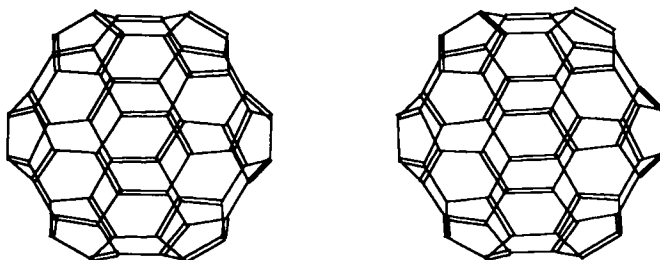


Fig. 16. Stereo-view of the unique C_{60} isomer with $p = 12$ and $q = 0$ ("pillow" isomer).

Metal atoms or ions can be trapped endohedrally ; the tris-potassium or tris-rubidium derivative is the organic superconductor with the highest critical temperature obtained so far.

One can interconvert fullerene isomers (at least theoretically) via the Stone-Wales rearrangement [60] which automerizes pyracylene into itself, as shown in the first row of Fig. 17. A generalization of this rearrangement was discussed [58], and the underlying idea is presented in the remaining rows of Fig. 17. From the 1812 C_{60} isomers, 31 do not possess an arrangement of pentagons that would allow a Stone-Wales rearrangement to be performed [61] ; by means of the above generalization (rows 2 and 3 in Fig. 17), all of these isomers are able to be converted into other isomers that can then undergo the normal Stone-Wales rearrangement.

The fullerene nomenclature adopted till now by IUPAC [62] is based on the "spiral code" [63]. However, there are fullerenes for which this code is not applicable. Other nomenclature proposals have been published [64]. The Baeyer nomenclature of buckminsterfullerene based on the "best Hamiltonian circuit" [65,66] is so complicated that it was reported erroneously several times in the literature (review in [68]). Several topological, quantum-chemical or graph-theoretical invariants can discriminate each of the 1812 isomers of buckminsterfullerene, or each of the 558 isolated-pentagon fullerenes with 60 through 96 carbon atoms [59]. Among the topological invariants, the most interesting ones are based on the dual graph of the fullerene and on its reduced dual which considers only the 12 pentagons.

6.2. Nanotubes and capsules

Although theoretically nanotubes can be infinitely long, we prefer to discuss them together with their capped counterparts (which are molecules) called capsules, as they are obtained experimentally. History repeats itself, as nanotubes had shared with fullerenes the fate of being initially detected by physical methods, and only later characterized chemically : they were first obtained in microscopic amounts by Iijima [67-69], and later in macroscopic amounts by Ebbesen et al. [70-72]. Electron-microscopic investigations revealed the multi-shell (onion-type) structure of these nanotubes which consist of hexagons, as if a graphene sheet was rolled into a cylinder and the dangling bonds became connected. The connection may give rise to a helical pitch [73, 74]. Capped nanotubes can be opened at their ends and thinned through oxidation which attacks preferentially the five-membered rings. Metal atoms can enter the nanotubes, leading to high expectations for molecular electronic devices.

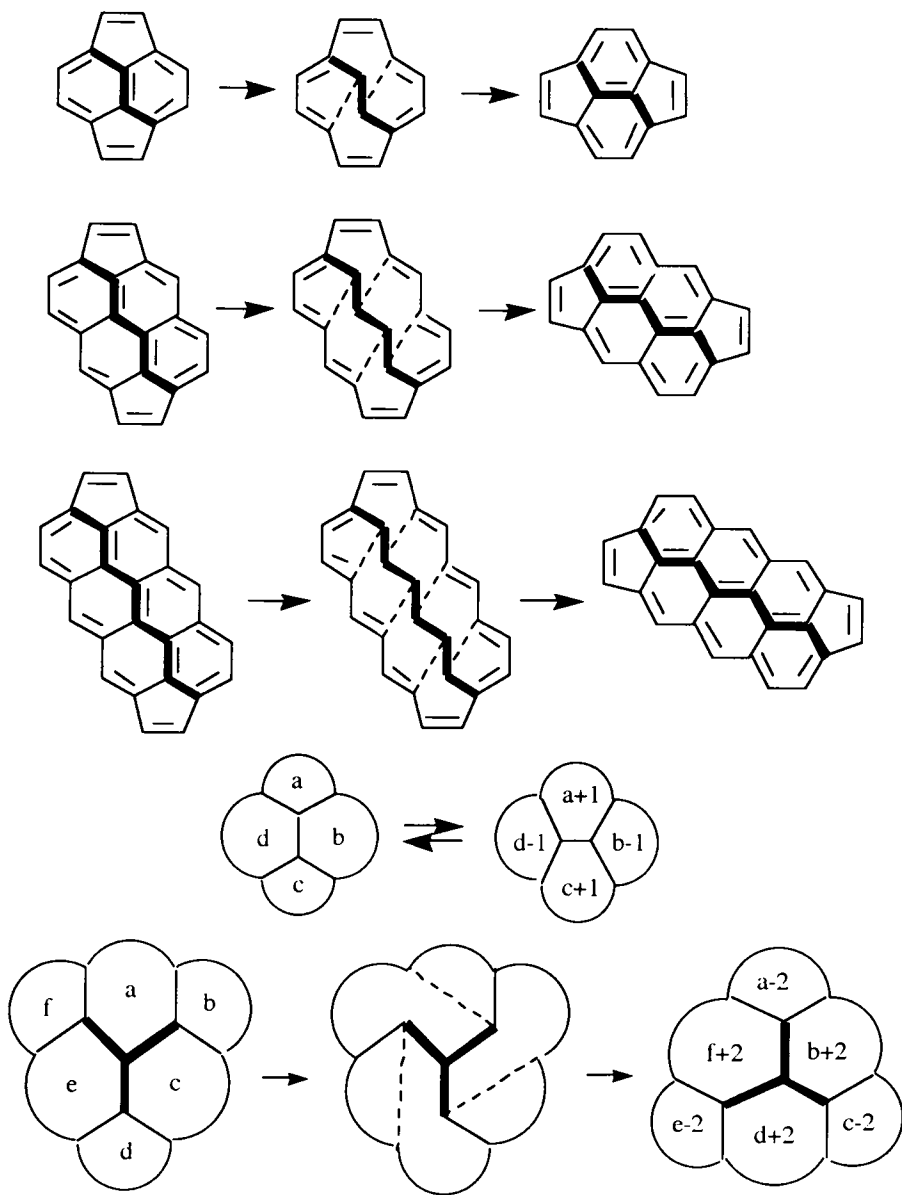


Fig. 17. Generalized Stone-Wales rearrangement : the first row indicates the pyracylene automerization, the next two rows the generalization for proper fullerenes, and the last two rows the generalization for any size of polygons of the cage.

In addition to these experimentally certified systems, one may conceive that the dangling bonds at the two ends of a carbon cylinder, instead of forming capsules by being plugged with fullerenic hemispheric caps, become reconnected forming a torus. Calculations show that such tori [75, 76] may have energies as low as finite graphitic sheets because tori have no dangling bonds. Interestingly, just as constitutional formulas are the same for stereoisomeric structures, the adjacency matrix of systems resulted by changing the order of the two foldings leading from a parallelogram of benzenoid rings in a graphene sheet to two tori does not distinguish between the two different topological stereoisomers of these two tori.

It is conceivable that a parallel bundle of linear polyynes chains (the carbon allotrope named chaoite or carbyne) may form, via thermally-allowed $(4 + 2)\text{-}\pi$ -electron cycloadditions, six-membered rings which lead to graphene sheets and/or nanotubes. However, the mechanism of nanotube formation in a plasma produced by arc discharge appears to require the presence of a transition metal atom at the rim of the growing nanotube [77].

6.3. Carbon cages and nanotubes including oxygen, nitrogen, or boron heteroatoms

If some or all of the carbon atoms in 5-membered rings of fullerenes are deleted, and if their adjacent carbon atoms (to which they had been attached, atoms that are left with a dangling bond) are replaced by oxygen heteroatoms, one obtains "fullerocoronands", one of which is shown as a stereo-view in Fig. 18. A detailed description of such systems has been published [78].

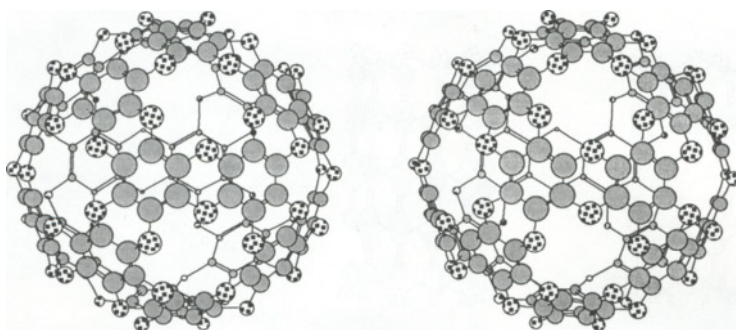


Fig. 18. Stereo-view of a fullerocoronand.

Capped carbon nanotubes (capsules) submitted to the same operation would lead to nanotubes having, at one or both their ends, hemisphere(s) with holes bordered by coronand systems as seen in Figures 19 and 20 [55].

The replacement of carbon atoms in fullerenes by boron and/or nitrogen atoms has been repeatedly discussed in the literature, both theoretically [79-81] and experimentally. Smalley's group at the Rice University in Houston has successfully obtained polyaza-fullerenes from a mixture of graphite and boron nitride under the same conditions in which graphite yields fullerenes [82]; by adding azides to buckminsterfullerene, Prato et al. [83] synthesized azafullerenes.

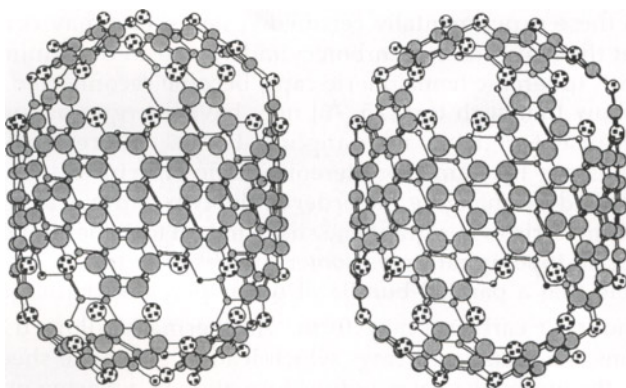


Fig. 19. Doubly capped nanotube with oxygen heteroatoms bordering holes.

Carbon nanotubes with open ends have dangling bonds at these ends ; this energetically unfavorable situation may be remedied on replacing the carbon atoms having dangling bonds by oxygen or nitrogen heteroatoms. For one extreme case of helical pitch in which the symmetry axis of the cylinder is parallel to C—C bonds in hexagons, such a replacement leads to the formation of pyranic and pyridinic rings, respectively, as seen in Figure 20 [84]. The resulting system has an electronic structure which is *para*-quinonoid, however.

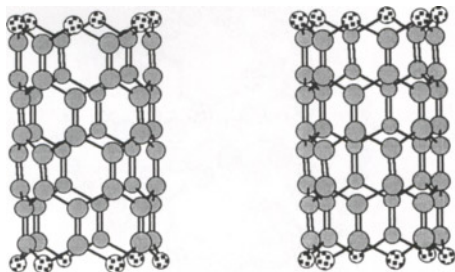


Fig. 20. Nanotube with heteroatoms in pyranic or pyridinic rings.

For the other extreme case of helical pitch in which the cylinder axis is perpendicular to C—C bonds in benzenoid rings, the same replacement leads [84] to furanic rings for oxygen heteroatoms (Figures 21 and 22), and to pyridazinic rings for nitrogen heteroatoms (Fig. 23).

Fullerocoronands of nanotubes whose ends have holes bordered by oxygen or nitrogen heteroatoms would be able to coordinate a metal cation, e. g. sodium, in the hole. If the anions are small enough to enter the cage or the nanotube, electric dipoles would result ; otherwise, with the coordinated cations attached to the face of the carbon system and the anions outside, the positively charged cage or nanotube should be retained on cationic ion exchange columns, e. g. on polystyrene-sulfonate resins.

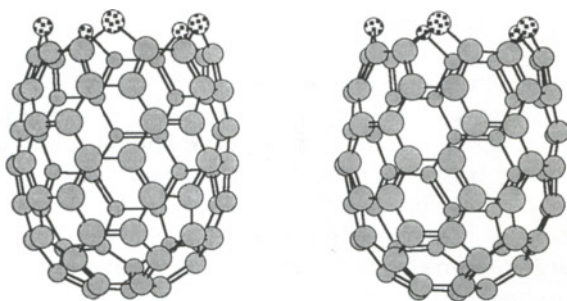


Fig. 21. Stereo-view of a nanotube with oxygens in furanic rings at one open end.

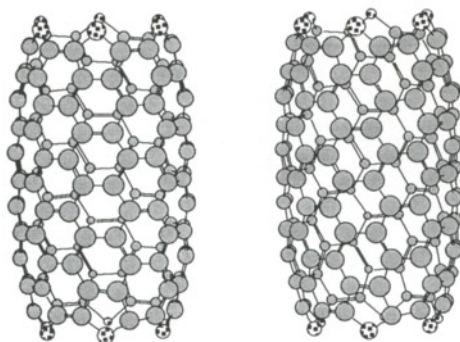


Fig. 22. Stereo-view of a nanotube with oxygens in furanic rings at both open ends.

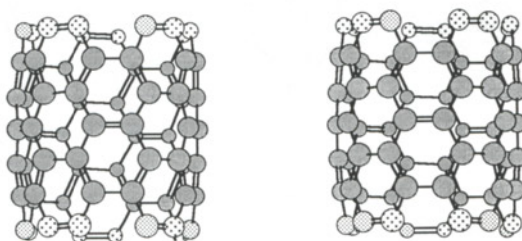


Fig. 23. Stereo-view of a nanotube with pyridazinic rings at both open ends.

7. MOLECULES WITH sp - AND sp^2 -HYBRIDIZED CARBON ATOMS

7.1. Cages with sp - and sp^2 -hybridized carbon atoms

Hypothetical carbon cages with acetylenic fragments replacing C—C bonds (belonging to 5-membered rings, 6-membered rings, or to both of these) in fullerenes were investigated theoretically by Baughman et al. [85]. Any C—C bond can thus be lengthened becoming C—C≡C—C. For C_{120} (Fig. 24) and C_{180} fullereneynes (Fig. 25), the binding energy (calculated using a tight-binding Hamiltonian which had been applied successfully to large carbon clusters) is about one eV per atom lower than for graphite, but these systems are expected to be stable.

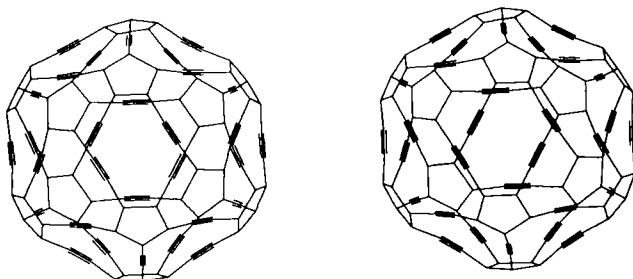


Fig. 24. Stereo-view of the C_{120} fullereneyne.

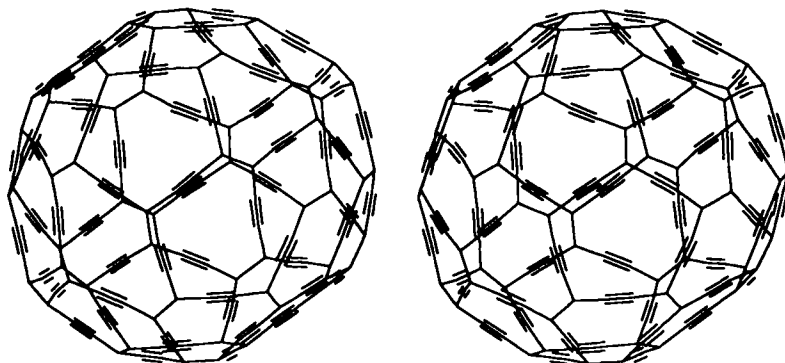


Fig. 25. Stereo-view of the C_{180} fullereneyne.

7.2. Molecules with sp -hybridized carbon atoms

The initial incentive for looking at the mass spectrum of carbon clusters, which brought about the serendipitous discovery of fullerenes, was Kroto and Walton's hypothesis [39] that the cosmic dark clouds might contain polyacetylenic aggregates. Indeed, up to about 30 carbon atoms, polyacetylenic rings seem to be prevalent [86, 87].

7.3. Covalently-bonded nested cages with sp - and/or sp^3 -hybridized carbon, or carbon and silicon atoms

Osawa et al. [88] imagined a nested buckminsterfullerenic cage with sp^3 -hybridized carbon atoms, connected by covalent bonds with an outer counterpart with sp^3 -hybridized silicon atoms ($C_{60}@Si_{60}$, an "inverse superatom"), and calculated the energy of such a strained arrangement (Fig. 26) in which the silicon atoms depart strongly from tetrahedral geometry, being pyramidalized.

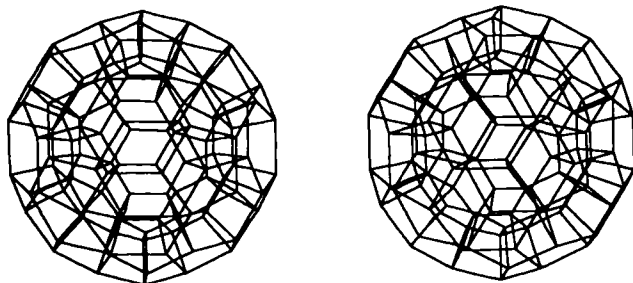


Fig. 26. Stereo-view of the "inverse superatom" $C_{60}@Si_{60}$ of Osawa et al.

A permuted system with a Si_{60} silicon cage *inside*, and various carbon cages having acetylenic fragments (fullerenynes) on the outside were investigated by Balaban [89]. Among combinations that fitted best such nested cages, minimizing steric strain and preserving icosahedral symmetry (atoms are no longer pyramidalized, but bent acetylenic fragments lead to strain), we present in Figures 27 and 28 two examples with two isomeric outer fullerenynes.

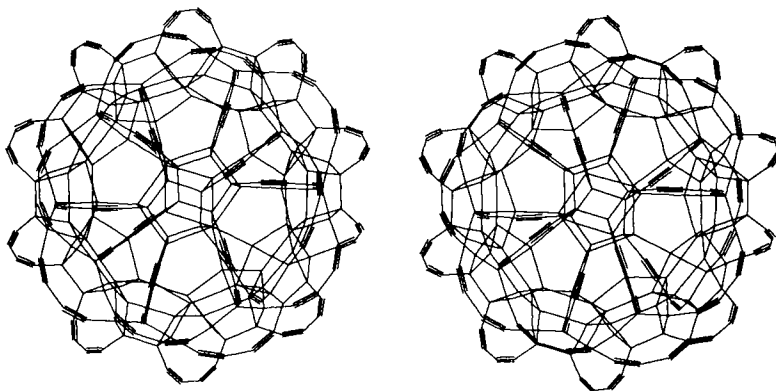


Fig. 27. Stereo-view of a reverse "inverse superatom" $Si_{60}@C_{180}$.

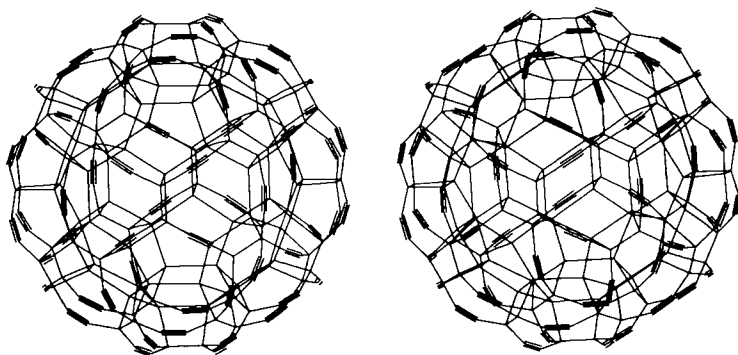


Fig. 28. Stereo-view of a reverse "inverse superatom" $\text{Si}_{60}@\text{C}_{180}$.

Finally, double carbon cages with C_{60} inside and various carbon cages having acetylenic fragments (fullerenynes) on the outside were explored by Balaban [90]. Again, it was necessary to explore combinations that fitted best such nested cages, minimizing steric strain and preserving icosahedral symmetry ; no pyramidalization is present, but there are again bent acetylenic fragments, as seen in the example displayed in Fig. 29.

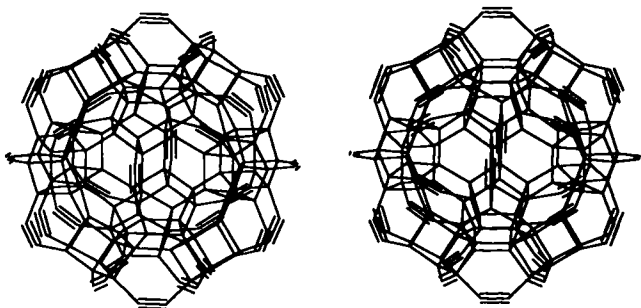


Fig. 29. Double carbon cage $\text{C}_{60}@\text{C}_{120}$ (C_{60} cage inside an outer C_{120} fullerene).

8. CONCLUSIONS : FROM RADIOASTRONOMY TO REMEDYING DANGLING BONDS IN CARBON NETS

Carbon is the most versatile element in the periodic system, and is the only one which is able to form the basis of life in the universe. Its elemental allotropes are also quite varied, both under the form of infinite nets and of molecules. In each of these, carbon atoms may be sp -, sp^2 -, or sp^3 -hybridized. Infinite nets have dangling bonds at the margins, normally satisfied by hydrogen atoms in a negligible ratio relative to carbon atoms. Two strategies are available for remedying the energetic cost of dangling bonds : (I) folding and reconnecting these bonds, or (II) replacing the

carbons possessing dangling bonds by heteroatoms. In both cases, the result is usually the conversion of an infinite net into a molecule. By folding and reconnecting, a portion of a planar graphene sheet leads to a molecule which is a cage, a nanotube, a capsule or a torus ; analogously, a linear chain leads to a ring. When heteroatoms replace carbon atoms at the borders of a piece of a 1-, 2-, or 3-dimensional carbon net, or at the borders of a hole in a 2- or 3-dimensional carbon net, stable molecules are the result.

On our planet, we live at the bottom of an ocean of air. The two "windows" of air for electromagnetic radiations are in the visible and radiofrequency regions. The former window has been used by living organisms since immemorial times, but the latter window has only recently allowed mankind to obtain information on our universe. The discovery by radioastronomic means of quasars, pulsars and (most relevant for the present discussion) polyacetylenic nitriles and dinitriles indicates that carbon-containing cages and chains are ubiquitous in the cosmos.

The cosmic abundance of nuclides having magic numbers of protons and neutrons (e. g. ^4He , ^{16}O , ^{40}Ca) is higher than for nuclides that have only one magic-number type of baryons (e. g. ^3He , ^{18}O) ; nuclides with even non-magic numbers of protons and/or neutrons, such as carbon, are less abundant, and in turn those with odd numbers of such baryons are least abundant.

Quantum chemistry explains the high stability of dinitrogen, N_2 , by the fact that it has a molecular closed shell of electrons (all bonding levels are filled, all antibonding levels empty, and no nonbonding levels). Molecules which are isoelectronic with N_2 , such as CO , NO^+ , CN^- , C_2^{2-} , and even O_2^{2+} , also possess enhanced stability and are formed at high temperature from any mixture containing the corresponding elements ; polyacetylenes and polyacetylenic dinitriles may owe their presence in intergalactic clouds to such factors. As a side-line, just as O_2^{2+} was predicted and then confirmed later [91] to be a bonded state (detectable by mass spectrometry of $^{16}\text{O}=^{17}\text{O}$), the ethane dication $\text{C}_2\text{H}_6^{2+}$ was investigated theoretically, but it was found that the lowest energy does not correspond to the structure which is isoelectronic with diborane ; instead, a carbenium-carbonium structure $[\text{H}_2\text{C}-\text{CH}_4]^{2+}$ has the lowest energy [92].

The spectacular difference between the two main carbon allotropes, graphite and diamond, as well as the recent discovery of fullerenes and nanotubes, make research in this area to be one of the most active at present.

REFERENCES

1. A. T. Balaban, C. C. Rentea and E. Ciupitu, *Rev. Roum. Chim.* 13 (1968) 231.
2. A. T. Balaban, *Computers Math. Applic.* 17 (1989) 397, reprinted in *Symmetry 2 Unifying Human Understanding* (Ed. I. Hargittai), Pergamon Press, 1989, p. 397.
3. J. R. Dias, *Carbon* 22 (1984) 107.
4. M. Randić, *J. Am. Chem. Soc.* 99 (1977) 444 ; *Tetrahedron* 33 (1977) 1905.
5. N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, FL, 2nd ed., p. 210.
6. W. C. Herndon, *J. Am. Chem. Soc.* 95 (1973) 2404 ; *Isr. J. Chem.* 20 (1980) 270.
7. J. Barriol and J. Metzger, *J. chim. phys.* 47 (1950) 17 ; 57 (1960) 848.
8. H. Zhu, A. T. Balaban, D. J. Klein and T. P. Zivkovic, *J. Chem. Phys.* 101 (1994) 5281.

9. J. K. Burdett and S. Lee, *J. Am. Chem. Soc.* 107 (1985) 3063.
10. J. K. Burdett, *Structure & Bonding* 65 (1987) 29 ; *Acc. Chem. Res.* 21 (1988) 189.
11. R. Hoffmann, O. Eisenstein and A. T. Balaban, *Proc. Natl. Acad. Sci. U.S.A.* 77 (1980) 5588.
12. R. Hoffmann, T. Hughbanks, M. Kertesz and P. H. Bird, *J. Am. Chem. Soc.* 105 (1983) 4831.
13. M. L. Cohen, *Phys. Rev. B* 32 (1985) 7988 ; A. Y. Liu, M. L. Cohen, K. C. Hass and M. A. Tamor, *Phys. Rev. B* 43 (1991) 6742 ; A. Y. Liu and M. L. Cohen, *Phys. Rev. B* 45 (1992) 4579.
14. A. T. Balaban, D. J. Klein and X. Liu, *Carbon* 32 (1994) 357.
15. M. Ge and K. Satler, *Chem. Phys. Lett.* 220 (1994) 192.
16. F. P. Bundy, *Nature* 241 (1973) 6930 ; *J. Geophys. Res.* 85 (1980) 6930.
17. R. C. De Vries, *Annu. Rev. Material Sci.* 17 (1987) 161.
18. D. A. Bochvar and E. G. Halpern, *Dokl. Akad. Nauk SSSR* 209 (1973) 40 ; M. V. Nikerov, D. A. Bochvar and I. V. Stankevich, *Izv. Akad. Nauk SSSR, Ser. Khim.* (1981) 1177 ; *Zh. Strukt. Khim.* 23 (1982) 13 ; 23 (1982) 16 ; 23 (1982) 177.
19. R. A. Davidson, *Theor. Chim. Acta* 58 (1981) 193.
20. E. Osawa, *Kagaku (Kyoto)* 25 (1970) 85.
21. A. T. Balaban, D. J. Klein and W. A. Seitz, *Int. J. Quantum Chem.* 60 (1996) 1065.
22. M. Kertesz and R. Hoffmann, *J. Solid State Chem.* 54 (1984) 313.
23. A. T. Balaban and D. J. Klein, *Carbon* (in press).
24. P. S. De Carli and J. C. Jamieson, *J. Chem. Phys.* 31 (1959) 1675 ; *Science* 133 (1961) 1821.
25. M. I. Heggie, *Carbon* 30 (1990) 71 ; A. E. De Vita et al. *Nature* 379 (1996) 523.
26. A. T. Balaban, D. J. Klein and C. A. Folden, *Chem. Phys. Lett.* 217 (1994) 266.
27. I. V. Stankevich, M. V. Nikerov and D. A. Bochvar, *Russ. Chem. Rev.* 53 (1984) 640.
28. K. M. Merz Jr., R. Hoffmann and A. T. Balaban, *J. Am. Chem. Soc.* 109 (1987) 6742.
29. J. Elguero, C. Foces-Foces and A. Llamas-Saiz, *Bull. Soc. Chim. Belg.* 101 (1992) 795.
30. H. R. Karfunkel and T. Dressler, *J. Am. Chem. Soc.* 114 (1992) 2285.
31. M. O'Keefe, G. B. Adams and O. F. Sankey, *Phys. Rev. Lett.* 68 (1992) 2325.
32. V. E. Strel'nitskii, V. G. Padalka and S. I. Vakula, *Sov. Phys. Tekh. Phys.* 23 (1978) 222 ; A. S. Bakai and V. E. Strel'nitskii, *Sov. Phys. Tekh. Phys.* 26 (1981) 1425.
33. I. V. Stankevich, M. V. Nikerov and D. A. Bochvar, *Russ. Chem. Rev.* 53 (1984) 640.
34. R. Biswas, R. M. Martin, R. J. Needs and O. H. Nielsen, *Phys. Rev. B* 30 (1984) 3210 ; 35 (1987) 9559.
35. R. L. Johnston and R. Hoffmann, *J. Am. Chem. Soc.* 111 (1989) 810.
36. A. M. Sladkov and Yu. M. Mikulin, *Usp. Khim.* 51 (1982) 736.
37. A. G. Whittaker, *Science* 200 (1978) 763 ; A. G. Whittaker, E. J. Watts, R. S. Lewis and E. Anders, *Science* 209 (1980) 1512.
38. H. W. Kroto, J. R. Heath, J. C. O'Brien, R. F. Curl and R. E. Smalley, *Nature* 318 (1985) 162.
39. H. W. Kroto, *Pure Appl. Chem.* 62 (1990) 707 ; *Science* 242 (1988) 1139 ; *Angew. Chem. Int. Ed. Engl.* 31 (1992) 111 ; H. W. Kroto, W. Allaf and S. P. Balm, *Chem. Rev.* 91 (1991) 1213 ; see also D. J. Klein, T. G. Schmalz, G. E. Hite and W. A. Seitz, *J. Am. Chem. Soc.* 110 (1986) 1113.
40. W. Krätschmer, L. D. Lamb, K. Fostiropoulos and D. R. Huffman, *Nature* 347 (1990) 354.

41. T. Braun, A. Schubert, H. Maczelka and L. Vasvari, *Fullerene Research 1985-1993, A Computer-Generated Cross-Indexed Bibliography of the Journal Literature*, World Scientific, Singapore, 1995.
42. G. S. Hammond and G. S. Kuck (Eds.), *Fullerenes. Synthesis, Properties, and Chemistry of Large Carbon Clusters*, ACS Symposium Series No. 481, American Chemical Society, Washington, DC, 1992.
43. C. L. Renschler, J. J. Pouch and J. J. Cox (Eds.), *Novel Forms of Carbon*, Proc. 270 Symposium Materials Research Society, Pittsburgh, PA, 1992.
44. W. E. Billups and M. A. Ciufolini (Eds.), *Buckminsterfullerenes*, VCH Publishers, New York, 1993.
45. A. Hirsch, *The Chemistry of the Fullerenes*, Georg Thieme Verlag, Stuttgart, 1994.
46. H. Aldersey-Williams, *The Most Beautiful Molecule : the Discovery of the Buckyball*, Wiley, New York, 1995 ; J. Baggott, *Perfect Symmetry*, Oxford Univ. Press, 1994.
47. J. Cioslowski, *Electronic Structure Calculations on Fullerenes and Their Derivatives*, Oxford University Press, New York, 1995.
48. P. W. Stephens (Ed.) *Physics and Chemistry of Fullerenes*, World Scientific, Singapore ; C. Taliani and G. Ruani, *Fullerenes : Status and Perspectives*, World Scientific, Singapore.
49. P. W. Fowler and D. E. Manolopoulos, *An Atlas of Fullerenes*, Internat. Series of Monographs on Chemistry No. 30, Clarendon Press, Oxford, 1995.
50. F. Wudl, *Acc. Chem. Res.* 25 (1992) 157.
51. V. I. Sokolov and I. V. Stankevich, *Russ. Chem. Rev.* 62 (1993) 419.
52. D. J. Klein and T. G. Schmalz, in *Quasicrystals, Networks and Molecules with Fivefold Symmetry*, Ed. I. Hargittai, VCH Publishers, New York, 1990, chapters 14-17.
53. D. J. Klein, in *From Chemical Topology to Three-Dimensional Geometry*, Ed. A.T. Balaban, Plenum Press, New York, 1997.
54. P. W. Fowler, in *From Chemical Topology to Three-Dimensional Geometry*, Ed. A. T. Balaban, Plenum Press, New York, 1997.
55. A. T. Balaban, *Bull. Soc. Chim. Belg.* 105 (1996) 383.
56. Y. D. Gao and W. C. Herndon, *J. Am. Chem. Soc.* 115 (1993) 8459.
57. D. J. Klein and X. Liu, *Int. J. Quantum Chem, Quantum Chem. Symp.* 28 (1994) 501.
58. A. T. Balaban, T. G. Schmalz, H. Zhu and D. J. Klein, *Theochem* 363 (1996) 291.
59. A. T. Balaban, X. Liu, D. J. Klein, D. Babic, T. G. Schmalz, W. A. Seitz and M. Randic, *J. Chem. Inf. Comput. Sci.* 35 (1995) 396.
60. A. J. Stone and D. J. Wales, *Chem. Phys. Lett.* 128 (1986) 501.
61. D. Babic and N. Trinajstic, *Comput. Chem.* 17 (1993) 271 ; *Chem. Phys. Lett.* 237 (1995) 239.
62. A. L. Goodson, C. L. Gladys and D. E. Worst, *J. Chem. Inf. Comput. Sci.* 35 (1995) 969.
63. D. E. Manolopoulos, J. C. May and S. E. Down, *Chem. Phys. Lett.* 181 (1991) 105 ; P. W. Fowler, *Chem. Phys. Lett.* 131 (1986) 444 ; D. E. Manolopoulos and P. W. Fowler, *Chem. Phys. Lett.* 204 (1993) 1 ; P. W. Fowler, D. E. Manolopoulos, D. B. Redmond and R. P. Ryan, *Chem. Phys. Lett.* 202 (1993) 1113.
64. R. Taylor, *J. Chem. Soc. Perkin Trans 2* (1993) 813.
65. D. Babic, A. T. Balaban and D. J. Klein, *J. Chem. Inf. Comput. Sci.* 35 (1995) 515.
66. A. T. Balaban, D. Babic and D. J. Klein, *J. Chem. Educ.* 72 (1995) 693.
67. S. Iijima, *Nature* 354 (1991) 56 ; S. Iijima and T. Ichihashi, *Nature* 363 (1993) 603.
68. P. M. Ajayan and S. Iijima, *Nature* 361 (1993) 333.

69. P. M. Ajayan, T. Ichihashi and S. Iijima, *Chem. Phys. Lett.* 202 (1993) 384.
70. T. W. Ebbesen and P. M. Ajayan, *Nature* 358 (1992) 220.
71. D. Dujardin, T.W. Ebbesen, H. Hiura and K. Tanigaki, *Science* 265 (1994) 1850.
72. T. W. Ebbesen, P. M. Ajayan, H. Hiura and K. Tanigaki, *Nature* 367 (1994) 519 .
73. C. T. White, D. H. Robertson and J. W. Mintmire, *Phys. Rev. B* 47 (1993) 5485.
74. M. S. Dresselhaus, G. Dresselhaus and R. Saito, *Phys. Rev. B* 45 (1992) 6234 ; R. Saito, M. Fujita, G. Dresselhaus and M. S. Dresselhaus, *Phys. Rev. B* 46 (1992) 1804 .
75. E. C. Kirby, in *From Chemical Topology to Three-Dimensional Geometry*, Ed. A. T. Balaban, Plenum Press, New York, 1997; *Croat. Chem. Acta* 66 (1993) 13.
76. D. J. Klein, *J. Chem. Inf. Comput. Sci.* 34 (1994) 453.
77. A. Thess, R. Lee, P. Nikolaev, H. Dai, P. Petit, J. Robert, C. Xu, Y. H. Lee, S. G. Kim, A. G. Rinzler, D. T. Colbert, G. E. Scuseria, D. Tomanek, J. E. Fischer and R. E. Smalley, *Science* 273 (1996) 483.
78. A. T. Balaban, H. Zhu and D. J. Klein, *Fuller. Sci. Technol.* 3 (1995) 133.
79. A. T. Balaban, *Chem. Brit.* 28 (1992) 1090.
80. R. H. Wentorf Jr., *J. Chem. Phys.* 26 (1957) 956.
81. X. F. Xia, D. A. Jelski, J. R. Bowser and T. F. George, *J. Am. Chem. Soc.* 114 (1992) 6493.
82. D. M. Poirier, T. R. Ohno, G. H. Kroll, Y. Chen, P. J. Benning, J. H. Weaver, L. P. F. Chibante and R. E. Smally, *Science* 253 (1991) 646.
83. M. Prato, Q. C. Li, F. Wudl and V. Lucchini, *J. Am. Chem. Soc.* 115 (1993) 1148.
84. A. T. Balaban, *Math. Chem. (MATCH)* 33 (1996) 25.
85. R. H. Baughman, D. S. Galvao, C. Cui, Y. Wang and D. Tomanek, *Chem. Phys. Lett.* 204 (1993) 8.
86. F. Diederich, Y. Rubin, C. B. Knobler, R. L. Whetten, K. E. Schriver and K. H. Houk, *Science* 245 (1989) 1088.
87. J. Hunter, J. Fye and M. F. Jarrold, *J. Phys. Chem.* 97 (1993) 3460.
88. S. Osawa, M. Harada and E. Osawa, *Fuller. Sci. Technol.* 3 (1995) 225.
89. A. T. Balaban, W. A. Seitz and D. J. Klein, *Bull. Soc. Chim. Belg.* 104 (1995) 525.
90. A. T. Balaban, W. A. Seitz and D. J. Klein, *Fuller. Sci. Technol.* 4 (1996) 467.
91. A. T. Balaban, *Izv. Akad. Nauk SSSR, Otd. Khim. Nauk* (1960) 2064.
92. P. R. Schleyer, A. J. Kos, J. A. Pople and A. T. Balaban, *J. Am. Chem. Soc.* 104 (1982) 3771.

Protein transmembrane structure: recognition and prediction by using hydrophobicity scales through preference functions

Davor Juretić,^a Bono Lučić,^b Damir Zucić^c and Nenad Trinajstić^b

^aDepartment of Physics, Faculty of Science, N. Tesle 12, HR-21001 Split, Croatia

^bThe Rugjer Bošković Institute, P.O.B. 1016, HR-10001 Zagreb, Croatia

^cFaculty of Electrical Engineering, Kneza Trpimira 2b, University of Osijek, HR-3100 Osijek, Croatia

1. INTRODUCTION

The problem of structure prediction for proteins involves secondary structure prediction based on sequence analysis as the first step [1]. Secondary structure prediction algorithms [2,3] that worked reasonably well with soluble proteins were considered inadequate for membrane proteins [4]. Recently, different artificial neural network algorithms have been used to predict secondary structure in globular soluble proteins [5-7] and sequence location of transmembrane segments (TMS) in integral membrane proteins [8,9]. When a small data base of structural features is used to train such algorithms there is always a danger of overtraining and that is precisely the case with integral membrane proteins. The structure of only several membrane proteins is known with high enough resolution for unambiguous assignment of secondary structure features [10-15]. Enlarging the data base by the use of, for example, the SWISS-PROT sequence data base [16] assignments of potential TMS as the 'standard of truth' is also connected with serious problems: in general, secondary structure information is not provided and erroneous assignments may be present in the data base.

Training is not necessary for simpler algorithms for analysis of hydrophobicity profiles [17-21]. However, recent improvements of sliding window algorithms [22] optimize all variable parameters by using the very restricted number of integral membrane proteins of known structure. Such procedure leads to overtraining and to a significant drop in prediction quality for unrelated proteins. In general, overprediction of putative TMS, and a need for subjective decision about their location and length is a common deficiency of hydrophobicity plots. As often observed [18], hydrophobicity alone is not enough to detect membrane spanning domains and their secondary structure conformation, because folding into the TMS conformation is controlled by the primary structure context. Sequence folding codes may be simpler for globular membrane proteins [23] than for globular soluble proteins, but paucity of known membrane protein structures is still making it very difficult to recognize such codes. Recognition of putative TMS from hydrophobicity plots may seem to be easy, but prediction of such segments must be accompanied with prediction accuracy assessment to be meaningful. In

spite of these shortcomings hydrophobicity plots are still considered among most promising approaches to successful structure prediction schemes [24].

For a large number of deduced sequences, coming out daily from different genome projects, theoretical sequence analysis is the only possible method for predicting TMS and deciphering transmembrane topology. At present membrane-embedded domains can be predicted with good accuracy but this is not the case with the secondary structure of these domains. This is important deficiency of the sliding window methods based on sequence hydrophobicity, because only the secondary structure information can serve as the starting point for predicting protein assembly into the final three-dimensional structure. In this report we shall describe a theoretical method based on hydropathy analysis that accurately predicts not only the sequence location of transmembrane segments but their secondary structure conformation as well.

The conformation of TMS seems to be α -helical for most membrane proteins [12,13,25-27], but there are some proteins, such as porins, that have TMS in the β -sheet conformation [28]. There may also exist proteins with both helical and β -strand transmembrane segments or with transmembrane helical segments combined with still unknown topology of membrane buried β -strands [29-31]. This work is focused on the prediction of transmembrane helical segments (TMH), but our algorithms do allow the prediction of transmembrane or surface attached β -strands (TMBS) as well.

Our approach is to associate given amino acid type both with its secondary structure conformation and with hydrophobicities of its sequence neighbors in a carefully selected reference set of membrane and soluble proteins. Conformational preferences are then calculated. Since preferences depend not only on amino acid type but also on amino acid attributes and local sequence context, our predictor is using preference functions [32]. Very high preferences for the α -helix conformation (often higher than 4.0) are then associated with residues known to have very hydrophobic sequence environment inside transmembrane segments with known sequence location. Based on chosen scale of 20 amino acid attributes (such as hydrophobicity, polarity, statistical preferences), secondary structure conformation is first predicted as α -helical, β -sheet, turn or undefined conformation in given protein sequence and secondly those segments are selected that have high preference for the membrane-embedded conformation.

In effect, our method predicts 6 different secondary structure conformations: α -helical, β -sheet, turn, undefined, TMH and TMBS. Only primary structure segments with predicted long uninterrupted stretches of α -helical residues with high maximum preference for helical configuration are considered as candidates for the TMH. Longer β -strands are also predicted and, at least in porins, are never confused with TMH. We have no false positive predictions of TMH in porins, but we do have false positive predictions of TMH in some soluble proteins.

By using the cross-validation statistical procedure and Kyte-Doolittle hydropathy scale, the prediction results for TMH in the training data base of 63 membrane proteins common to us and to Rost *et al.* [9] and also to Jones *et al.* [33] were similar in accuracy by all three methods. When training data base is enlarged to 168 proteins, we maintain the 95% accuracy for predicted transmembrane helices and almost 80% (78.6%) of proteins are predicted with 100% correct transmembrane topology. When 168 proteins are divided in the above mentioned training set of 63 proteins and an independent test set of 105 proteins, all performance parameters for TMH prediction associated with a set of 105 proteins exhibited a decrease which was smaller in our case than for Rost *et al.* [9].

2. METHODS

2.1. Selecting protein data bases for training and for testing

Rost *et al.* training list of proteins [9] and SWISS-PROT sequence data base [16] release 29 and 31 were used to select training and testing sets of proteins. We examined more than 4000 proteins with transmembrane domains mostly selected from the SWISS-PROT data base release 29. A total of 168 integral membrane proteins were finally selected. These are in alphabetical order with the SWISS-PROT release code or letter 'r' (for the Rost *et al.* proteins [9]). Letter 's' is added when appropriate to indicate that signal sequence has been removed:

4f2_human(r), 5ht3_mouse(rs), alaa_human(r), a2aa_human(r), a4_human(rs), aalr_canfa(r), aa2a_canfa(r), ach1_xenla(29s), acm5_human(29), adt_ricpr(r), adt2_yeast(29), ag22_mouse(29), aqp1_human(29), athb_rat(29), athp_neucr(29), atm1_yeast(31), atn1_human(29), atp9_wheat(29), atpl_ecoli(31), b3at_human(29), bach_hallm(r), bacr_halha(r), c561_bovin(29), cadn_mouse(29s), car1_dicdi(29), cb2r_human(29), cb21_pea(r), cd2_human(29s), cd7_human(29s), cd72_human(29), cd8a_human(29s), cek2_chick(rs), cgcc_bovin(29), cic1_cypca(29), cikl_drome(29), cox2_parli(29), cox9_yeast(29), cp5a_cantr(29), cxb5_rat(29), cyda_ecoli(29), cydb_ecoli(29), cyf_brara(29), cyoa_ecoli(rs), cyob_ecoli(r), cyoc_ecoli(r), cyod_ecoli(r), cyoe_ecoli(r), dhg_ecoli(31), dhsc_bacsu(29), divb_bacsu(29), dm5c_ecoli(29), dsbb_ecoli(31), edg1_human(r), egf_mouse(31), exbb_ecoli(29), fce2_human(r), fixl_rhime(29), fmlr_rabit(29), frdd_provu(29), ftsl_ecoli(29), ftsh_ecoli(29), furi_human(29s), g2lf_human(29), gaal_bovin(29), gasr_human(29), gesr_human(31s), ghr_human(29s), glp_pig(r), glpa_human(rs), glpc_human(r), glra_rat(rs), gmcr_human(rs), gp1b_human(rs), gpt_crilo(r), grhr_human(29), ha21_human(29s), hb23_mouse(29), hema_cdvo(r), hema_measa(r), hema_pi4ma(r), hg2a_human(r), hly4_ecoli(29), hmdh_human(29), iggb_strsp(r), il2a_human(rs), il2b_human(rs), imma_citfr(29), isp6_yeast(29), ita5_mouse(r), itb1_human(29s), kdgl_ecoli(31), kgtp_ecoli(29), lacy_ecoli(r), lech_human(r), leci_mouse(r), lep_ecoli(r), lha1_rhosh(29), lhb4_rhopa(29), ly49_mouse(29), m49_strpy(29s), magl_mouse(rs), malf_ecoli(r), malg_ecoli(29), mas6_yeast(31), mdr3_human(31), melb_ecoli(29), mepa_mouse(31s), mota_ecoli(29), motb_ecoli(r), mpcp_rat(29), mprd_human(rs), myp0_human(rs), mypr_human(29), nals_bovin(29), nep_human(r), ngfr_human(rs), nk1i_mouse(29), nntm_bovin(31), nram_iabda(29), och1_yeast(29), oec6_spiol(29), oppb_salty(r), oppc_salty(r), ops1_calvi(r), ops2_drome(r), ops3_drome(r), ops4_drome(r), opsb_human(r), opsd_human(r), opsg_human(r), oprs_human(r), pigr_human(r), psaa_pynth(31), psab_pynth(31), psbi_horvu(29), ptgb_ecoli(31), ptma_ecoli(r), sece_ecoli(r), secy_ecoli(29), spc2_canfa(29), spir_spime(29), stub_drome(29), sy65_drome(29), syb1_human(29), synp_rat(29), tal6_human(29), tapa_human(29), tat2_yeast(31), tca_human(29), tcb1_rabit(r), tccl1_mouse(29), tcrb_bacsu(31), tee6_strpy(29s), tgfa_human(31s), thas_human(29), tnfa_bovin(29), tnr1_human(31), trbm_human(rs), trsr_human(r), tsa4_gjala(31s), ucp_rat(29), va34_vacce(29), vca1_human(29s), vglg_hrsva(29), vmt2_iaann(r), vnb_inbbe(r), vs10_rotbn(29), wapa_strmu(29s).

Of these proteins 80 had a single transmembrane helix, 6 had 2, 6 had 3, 14 had 4, 4 had 5, 13 had 6, 24 had 7, 5 had 8, 2 had 10, 3 had 11, 9 had 12, and 2 had more than 12 TMH. There were 662 expected transmembrane segments with 14359 residues in 'observed' transmembrane helix configuration among total number of 67155 residues. In the selection

process preference was given to proteins with transmembrane domains without associated label (such as 'putative'). Proteins with 'probable' or 'potential' transmembrane domains were also collected in the case when such segments were sandwiched between protein domains of known cytoplasmatic and extracellular identity. Signal sequences, claimed as such in the SWISS-PROT, were omitted.

In order to facilitate comparisons with other statistical methods the training data base of proteins contained a subset of the training data base selected by Rost *et al.* [9] that was already a subset of the training data base selected by Jones *et al.* [33]. The omission of some proteins first by Rost and then by us decreased the size of the original data base from 83 (Jones *et al.*, [33]) to 69 (Rost *et al.*, [9]) and to 63 (this work). While Rost omitted 14 proteins because they were less well determined experimentally than other proteins included for the training procedure by Jones, a few additional polypeptides were omitted by us because of their known X-ray structure that we later used for rigorous testing of our algorithm. Two proteins longer than 1000 amino acids were omitted, too. Some proteins from the 168 protein list have been taken without N-terminal or C-terminal amino acids in undefined conformation so that their final length is also less than 1000 residues. These are: *atn1_human* with omitted 23 C-terminal amino acids, *cic1_cypca* with only first 720 amino acids taken out of 1852, *egf_mouse* with omitted 400 N-terminal amino acids, *mdr3_human* with omitted 279 C-terminal amino acids and *ntnm_bov* with omitted 200 N-terminal amino acids.

We took care that all polypeptides selected by us show less than 30% similarity with any other polypeptide used in the training process. Of 10 proteins from the test list of membrane proteins with the best known structure only one (plant light-harvesting complex) had its twin (*cb21_pea*) in the data base of 168 proteins. The similarity was judged by the HSSP data base of [34]. An exception to that rule are several of 63 proteins selected by Rost *et al.* [9]. These are *hema_cdvo* and *hema_measa* with 40% similarity, *opsb* and *opsd* with 41% similarity, *opsb* and *opsg* with 37% similarity, *opsb* and *opsr* with 37% similarity, *opsd* and *opsg* with 36% similarity, *opsd* and *opsr* with 35% similarity, *ops4* and *ops3* with 68% similarity, *aa1r* and *aa2a* with 44% similarity and *opsg* and *opsr* with 97% similarity. All of 105 proteins selected by us from SWISS-PROT releases 29 and 31 were less than 30% similar to each other and less than 30% similar to any other protein of the complete set of 168 protein. The rule of less than 30% similarity among tested proteins was not maintained for some special data bases of such proteins, such as the above mentioned collection of integral membrane proteins of the α -class with known high resolution X-ray structure. On the other hand, with the exception of light-harvesting complex, we always made sure that no tested protein was more than 30% similar to any protein from the training list of proteins.

All potential transmembrane segments in the reference set of 168 integral membrane proteins were considered to be in the α -helix conformation during training process. Five residues next to each observed TMH were considered to be in the turn conformation, while all other residues were regarded as present in the undefined conformation. Soluble proteins and membrane proteins with solved structure were analyzed with the Kabsch and Sander program DSSP [35], which assigned the secondary structure. All helical conformations 'H', 'I', and 'G' were lumped into α -helical 'H', all beta into 'B', all turn into 'T', while all remaining residues were considered to be in the 'U' conformation. Transmembrane helices broken with turn residues in the SWISS-PROT data base or by the DSSP algorithm, but reported as transmembrane helices in the original papers, were considered to be unbroken string of 'H' residues.

Two sets of soluble globular proteins were selected from Protein Data Bank (PDB) for testing for false positive results. In the set SOLU1 of 187 such proteins resolution for each protein was equal or better than 3 Å. Secondary structure conformations were determined by the DSSP algorithm. In the set SOLU2 of 147 proteins (protein data set used in [7] plus 21 additional proteins) only proteins known with equal or better than 2.5 Å resolution were included. There was less than 25% pairwise similarity. Three different secondary structures were determined as described by Rost and Sander [7]. Both data sets are available in the Supplementary Material.

Two sets of β -class soluble proteins were used for training.

a) The first set of 37 such proteins SOLB1 has been selected from Protein Data Bank among soluble proteins known with equal or better than 3 Å resolution. When more than one chain was present in the protein only the first polypeptide chain denoted with the last letter '1' has been selected:

1acx, 1bbp1, 1cd4, 1fdl1, 1hnel, 1mcp1, 1paz, 1pfc, 1rbp, 1rei, 1sqt, 1ton1, 1trm1, 2alp, 2apr, 2aza1, 2fb41, 2fbj1, 2gch1, 2cna, 2gcr, 2i1b, 2ltn, 2pcy, 2pka1, 2ptn, 2rhe, 2rsp1, 2sga, 2sod1, 2tbv1, 3est, 3rp2, 3sgb1, 4ape, 4cms1, 5pep.

b) The second set of 39 such proteins SOLB2 has been selected from SOLU2 data set:

1azu, 1bbp_A, 1bds, 1bmv_1, 1bmv_2, 1cbh, 1cd4, 1cdt_A, 1fc2_D, 1fdl_H, 1mcp_L, 1rnh, 1sh1, 2alp, 2gcr, 2i1b, 2ltn_A, 2ltn_B, 2mev_1, 2mev_3, 2pab_A, 2pcy, 2pka_A, 2rsp_A, 2sod_B, 2stv, 3ait, 3ebx, 3hla_B, 3hmg_A, 4cms, 4cpa_I, 4rhv_1, 4rhv_3, 4sgb_I, 5er2_E, 5hvp_A, 6hir, 9api_B.

The data base of the best known 10 integral membrane proteins with transmembrane helical segments (BESTP) consisted of photosynthetic reaction center subunits H, L and M from *Rhodobacter viridis* [12,36] and *Rhodobacter sphaeroides* [25], plant light-harvesting complex LHC-II [13], light-harvesting protein LHA2 from *Rhodospseudomonas acidophila* [27], and two human class I histocompatibility antigens 1b14 [37] and 1a02 [38]. The X-ray structure for the single transmembrane segment of each histocompatibility antigen was not determined, but since all the rest of the three-dimensional structures of these proteins was determined by the X-ray crystallography, we considered the combination of experimental and theoretical methods used to describe their structure powerful enough to include these proteins among the best known integral membrane proteins.

The data base PORINS consisted of seven porins and two defensins all with known or proposed transmembrane β -strand structure. The porins with known X-ray structure were porin from *Rhodobacter capsulatus* [10,28] and porins PhoE and OmpF from *Escherichia coli* [39,11]. Porins with proposed transmembrane β -barrel topology were anion-selective porin Omp32 from *Comamonas acidovorans* [40], outer membrane protein OmpA from *Escherichia coli* K12 membrane (membrane-embedded fragment residues 1 to 177, [41,42]), and mitochondrial outer membrane porin from human B-lymphocytes [43] and from *Neurospora crassa* [44]. Two defensins of known structure were HNP-3 [45] and defensin from larvae of the dragonfly *Aeschna cyanea* [46].

2.2. Main performance parameters used to judge the prediction quality

a) Parameters for individual residues are composed of correct positive predictions p, correct negative predictions n, overpredictions o and underpredictions u for all residues found in the protein data base. One such parameter is the fraction of residues predicted in correct secondary conformation:

$$Q_3 = (p_1 + p_2 + p_3) / N$$

where secondary conformations are helix, beta and everything else (turn, undefined or coil) found in the data base having a total of N residues. Another such parameter [47] is

$$A_i = (N_i - o_i - u_i) / N_i$$

where i is the index of chosen secondary conformation with N_i residues from protein data base found in that conformation, while o_i and u_i are respectively overpredicted and underpredicted residues in that conformation. While lower bound for the Q parameter is 0, the A parameter can be large negative number for poor prediction. For α -helical, β -strand, turn, undefined and TMH conformation A parameters are respectively A_h , A_b , A_t , A_u and A_{TM} .

b) Parameters for TMH segments as prediction units. Parameter of the A type measures prediction accuracy for transmembrane segments instead of prediction accuracy for individual residues:

$$A_s = (N_s - o_s - u_s) / N_s$$

where s denotes transmembrane segment. There are N_s observed transmembrane segments, u_s underpredicted and o_s overpredicted segments. Even simpler performance measure is the fraction of correctly predicted TMH:

$$Q_s = N_{cs} / N_s$$

where N_{cs} is the number of correctly predicted TMH. There must be an overlap of at least 9 residues in the TMH conformation between predicted and observed TMH for the case of correctly predicted TMH.

c) Protein topology parameters. If there are n_c proteins with 100% correctly predicted transmembrane topology (all TMH correctly predicted in correct sequence positions) out of the total number of n tested proteins, than a very useful parameter is

$$Q_p = n_c / n$$

Our algorithm also reports absolute values of: a) residues correctly predicted, overpredicted and underpredicted in the TMH conformation, b) transmembrane helical segments predicted, correctly predicted, overpredicted and underpredicted as TMH, c) proteins recognized as membrane proteins and d) proteins recognized with 100% correct topology.

2.3. Hydrophobic moment profile

Hydrophobic moment profile is calculated as described by Eisenberg *et al.* [48] to collect information about possible amphipathic helices and strands. We used only the PRIFT scale (# 27 in Table 5) to find hydrophobic moments. Scales used for the calculation of hydrophobic moments were not normalized. The PRIFT scale produces high moments (sometimes higher than 2.0) for sequence segments known to be highly amphipathic. An ideal α -helix twist angle of 100 was used to associate α -helix hydrophobic moments with all

sequence positions. Less ideal angle of 162 (more appropriate to the β -barrel structure) was used to produce sequence profile of β -strand moments.

2.3.1. The training procedure for the preference functions method

The prediction is based on the method of preference functions [32]. The PREF suite of algorithms in the present version (PREF 3.0) consists of training and testing algorithms called PREF (PREference Functions) and SPLIT (predicted long helices are SPLIT into two or three TMH), respectively. The first obligatory step in the PREF algorithm is the choice of amino acids scale of 20 values. Secondly, data sets of proteins are selected to train the algorithm. Standard training procedure uses the Kyte and Doolittle hydrophobicity scale [17], 168 integral membrane proteins listed above and 37 soluble β -class proteins (SOLB1). With a chosen scale, sequence environment is calculated for each amino acid type associated with one of four secondary conformations (helix, sheet, turn and undefined) at each sequence position, as an average over hydrophobicity values of neighboring 10 amino acids. The amino acid attribute of the central amino acid residue in the sliding window is not taken into account to calculate sequence environment. This is being done for the whole data set of proteins. Collected sequence environments are grouped into nine classes so that about equal number of environments is collected into each class. For the best scales, histograms of frequency distributions for environments for the same amino acid type differ significantly for different secondary conformations. This is most easily seen if frequency distributions are replaced with corresponding Gaussian functions. For each amino acid type in each secondary conformation three Gaussian parameters are extracted from observed frequency distributions. These are: a) the number of sequence environments, b) average value for sequence environments and c) standard deviation for sequence environments. All such parameters ($3 \times 20 \times 4$ if four different folding motifs are considered) are collected in the file with Gaussian parameters (enclosed in the Supplementary Material, Table III).

2.3.2. The testing procedure

Preference functions are calculated in the SPLIT algorithm as described before (ref. [32]; equations (2) and (3)). For instance, up to the constant factor, the preference function for alanine in helix conformation is found as the ratio of the Gaussian function for alanine in helix conformation to sum of Gaussian functions for alanine in all four conformations. The constant divisor is the frequency of helix conformation in protein data set. For tested protein preference functions are evaluated for all amino acids and for all four conformations. Ratio of Gaussians, as probability to find conformational motif, can be very successful in detecting such motifs, when overlap of corresponding distributions is not too great, or, in other words, when different conformations can be associated with different scores or averages (as proved by Lupas *et al.* [49] in their statistical method for detecting coiled-coil structures).

2.3.3. Decision constants choice

The automatic choice of decision constants (DC) is the standard feature of the testing procedure by the SPLIT algorithm. In the first prediction loop, preliminary prediction results for tested protein are used for the automatic determination of decision constants for helix (dch), sheet (dce) and coil (dcc) conformation (turn or undefined). Each choice of decision constants is made sequentially and independently of previous choices in the following order: Constants $dch = 0.3$, $dce = -0.6$ and $dcc = 0$ are chosen when predicted helical conformation is

greater than 30% and percentage of charged amino acids is less than 20%. Constants $dch = -0.2$, $dce = 0.4$ and $dcc = 0$ are chosen when percentage of predicted sheet conformation is higher than 25%, while the percentage of predicted helical conformation is less than 15%. In the case if predicted helical conformation is higher than 25%, protein is longer than 300 amino acids and predicted number of transmembrane helices is higher than 6, then chosen decision constants are $dch = 0.4$, $dce = -0.2$ and $dcc = 0$. For all other possibilities decision constants are all set to zero. The algorithm is used without decision constants by setting initially all decision constants to zero, only when so noted in the text!

2.3.4. Collection of environments and smoothing procedure

Except when testing the window length influence on prediction performance a sliding window length of 11 residues was used throughout this report in such a way that central residue in the window was omitted from averaging procedure. Resulting sequence environments are then used for the evaluation of preference functions. In practice it is advantageous to smooth these preferences before comparing preference profiles. Seven residue preferences are smoothed for the 'H' conformation, five for the 'B' conformation and three for the 'U' or 'T' conformation. The smoothed value is always assigned to the residue in the middle of the sliding window. Corresponding decision constants are added to strings of smoothed preference values. Numerical values for smoothed preferences for four conformational states are then compared and secondary structure is assigned to the highest preference. In the remaining text whenever preferences are mentioned or reported it should be understood that we have in mind the sequence profile of smoothed preferences.

2.3.5. Filtering procedure

Unrealistic assignments of a single isolated residue assuming helical or beta sheet conformation, among two left and two right neighbors in nonregular conformations, are corrected by introducing nonregular ('U') conformation for such residues. Isolated residues in 'B' conformation surrounded by two left and two right residues in 'H' conformations are reassigned as residues in helical conformation. Similarly the BBHBB pattern is transformed into BBBBB. Two arginines neighbors or two proline neighbors are assigned to nonregular ('U') conformation whenever found with helix preference less than 3.0.

The essential part of the algorithm recognizes transmembrane helix conformation as the fifth possible conformation. The first appearance of the 'H' conformation is memorized and subsequent 'H' residues are counted if not interrupted by any other conformational assignment. The value for maximum helical preference is also memorized in each helical segment. String of helical residues is considered as possible transmembrane helix if found to be longer than 12 residues with maximum preference for helical conformation higher than 2.7. Residues predicted in the 'B' conformation as neighbors to candidate TMH segment are used to elongate it at both ends. Even shorter helical segments (from 9 to 14 residues in length) are memorized and fused together if less than six residues apart with at least one maximum helix preference higher than 2.7. Total number of predicted C caps is memorized at this stage and used as information about total number of predicted transmembrane helices in the protein. The percentage of predicted helical and sheet residues with respect to protein sequence is also calculated in order to determine decision constants for the next prediction cycle.

In order to compare predicted and observed transmembrane helices automatic extraction of observed transmembrane segments with probable helical conformation is also

performed. It was necessary to consider all uninterrupted transmembrane segments longer than 13 and shorter than 38 residues as potential transmembrane helical segments.

The main part of the filter is designed to reexamine potential transmembrane helical segments and to shorten or split TMH of unrealistic length. All candidates for predicted TMH are divided into five groups: short segments having 13 to 16 residues, normal length segments having 17 to 27 residues, long segments having 28 to 35 residues, very long segments having 36 to 54 residues and obviously wrong predictions of segments longer than 54 residues.

Short TMH are eliminated if their TMH preference peak is less than 2.7, and also in the case three of residues E, P, K, D, R are present in the segment with maximum helix peak less than 4.0. Normal length TMH are shortened from both ends in the case when any of charged amino acids: arginine, lysine, aspartic or glutamic acid are found inside first four and last four positions of the putative transmembrane segment. In addition, turn preference had to be greater than 1.0 for these amino acids for shortening to take effect. New N and C caps are positioned at the first residue inside segment (going from old helix caps in the direction of helix middle) that could remain in the helical conformation. We shall call this subroutine the CHARGE-BREAK subroutine. Also disregarded are TMH that are too short (shorter than 17 amino acids) after CHARGE-BREAK routine, and of not enough high helical preference peak (less than 2.7).

In the case if length of putative TMH remains equal or greater than 24 FILTER subroutine is applied. It shortens TMH on both ends until helix preference becomes too high. Helical preference is multiplied with number of residues reached from the cap residue position and resulting value compared with $(\text{TMH length} - 21)/2$. The FILTER creates new helix cap positions closer to the middle of TMH. The shift in the new cap positions is greater for lower helical preference and for longer TMH.

Long TMH's, having 28 to 35 residues, are shortened by using the TURN-BREAK subroutine. In brief, residues inside helix and next to each cap are examined with respect to their turn preferences. If maximum turn preference is greater than 1.0 then corresponding cap position is shifted to the position next to turn preference maximum in the direction of helix middle. In the case if remaining TMH is still longer than 24 residues, than CHARGE-BREAK and FILTER subroutine is applied. Predicted helical segments longer than 35 residues are broken into two or three segments with the TURN-BREAK subroutine and with a help of additional similar routine for finding maximum α -helix preference, while both TURN-BREAK and FILTER routine is used to shorten remaining segments that are still too long.

After ending the main filter routine the 'T' conformation is assigned to four residues next to each predicted helix cap. Also peaks in helix preference higher than 4.0 are examined for the whole sequence. Additional (overlooked) TMH is assigned as 15 residue segment centered around such peak if a) less than three of K, P, D, R, E residues are present in such a segment, if b) such segment is at least 20 residues removed from sequence terminals, and if c) TMH was not previously predicted in that position.

2.3.6. Predicting transmembrane β -strands (TMBS)

The SPLIT algorithm was optimized for predicting transmembrane α -helices by using the Kyte-Doolittle hydropathy scale to create profile of α -helix preferences. The digital version of prediction for transmembrane α -helices is designated as the TMH predictor. Predicted profile of β -strand preferences can be used to find sequence location of potential membrane-embedded or surface-attached β -strands. The score for potential membrane-attached β -strand

conformation is found by summing up β -sheet preference and β -sheet hydrophobic moment (calculated using PRIFT scale [50]) for each sequence position. The digital version of the prediction for potential membrane-embedded β -strands (TMBS predictor) is then found as collection of sequence segments at least 6 residues long with each residue-associated score higher than 2.0.

2.3.7. Adopted cross-validation technique

The prediction performance statistics is better for larger number of proteins tested. All proteins included in the training data base can be used for testing as well if the jack-knife or cross-validation technique is adopted (see ref. [7] concerning the necessity of using this statistical technique to estimate the prediction performance). We used 5-times cross-validation to obtain representative results for the reference set of 168 integral membrane proteins after extracting preference functions with the Kyte-Doolittle hydrophathy scale. The same set of 37 soluble proteins of the β -class (SOLB1) was always included in the training data base of proteins. It was noticed that prediction results are sensitive to the type of transmembrane topology. Therefore, for the 5-times cross-validation, all proteins were grouped according to expected number of transmembrane segments. We took care that each group of tested membrane proteins (33 or 34 proteins) has similar distribution of proteins with respect to their transmembrane topology as the total set of 168 membrane proteins. For instance smaller reference set of 135 proteins, used in the 'best' training procedure, is left when following 33 proteins are removed from the original reference set: cd72, cd7, cd8a, cek2, cp5a, egf, vca1, va34, tsa4, trsr, trbm, ghr, glp, glpa, glpc, gmcr, gp1b, atpl, exbb, cxb5, dsbb, atm1, bach, car1, cb2r, cyda, edg1, fmlr, opsb, athp, gpt, b3at, tat2. In some explicitly stated cases the 2-times cross validation procedure was used such that training set of 168 proteins was divided into 63 proteins selected by Jones *et al.* [33] and Rost *et al.* [9], and 105 proteins selected by us.

Both training and testing process take only several minutes on the PC equipped with the 486 processor in the case when up to 200 proteins are used. The FORTRAN source code, files with Gaussian parameters and protein data bases used in this report are available via Internet (see Supplementary Material).

3. RESULTS

3.1. Conformational preference for transmembrane α -helix is strongly dependent on sequence hydrophobic environment for most amino acid types

When only transmembrane segments, expected to be in the helical conformation, are used to train the algorithm to predict helical segments, then preference for the α -helix conformation ('H') is at the same time the preference for the TMH conformation. In our case the training part of the algorithm uses so small percentage of the observed 'H' conformations in soluble proteins (because only soluble proteins of the β -class are used) that we can still consider the 'H' conformational preference as the transmembrane helical segment conformational preference. It appears that some amino acid types passively acquire the conformation dictated by their neighbors (Figure 1), while others (mainly charged amino acids) are able to resist to some extent (Table 1). Extremely secure dependence of TMH preference on the hydrophobic sequence environment is found for 12 amino acid types (Table 1). Only

Arg, Lys, Asp and Glu have the F factor (a statistical measure for the dependence of preference on hydrophobicity of sequence neighbors) less than 50.

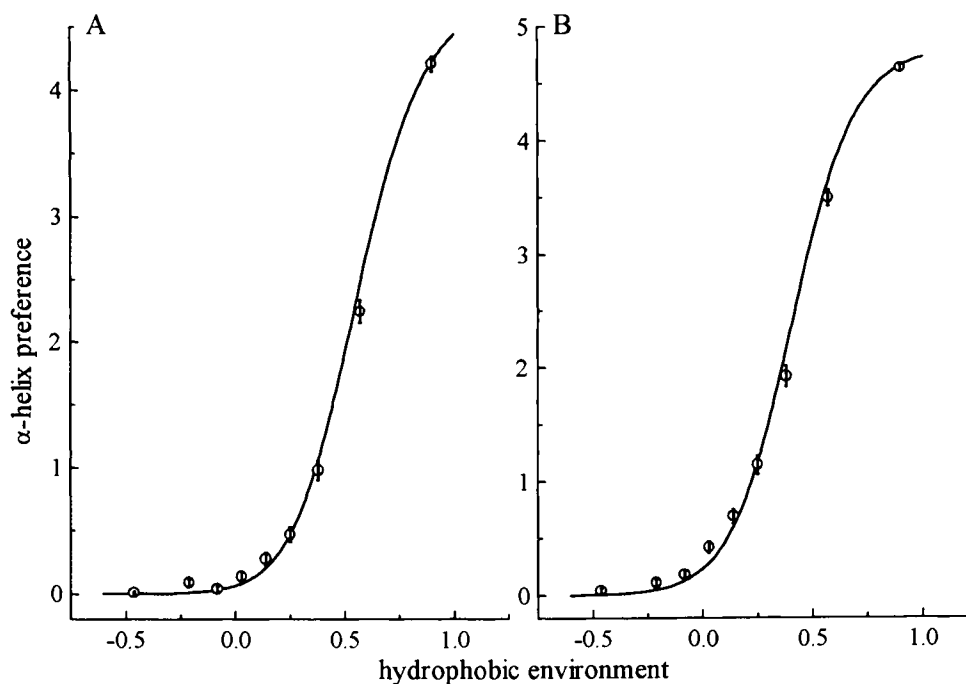


Figure 1: Very strong dependence of the α -helix conformational preferences on average hydrophobic sequence environment. Standard training procedure (Methods) was used. Observed preferences for glycine (Figure 1A) and leucine (Figure 1B) are shown as open points. Confidence limits, shown as bars above and below preference points, were calculated as described by Ptitsyn [51] so that it was 67.5% certain that observed preferences would fall between these values. The preference functions for leucine and glycine are shown as full lines.

From Figure 1 it is quite clear that linear approximation for the dependence of TMH preference on sequence environment is not so good as the preference function approximation. Similar results are obtained for other 18 amino acids (not shown). It is also clear that preference functions can be regarded as good but not the best nonlinear fit to observed preference points. For the four state model of secondary structure the preference function is obtained as the ratio of one Gaussian function to four Gaussian functions (Methods). Normal distribution (Gaussian function) is expected to be good fit for the histogram of sequence environments [32,52] due to averaging procedure used to produce histograms for each amino acid type and each secondary structure. However, cases of nonrandom distribution of amino acid types among sequence environments for particular secondary structure motifs have been observed [52] as well as the cases of too small number of sequence environments for the particular class of sequence environments, chosen amino acid type, secondary structure and

Table 1
 Statistical parameters derived for the linear approximation of the dependence of helix preference on hydrophobic environment

Amino acid type	Slope: b	Standard error in slope: s_b	F parameter $(b/s_b)^2$:
Ala	2.726	0.034	6360
Arg	0.088	0.033	7
Asn	0.601	0.052	131
Asp	0.210	0.038	31
Cys	2.441	0.050	2387
Gln	0.337	0.045	57
Glu	0.173	0.037	22
Gly	1.605	0.029	3040
His	0.739	0.070	111
Ile	3.624	0.029	15941
Leu	3.381	0.025	18042
Lys	0.192	0.038	25
Met	2.806	0.055	2568
Phe	3.393	0.035	9292
Pro	0.754	0.042	319
Ser	1.110	0.030	1340
Thr	0.880	0.030	870
Trp	2.815	0.078	1288
Tyr	2.302	0.059	1510
Val	3.317	0.025	16961

protein data set used for the training procedure (not shown). Both possibilities can produce less than ideal fit of preference function to experimental data, and are discussed in a recent paper [52].

3.2. Expected and predicted length distribution for transmembrane helical segments

The transmembrane segments (TMS) and transmembrane helical segments (TMH) are not necessarily identical in lengths. Our predicted TMH could be longer and could be shorter than usual length of TMS of 19-22 residues. Figure 2 illustrates in the form of histogram that expected lengths of TMS could also be different from expected 19-22 residues.

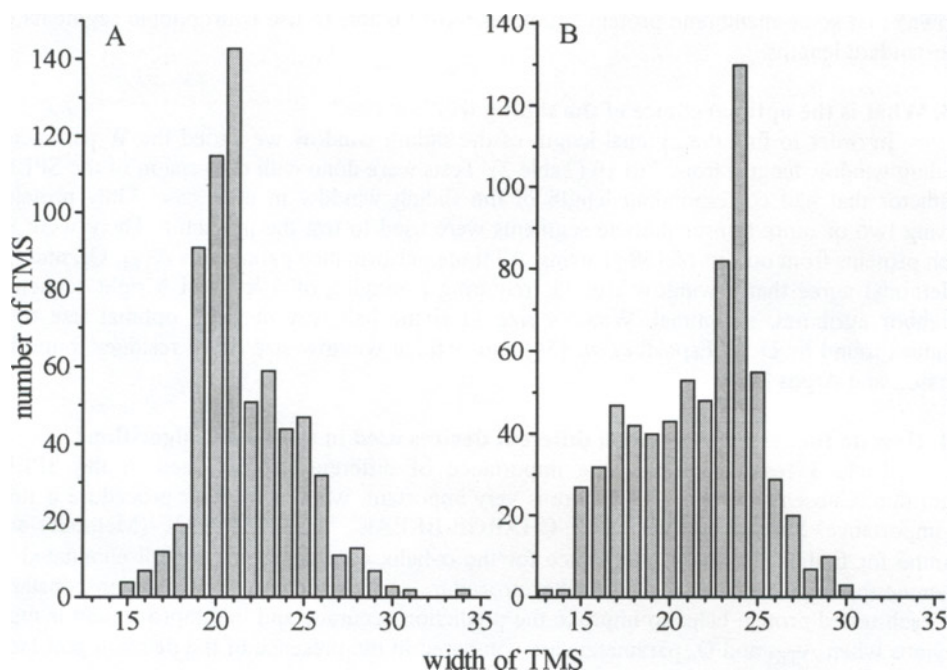


Figure 2: The length distribution of TMS in 168 proteins expected (Figure 2A) and predicted as TMH by us (Figure 2B). Two-times cross validation procedure (Methods) was used.

Both expected TMS and predicted TMH are often too short to span the membrane as α -helices or are so long that extramembrane parts in such segments must exist. Helical configurations other than α -helix should not be excluded for potential transmembrane segments [31]. For instance, it was pointed out [23] that 15 residue segment could span the bilayer as a 3_{10} helix. It is also possible that some transmembrane segments in α -class membrane proteins are in reality helical segments that pass through a part of membrane depth [53] or through whole membrane depth extending outside membrane or are in tilted orientation with respect to orthogonal direction from membrane surface. It appears from TMH length distribution in 10 integral membrane proteins of the best known structure too (not

Table 2
How results depend on the W parameter (sliding window length)^a

W	Q_s	A_{TM}	Q_p
7	94.0	0.689	57
9	95.2	0.700	64
11	95.5	0.704	67
13	94.7	0.697	58
15	93.1	0.694	59
17	93.0	0.693	57
19	92.1	0.675	57

^aPreference functions were extracted from the data base of 63 membrane proteins selected by Rost *et al.* [9] and 37 soluble proteins of the β -class (SOLB1, Methods) by using the PREF algorithm versions with sliding window length from 7 to 19 residues.

shown) that some membrane protein structures must be able to use hydrophobic segments of nonstandard lengths.

3.3. What is the optimal choice of the sliding window size?

In order to find the optimal length of the sliding window we varied the W parameter (sliding window length) from 7 to 19 (Table 2). Tests were done with the version of the SPLIT predictor that had corresponding length of the sliding window in each case. Only proteins having two or more transmembrane segments were used to test the predictor. There were 88 such proteins from our list of 168 proteins. All three performance parameters A_{TM} , Q_s , and Q_p (Methods) agree that a window size 11, requiring averaging of 5 left and 5 right sequence neighbor attributes, is optimal. Window size 11 is the half way between optimal size of 7 residues found by Degli Esposti *et al.* [54] and optimal window size of 15 residues found by Persson and Argos [55].

3.4. How do the results depend on different devices used in the SPLIT algorithm?

Table 3 results compare the importance of different devices used in the SPLIT algorithm. Chosen smoothing procedure is very important, while main filter procedure is next in importance. Subroutines 'FILTER', 'CHARGE-BREAK', 'TURN-BREAK' (Methods) and routine for finding maximum preference for the α -helix configuration were all eliminated to examine the importance of the main filter procedure. Automatic choice of decision constants for each tested protein helps to improve the prediction accuracy and the improvement is most obvious when A_{TM} and Q_p parameters are compared in the presence of the decision constants device (first row) and in its absence (fourth row).

Table 3
The dependence of prediction results on different devices used in the SPLIT algorithm^a

SPLIT algorithm	A_{TM}	Q_s	Q_p	# predicted TMH	# proteins with correct prediction
With no change	0.712	95.0	76.8	665	129
With no smoothing	0.646	87.0	57.8	613	97
Without main part of the filter	0.655	95.3	63.7	596	107
With all DC = 0	0.693	92.3	67.9	649	114
Without 'FILTER' subroutine	0.701	95.0	76.8	665	129
Without additional parts of the filter	0.705	94.1	74.4	659	125

^aEach device is separately eliminated from the algorithm before testing the prediction on the complete data set of 168 membrane proteins. The best Gaussian parameters file was obtained after 5-times cross validation procedure applied as described in the Methods section (the 'best' training procedure), but cross-validation was not performed. Refer to the Methods section for performance parameters.

'FILTER' subroutine alone seems to be important only in adjusting the positions of transmembrane helical caps. Additional parts of the filter, such as fusing short predicted helices, that may be the part of longer transmembrane helix, and extracting very short predicted helices with very high α -helix preference, are of minor importance. The Q_s parameter, or percentage of TMH that are correctly predicted, can be very misleading as the measure of prediction accuracy in the absence of a good filter, because it then increases together with the increase (overprediction) of residues predicted in the TMH conformation.

Table 4
Several scales of amino acid attributes used in this report

AA code	KYTDO ^a	MODKD ^b	CPREF ^c
Ala	1.8	1.10	0.6942
Arg	-4.5	-5.10	-1.4344
Asn	-3.5	-3.50	-0.7786
Asp	-3.5	-3.60	-1.1296
Cys	2.5	2.50	0.3427
Gln	-3.5	-3.68	-1.0870
Glu	-3.5	-3.20	-1.2480
Gly	-0.4	-0.64	-0.0549
His	-3.2	-3.20	-0.9697
Ile	4.5	4.50	1.7999
Leu	3.8	3.80	1.1403
Lys	-3.9	-4.11	-1.1850
Met	1.9	1.90	1.3557
Phe	2.8	2.80	1.3171
Pro	-1.6	-1.90	-0.5091
Ser	-0.8	-0.50	-0.2812
Thr	-0.7	-0.70	-0.2030
Trp	-0.9	-0.46	0.8475
Tyr	-1.3	-1.3	0.3693
Val	4.2	4.2	1.0138

a,b,c Scale acronyms are defined in Table 5.

3.5. What are the best scales of amino acid attributes?

As expected many different hydrophobicity scales are good predictors of transmembrane helical segments. The same scale is used during training and testing procedure. Each scale is normalized with average zero and standard deviation of one when called by the algorithm. As an example 20 values for the Kyte-Doolittle scale (KYTDO) are given together with modified Kyte-Doolittle scale (MODKD) and with normalized scale of constant preferences (CPREF) that were extracted from the reference data set of 168 membrane proteins (Table 4).

The list of 30 scales in Table 5 is our selection of the best predictor-scales from almost 100 scales that are available in the algorithm.

Table 5
Evaluation of hydrophobicity scales^a

Scale #	Acronym: Attribute	Reference	Performance parameters		
			A _{TM}	Q _s	Q _p
83	MODKD: Modified Kyte-Doolittle hydrophathy scale	This work (Table 4)	0.711	95.7	76.8
1	KYTDO: Hydrophathy values	[17]	0.704	95.9	78.6
100	CPREF: TMH preferences from training data base	This work (Table 4)	0.699	96.4	73.2
17	PONG1: Surrounding hydrophobicity scale	[47]	0.680	94.1	68.5
26	EISEN: Consensus hydrophobicity scale	[18]	0.675	95.0	67.8
9	VHEBL: Hydrophathy scale for membrane proteins	[56]	0.672	95.2	68.5
35	NNEIG: Self-consistent hydrophobicity scale	[50]	0.671	93.7	66.7
29	CHOTH: Proportion of residues 95 percent buried	[57]	0.670	93.8	66.7
30	ROSEF: Mean fractional area loss	[58]	0.666	93.7	63.1
52	EDE25: Optimal predictors for width 25	[22]	0.660	94.6	66.7
53	EDE21: Optimal predictors for width 21	[22]	0.660	94.6	65.5
4	ENGEL: Hydrophathy values	[59]	0.659	94.7	64.9

Continued on next page.

Table 5 - Continued

49	HEIJN: Hydrophobicity scale for TMS	[60]	0.658	94.4	63.7
71	GRANT: Polarity scale	[61]	0.658	93.2	62.5
44	DEBER: M/A ratio in membrane transport proteins	[62]	0.656	93.2	62.5
7	GUY-M: Average of four hydrophobicity scales	[63]	0.655	92.1	64.3
70	WOESE: Polarity scale	[64]	0.652	94.0	61.9
3	PONNU: Surrounding hydrophobicity scale	[65]	0.652	91.5	59.5
8	KRIGK: Ethanol to H ₂ O hydrophobicity scale	[66]	0.650	93.1	60.1
28	HOPPW: Antigenic determinant scale	[67]	0.649	94.1	63.1
5	JANIN: Free energy of transfer from protein interior	[68]	0.645	92.9	61.3
16	CIDAB: Hydrophobicity scale for proteins of α/β class	[69]	0.645	90.5	63.7
31	GUYFE: Transfer free energy for 6 layers in proteins	[63]	0.645	91.4	61.9
42	MIJER: Average contact energy	[70]	0.645	91.2	61.9
12	GIBRA: Solvent accessibility in proteins	[71]	0.645	92.1	60.1
2	FAUPL: Solution hydrophobicities	[72]	0.643	92.6	60.1
19	PONG3: Combined membrane hydrophobicity scale	[47]	0.642	94.9	63.7
27	PRIFT: Statistical scale for amphipathic helices	[50]	0.642	93.2	61.3
21	ROSEM: Self-solvation free-energy changes	[73]	0.639	92.1	61.9
78	CASSI: Structure-derived hydrophobicity scale	[74]	0.635	93.4	58.3

^aFor a chosen scale of amino acid attributes each of 168 membrane proteins was tested once without being used in the training procedure as described in the Methods section.

3.6. The prediction results with Kyte-Doolittle preference functions

Full details of prediction results for each of 168 reference membrane proteins are enclosed in the Supplementary Material (Table IV). We used cross-validation (5-fold, Methods) and the KYTDO scale (# 1). All of 168 proteins were correctly predicted as membrane proteins having at least one transmembrane segment. With 100% correct transmembrane topology 130 proteins were predicted. A total of 631 transmembrane helices were correctly predicted out of a total number of 662 expected transmembrane segments. Only 36 TMH were overpredicted and 31 underpredicted. Of individual residues in TMH configuration 12273 out of 14374 were correctly predicted, 2033 overpredicted and 2101 underpredicted. The performance parameters (Methods) are then: $A_{TM} = 0.712$, $Q_s = 95.3\%$, $Q_p = 77.4\%$, $A_s = 0.898$.

As an example of complete information provided by the predictor the predicted preference profiles and hydrophobic moment profiles for the *gef_ecoli* protein (outside reference list of 168 proteins and without assigned transmembrane domain in the SWISS-PROT data base) are given in Table 6 as unmodified output file. The *gef* protein can stimulate cell killing [75] after overexpression and oligomerization in the membrane environment. In addition to predicted α -helix transmembrane segment from residues 6 to 24 there is also the 31-46 segment predicted in the β -strand conformation. The 31 to 45 segment may be another potential membrane-embedded segment possibly involved in dimerization or oligomerization process in the membrane environment that can lead to cell killing.

Table 6

Complete prediction results for the *gef_ecoli* protein by using the Kyte-Doolittle hydropathy scale through preference functions^a

	AA	PS	PTM	PH	PB	PT	PU	MA	MB	H-T	PB+MB-2
1	M	C	U	0.00	0.00	0.33	1.58	ND	ND	-0.33	ND
2	K	C	U	0.02	0.29	0.62	1.49	ND	ND	-0.60	ND
3	Q	C	U	0.06	0.54	0.81	1.42	1.38	1.08	-0.75	-0.38
4	H	C	U	0.48	0.79	1.06	1.33	1.21	0.33	-0.57	-0.88
5	K	C	T	1.02	0.82	1.30	1.12	1.07	0.21	-0.27	-0.97
6	A	H	M	1.66	0.88	1.31	0.89	1.26	0.39	0.35	-0.73
7	M	H	O	2.32	0.74	0.86	0.60	1.17	0.34	1.47	-0.91
8	I	H	O	3.00	0.53	0.49	0.32	0.61	1.00	2.51	-0.47
9	V	H	O	3.65	0.28	0.19	0.10	0.53	0.52	3.46	-1.19
10	A	H	O	4.17	0.11	0.08	0.02	0.79	0.60	4.10	-1.29
11	L	H	O	4.57	0.02	0.04	0.01	0.73	0.74	4.53	-1.23
12	I	H	O	4.71	0.01	0.02	0.00	0.92	0.92	4.69	-1.07
13	V	H	O	4.75	0.01	0.02	0.00	1.07	0.59	4.73	-1.41
14	I	H	O	4.75	0.01	0.02	0.00	1.11	0.60	4.73	-1.39
15	C	H	O	4.74	0.01	0.02	0.00	0.97	0.56	4.72	-1.43
16	I	H	O	4.74	0.01	0.05	0.01	0.96	0.56	4.68	-1.43
17	T	H	O	4.72	0.01	0.06	0.01	1.11	0.38	4.66	-1.61
18	A	H	O	4.64	0.01	0.06	0.01	1.22	0.55	4.57	-1.44
19	V	H	O	4.33	0.06	0.05	0.01	1.07	0.48	4.28	-1.46

Continued on next page.

Table 6 - Continued

20	V	H	O	3.93	0.21	0.15	0.05	1.67	0.14	3.78	-1.65
21	A	H	O	3.58	0.43	0.44	0.21	1.72	0.16	3.13	-1.42
22	A	H	O	3.15	0.74	0.72	0.42	1.05	0.55	2.43	-0.71
23	L	H	O	2.50	1.07	0.85	0.59	1.17	0.48	1.65	-0.45
24	V	H	M	1.88	1.21	0.94	0.65	0.99	0.48	0.94	-0.30
25	T	C	T	1.31	1.26	1.34	0.76	0.97	0.44	-0.03	-0.30
26	R	C	T	0.98	1.20	1.77	0.90	1.02	0.23	-0.78	-0.57
27	K	C	T	0.71	1.19	1.87	1.07	0.81	0.52	-1.16	-0.29
28	D	C	T	0.41	1.12	1.46	1.14	0.83	0.51	-1.05	-0.37
29	L	C	U	0.18	1.05	1.03	1.26	0.65	0.34	-0.85	-0.61
30	C	C	U	0.18	1.26	0.77	1.34	0.95	0.59	-0.59	-0.15
31	E	B	E	0.17	1.37	0.76	1.35	0.89	0.63	-0.58	0.01
32	V	B	E	0.16	1.43	0.91	1.29	0.77	0.68	-0.75	0.12
33	H	B	E	0.11	1.27	0.92	1.24	0.83	0.77	-0.81	0.04
34	I	B	E	0.14	1.46	1.07	1.25	0.70	0.74	-0.93	0.20
35	R	C	E	0.16	1.28	1.02	1.28	0.45	1.27	-0.86	0.55
36	T	B	E	0.24	1.33	1.30	1.24	0.52	1.34	-1.06	0.67
37	G	B	E	0.23	1.40	1.29	1.19	0.59	1.16	-1.06	0.55
38	Q	B	E	0.31	1.44	1.39	1.08	0.57	1.10	-1.08	0.54
39	T	B	E	0.44	1.63	1.21	1.14	0.79	0.91	-0.77	0.54
40	E	B	E	0.49	1.64	1.10	1.07	0.53	0.91	-0.61	0.55
41	V	B	E	0.64	1.83	1.07	1.06	0.64	0.86	-0.43	0.69
42	A	B	E	0.65	1.77	1.02	1.00	0.95	0.55	-0.36	0.32
43	V	B	E	0.63	2.03	1.06	0.96	0.90	0.44	-0.44	0.47
44	F	B	E	0.63	1.78	1.08	1.00	0.71	0.36	-0.45	0.14
45	T	B	E	0.52	1.97	1.17	1.01	0.65	0.33	-0.65	0.30
46	A	B	B	0.38	1.55	1.17	1.12	1.16	0.40	-0.79	-0.05
47	Y	C	E	0.34	1.20	0.95	1.25	1.09	0.95	-0.61	0.15
48	E	C	U	0.15	0.81	0.74	1.39	0.61	1.15	-0.59	-0.04
49	S	C	U	0.03	0.22	0.52	1.51	ND	ND	-0.49	ND
50	E	C	U	0.04	0.16	0.39	1.54	ND	ND	-0.36	ND

^aOne letter amino acid codes are used in the second column (AA). Predicted structure (PS) in the third column can be α -helix (H), β -sheet (B) or coil (C) structure that includes turn and undefined structure. Residues predicted in the transmembrane helix configuration (PTM) in the fourth column are labeled with letter 'M' except for highly probable TMH conformation when letter 'O' is used. Residues with a potential to form transmembrane β -strands are labeled with letter 'E' in the fourth column. The coil (C) conformation from third column is specified as undefined (U) or turn (T) conformation in the fourth column. Fifth to eighth column contain smoothed preferences for α -helix (PH), β -sheet (PB), turn (PT) and undefined (PU) conformation. The columns 9 and 10 contain numerical values for hydrophobic moments calculated in the case of assumed α -helix configuration (MA) and for moments calculated for assumed β -sheet configuration (MB). Last two columns contain PH-PT difference of preferences (H-T) that helps in visual identification of predicted transmembrane helices and PB+MB-2.0 scores that help in prediction of potential membrane-embedded β -strands.

Since interaction of transmembrane helices is not directly taken into account by us it is possible that our prediction for proteins expected to have large number of transmembrane helices systematically err on the side of underprediction. One such example may be the calcium channel subunit *cic1_cypca* in which fourth and tenth potential transmembrane segments are not predicted. Another such example is the human erythrocyte anion exchanger *b3at_human* in which our prediction of 13 transmembrane helices is associated with one underpredicted TMH (residues 460-479) according to the 14 TMH topological model of Wang *et al.* [76]. Underpredicted segment has three Glu residues and not enough high preference peak, so that it is rejected by algorithm's filter procedure, but can be recognized from preference profile (not shown) as potential TMH segment. Earlier models for a monomer of the Band 3 dimer [77] predicted only 12 membrane-spanning α -helices, but one of the authors [53] later observed that 'inner' helices can be easily overlooked when sufficiently long hydrophobic segments are sought, since such helices can span the membrane only partially and without direct contact with lipid environments.

Binding of ligands or cofactors is not taken into account too, but it can conceivably change the potential for formation of regular secondary structure for sequence segment that interacts with a ligand or cofactor. Underprediction was seen for the thromboxane A synthase (*thas_human*), a member of the P-450 family, which probably binds heme-thiolate at the position 479. The fifth transmembrane segment, that is underpredicted both by us and Rost *et al.* [9], starts with residue 480 in the thromboxane A synthase topological model reported by the SWISS-PROT data base.

Gross errors in the topological models adopted by the authors and by the SWISS-PROT or some other data base can be easily detected by our algorithm. We have very strong prediction of three transmembrane helical segments (14-36, 139-160 and 166-189) for the TOLQ protein from *Escherichia coli*. Only the first TMS from residues 23 to 43 is correctly predicted according to the SWISS-PROT assignment of the bitopic transmembrane topology for that protein. Interestingly, very similar protein *exbb_ecoli* has SWISS-PROT release 29 assignment of three transmembrane segments too. Two commonly used methods for predicting transmembrane helices, that of Eisenberg *et al.* [18] and that of Rao and Argos [20] also predict three transmembrane segments for *tolq_ecoli*, while Rost *et al.* method [9] predicts four transmembrane segments for that protein. Small number of homologues for that protein (only 3) and a need to filter predicted 'transmembrane segment' having 66 residues is the likely cause for the proposed four helix model by the automatic *E-mail* service of Rost *et al.* [9].

3.7. Testing for false positive predictions in membrane and soluble proteins of crystallographically known structure

Ten integral membrane proteins of well known structure (BESTP, Methods) have been tested first. Only the Kyte-Doolittle and our modification of the Kyte-Doolittle scale (MODKD, # 83) were able to predict all of these ten membrane proteins with 100% correct transmembrane topology, i.e. all transmembrane helices were correctly predicted at their observed sequence locations and there were no overpredicted TMH (Table 7). Only the Chothia buried surface scale (CHOTH, # 29) did not recognize one of ten membrane proteins as the membrane protein (the subunit H from the photosynthetic reaction center from *R. viridis*). Nine long extramembrane helices in these 10 proteins were not predicted as TMH by any of 12 tested amino acid scales. That these sensitive tests of our predictor do not depend on the chosen training procedure was checked by using different training procedures. After

training the algorithm on 63 proteins selected by Rost *et al.* [9] or on 105 proteins selected by us with the addition of 37 β -class soluble proteins (SOLB1) the results were very similar (not shown). Another sensitive test was made possible when the crystal structure of cytochrome c oxidase from *Paracoccus denitrificans* [14] became known during work on this report.

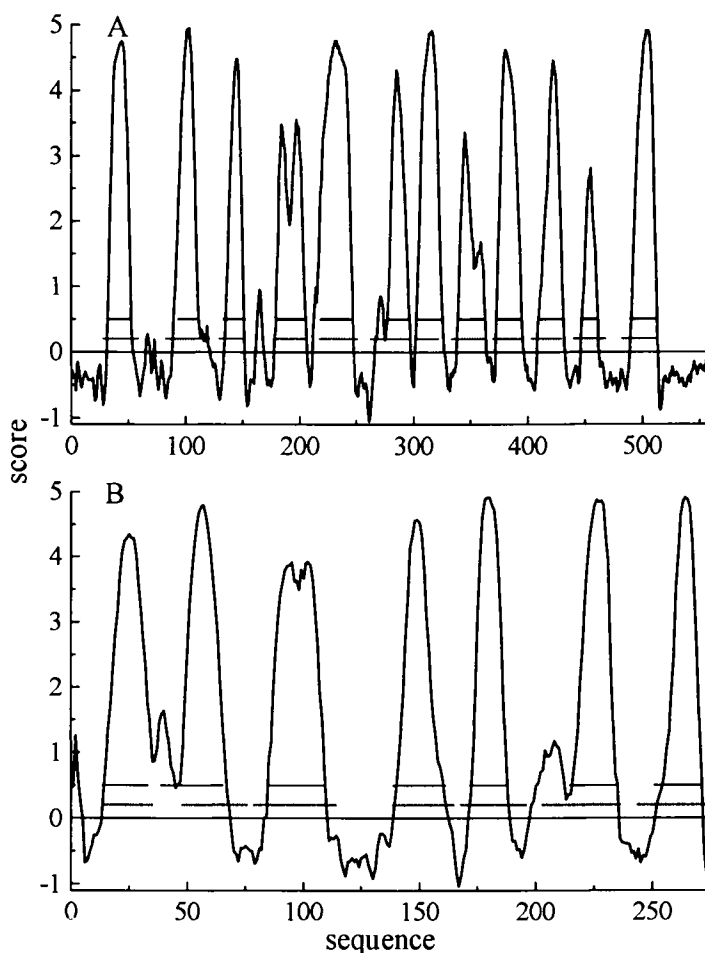


Figure 3: Score profiles for *cx1b_parde* (Figure 3A) and for *cox3_parde* (Figure 3B) of cytochrome oxidase from *Paracoccus denitrificans* [14] are obtained by subtraction of turn preferences from α -helix preferences (full line). Digital predictions, as outcome of the best training procedure for the SPLIT algorithm with Kyte-Doolittle hydropathy scale (Methods), are shown as bold horizontal bars at the score level 0.5. Observed location of TMH segments are shown as bold horizontal bars at the score level 0.2.

With our best file of Gaussian parameters ('best' training procedure, Methods) we correctly predicted all of 12 TMH in *cx1b_parde* (Figure 3A) and all of 7 TMH in *cox3_parde* (Figure 3B) without single overpredicted TMH. Subunit IV was not tested, while in the *cox2_parde* two TMH were predicted correctly and two overpredicted. Predicted 'TMH' at residues 12 to 30 is the signal sequence. Predicted 'TMH' at residues 192 to 216 has atypical flat profile with maximum height less than half of other peaks. Three observed β -strands: 190-194, 200-204 and 209-216 at that position are not seen by the algorithm when it makes the automatic choice of decision constants (Methods) such that β -structure is depressed. Setting all decision constants to zero eliminated this erroneous TMH prediction. When these 3 polipeptides are added to 10 considered above, the total score is 49 correctly predicted TMH out of the total number of 49 observed TMH (not counting the signal sequence), with one, easy to detect, overprediction. This result did not change in the case when 105 integral membrane proteins were used to train the algorithm and to extract corresponding file with Gaussian parameters, but one TMH was overpredicted in the *cox3* at sequence segment 198-213 when 63 or all of 168 proteins were used in the training process. Setting all decision constants to zero eliminated this overprediction as well in both cases. Standard training procedure with the MODKD scale (Table 4) produced 100% correct topology for subunits *cx1b* and *cox3* and the same two overpredicted TMH in the *cox2*.

Table 7

Test of best 12 amino acid attributes in predicting TMH in membrane proteins of known structure^a

Scale #	A _{TM}	# correct TMH pred.	# predicted TMH	# correct M. P. pred.	# predicted M. P.
1	0.695	28	28	10	10
83	0.693	28	28	10	10
52	0.682	27	28	8	10
53	0.679	27	28	8	10
29	0.676	27	27	9	9
17	0.644	27	30	7	10
35	0.626	27	29	7	10
100	0.616	27	29	7	10
30	0.603	27	31	6	10
4	0.547	27	31	5	10
9	0.527	27	32	5	10
26	0.523	27	32	6	10

^aTested proteins (BESTP, Methods) had 28 observed TMH with 717 residues in the TMH conformation. Standard training procedure was used with each choice of amino acid attribute. Code numbers for amino acid scales are listed in Table 5.

By using our standard training procedure the tests were performed on membrane proteins of known or partially known structure with transmembrane β -strands and on soluble proteins of known structure. For seven tested porins and two defensins (PORINS, Methods) we tested 12 best scales used in Table 7. Only the scale # 4 [59] predicted one transmembrane segment in the α -helix conformation (residues 119 to 133 in the porin sequence from *Rhodobacter capsulatus*). For two different sets of soluble proteins SOLU1 and SOLU2 (Methods) prediction results are collected in Table 8 as percentage of proteins falsely predicted to be membrane proteins. The best scales for TMH prediction in membrane proteins still falsely predicted 11-12% of soluble proteins as being membrane proteins with at least one transmembrane helix.

Table 8
The prediction performance of 12 best amino acid scales (Table 5) on soluble proteins^a

Scale #	SOLU1 ^b	SOLU2 ^b
17	11.2	12.2
53	11.2	12.2
30	11.2	13.6
52	11.2	13.6
83	12.8	13.6
100	13.4	14.3
35	15.0	16.3
1	13.9	17.7
29	19.3	17.7
9	21.4	19.0
26	20.3	19.7
4	25.7	23.8

^aOnly the percentage of proteins predicted with one or more transmembrane helices is reported. Code numbers for amino acid scales are listed in Table 5.

^bData base of soluble proteins of known structures (see Methods).

3.8. Cross-validation, overtraining and sensitivity to the choice of protein data base

After standard training procedure tests were performed separately on the subsets of 80 proteins having only one observed TMH and 88 proteins having more than one TMH. All performance parameters registered higher prediction accuracy for 80 proteins having only one transmembrane segment. The best result of $A_{TM} = 0.778$, $Q_s = 97.5\%$ and $Q_p = 92.5\%$ was achieved in the 2 times cross-validation procedure when training was done on 88 proteins having more than one TMH and 37 soluble proteins of β -class. Interestingly, training and testing on the same data set of 80 membrane proteins (with 37 soluble β -class proteins included as always in the training procedure) produced huge overprediction of predicted TMH and very poor performance parameters $A_{TM} = 0.285$ and $Q_p = 52.5\%$. The percentage of accurately predicted transmembrane helices remained the same: $Q_s = 97.5\%$ or 78 correctly predicted TMH of 80 observed, but total number of predicted TMH jumped from 84 to 139, while number of overpredicted TMH jumped from 6 to 61! Commonly used Q_s parameter gives obviously wrong picture of the prediction performance in this case. More surprising result is such extreme advantage of cross-validation procedure versus training and testing on the same data set of integral membrane proteins. Slight advantage of the cross-validation procedure is seen too when all of 168 reference proteins are used for training and for testing (compare performance parameters in the first two rows of Table 9). Needless to say, we always expect a

decrease in the prediction performance when training is no longer performed on the same data set that is used for testing procedure.

Table 9
Different training procedures^a

	A_{TM}	Q_s	A_s	Q_p	# prot tested
a) Five-times cross validation (Supplementary Material Table IV).	0.712	95.3	0.898	77.4	168
b) No cross-validation. All of 168 membrane proteins used to train and to test.	0.709	94.7	0.891	76.2	168
c) No cross-validation. Best training procedure (Methods).	0.712	95.0	0.896	76.8	168
d) Two-times cross validation: 63 proteins to train and 105 to test and vice versa.	0.704	95.9	0.903	78.6	168
e) Train on 105 proteins. Test on 63.	0.740	97.9	0.934	84.1	63
f) Train on 63 proteins. Test on 105.	0.682	94.7	0.885	75.2	105
g) Train on 105 proteins. Test on 105.	0.693	94.0	0.878	73.3	105
h) Train on 63 proteins. Test on 63.	0.737	97.5	0.905	76.2	63
i) No cross-validation. Soluble proteins SOLB2 used instead of SOLB1 during training procedure.	0.705	94.4	0.890	74.4	168

^aThe Kyte-Doolittle scale is used in each case. See Methods for performance parameters.

The clue is offered by such training procedure when only 16 residues next to each side of a transmembrane segment are used to extract sequence environments. Then it becomes possible to use 80 proteins having single TMH both for training and for testing and to obtain high performance parameters: $A_{TM} = 0.777$, $Q_s = 97.5\%$ and $Q_p = 92.5\%$. It would seem that dominant contribution of sequence environments from extramembrane parts of membrane proteins with single TMH must be reduced if balanced training is to be achieved. This can be done either directly by omitting residues from the training process that are far removed from expected transmembrane segments or indirectly by choosing the training data base of membrane proteins with balanced contribution of residues in transmembrane and in extramembrane positions.

That need for balanced training is not the whole explanation becomes clear when the PREF algorithm is modified in such a way that is always collects exactly the same number of environments associated with different secondary structure motifs. Again poor prediction results are obtained when bitopic proteins are used for training and for testing (not shown). When all of 168 membrane and 37 soluble proteins are used in a balanced training procedure prediction results for 168 proteins remain similar for the TMH prediction ($A_{TM} = 0.702$), but

overall prediction of all secondary structures is dramatically improved ($Q_3 = 0.775$) meaning that turn and undefined residues are much better predicted.

The extraction of preference functions, as the training procedure, is not a very powerful training procedure and it is not expected to lead to overtraining. We shall test this assumption by performing still another two-times cross-validation test in which 168 membrane proteins are divided into 63 proteins used by Rost *et al.* [9] and 105 proteins used by us. Table 9 lists performance results for different combinations of training and testing procedures.

Table 9 results indicate that extraction of preference functions as the part of the training procedure does not lead to overtraining, because training on an independent set of unrelated proteins can produce even better results. It is still possible that either automatic or subjective choice of filter parameters leads to overtraining. All our filter parameters were trained on the subset of 63 proteins and with the choice of the sliding window length of $W = 9$ residues. A drop in prediction accuracy when $W = 11$ (window length used in all presented results) is used, for the same subset of 63 proteins, was indeed observed (not shown). Since $W = 11$ seems to be optimal for much larger group of transmembrane segments (Table 2) it is indeed possible to increase apparent prediction accuracy by variation of filter parameters. To avoid such a danger we did not try to optimize filter parameters for a final choice of sliding window length ($W = 11$) and protein data base (168 proteins).

Having a larger reference set of nonhomologous proteins for extracting preference functions will not increase prediction accuracy. Safe lower limit is difficult to estimate, but is probably no more than 30-40 such proteins. In terms of residues considered for extraction of preference functions only about 4500 residues were enough to achieve very high prediction accuracy ($A_{TM} = 0.777$) in the case of bitopic (single-span) membrane proteins. A different set of soluble proteins in the training list of proteins may change slightly the prediction performance (last row in Table 9).

3.9. Comparisons with other methods

An automated FTP service was used to obtain the predictions for all of our 168 integral membrane proteins by using the Rost *et al.* method [9]. A total of 11870 residues were correctly predicted in the TMH conformations, 2436 residues were overpredicted, 2512 residues were underpredicted, while 50335 residues were correctly predicted not to be in the TMH conformation. One of many different performance parameter that can be constructed by using these data is the A_{TM} parameter (Methods). Its value is $A_{TM} = 0.656$, which is inferior to our value of 0.712 (Table 9) for the same parameter. However, when tested on the subset of 63 proteins used by Rost *et al.* [9] the A_{TM} parameter, calculated from predictions returned by automated service, becomes 0.733, which is comparable to our value of $A_{TM} = 0.740$ for the same subset of proteins (Table 9). Similar test on the subset of 105 proteins, never before seen in the training process for the neural network algorithm, gave quite a low value of $A_{TM} = 0.610$ for the Rost *et al.* method [9]. That value is lower than our value of $A_{TM} = 0.682$ for the same subset of 105 proteins (Table 9). All of 63 proteins selected by Rost *et al.* [9] are also predicted as membrane proteins, but their method does not recognize 2 out of 105 membrane proteins selected by us. Underprediction of membrane proteins is due to serious underprediction of transmembrane helices: 50 of observed 419 TMH are underpredicted and 11 overpredicted by Rost *et al.* [9]. For comparison our Table 9 results (row f) for A_s are obtained for the case of 21 underpredicted and 25 overpredicted TMH in the same test set of 105 proteins.

The prediction results for three commonly used prediction methods: that of Rao and Argos [20], that of Eisenberg *et al.* [18], and that of Rost *et al.* [9] can be compared with our results listed in Table 7 for the data set of 10 best known membrane proteins with observed 717 residues in the membrane-spanning helix conformation (Supplementary Material, Table V). Eisenberg's algorithm overpredicts 5 helices in the subunits M, and L from the photosynthetic reaction center, and has correspondingly low performance parameter for all ten proteins: $A_{TM} = 0.470$ (195 underpredicted and 185 overpredicted residues). Rao and Argos algorithm [20] has better performance of $A_{TM} = 0.562$, but large number of residues (314) is still underpredicted or overpredicted. Rost *et al.* neural network algorithm [9] used some subunits of the photosynthetic reaction center for training and achieved a much better result: $A_{TM} = 0.702$ with 108 underpredicted and 106 overpredicted residues. Only one helix was underpredicted (the N-terminal transmembrane helix from the light harvesting center).

Residues in three transmembrane helices of the LHC-II are underpredicted by all four methods, the likely reason being increase in helix preference due to binding of chlorophylls which is not taken into account by these methods. The whole first helix is underpredicted in Rost *et al.* [9] and Rao and Argos method [20]. It may seem strange that Rost *et al.* method [9] can predict 7 of 35 residues in the first transmembrane helix of the cb21_pea protein, but cannot repeat even such partial success when shorter but otherwise identical LHC-II polypeptide is tested. Filter elimination of signal sequences (see Discussion) and/or too short potential transmembrane segments in the Rost *et al.* [9] procedure becomes critical when polypeptides lacking complete N-terminal section in front of a potential TMS are considered.

Our result $A_{TM} = 0.695$ (Table 7, Kyte-Doolittle scale) for all of 10 membrane proteins becomes $A_{TM} = 0.714$ (56 underpredicted and 23 overpredicted residues, all 11 TMH correctly predicted) when only H, M and L subunits of the photosynthetic reaction center from *Rhodospseudomonas viridis* are considered. This can be compared with Fasman and Gilbert [78], and Ponnuswamy and Gromiha [47] evaluation of many different methods for predicting transmembrane helices when these same three polypeptides are used as very restricted 'standard of truth'. The Kyte-Doolittle [17], Sieved Kyte-Doolittle [21] and Klein-Kanehisa-DeLisi procedure [19] are associated with prediction accuracy lower than $A_{TM} = 0.7$, while von Heijne [79], Engelman-Steitz-Goldman [59], Esposti-Crimi-Ventruoli [54], and Ponnuswamy-Gromiha procedure [47] are associated with higher prediction accuracy.

We have also compared two powerful prediction methods, that of Jones *et al.* [33] and Rost *et al.* [9] with our own (JLT) by testing greater number of proteins whose expected transmembrane structure is taken from the SWISS-PROT data base. For 83 proteins used by Jones *et al.* [33] one can extract A_s and Q_p performance parameters as $A_s = 0.928$ and $Q_p = 79.5\%$. For 69 proteins tested by Rost *et al.* [9] A_s and Q_p parameters are 0.896 and 79.7%, respectively. For 63 proteins tested by us these parameters are $A_s = 0.934$ and $Q_p = 84.1\%$ (Table 9).

Overprediction of transmembrane segments in large eukaryotic proteins having single transmembrane anchoring segment is common deficiency of many prediction methods [33]. Our algorithm overpredicts six and underpredicts two transmembrane segments in the data base of 80 membrane proteins expected to have single transmembrane helix. For instance, in the case of epidermal growth factor receptor precursor: egfr_human Jones *et al.* [33] overpredicts two transmembrane segments. Our method adds to correct prediction of the segment 646 to 668 an incorrect prediction for residues 777 to 798. Rost *et al.* prediction 648-666 without overpredicted segments is even better [9]. The price paid for reduced overprediction in single-

span proteins is seen much better when Rost *et al.* method [9] is tested on never-before-seen data set of 105 membrane proteins containing 48 single-span proteins. Then two proteins: *fish_ecoli* and *spir_spime* are not predicted as membrane proteins, because Rost *et al.* [9] do not find a single transmembrane segment in these proteins. Cell division protein *fish* is strongly predicted with two transmembrane helices at correct sequence location by our method (Supplementary Material, Table IV). Spiralin is predicted by us as membrane protein, but with transmembrane segment at the N-terminal (residues 3 to 21) instead from residues 165 to 184 (SWISS-PROT assignment).

Underpredictions of the last transmembrane segment in G-protein coupled receptors with seven transmembrane segments are also commonly seen by our and other methods [33]. This is the case with *a2aa_human*, *aa2a_canfa*, *acm5_human*, *car1_dicdi*, *ops1_calvi* and *ops2_drome* for our prediction. The seventh helix in the superfamily of seven-helix protein G-coupled receptors contains retinyl-lysine in the case of opsins or may be adjacent to a potential acylation site [80]. As a rule it can be recognized for potential TMH from preference profile as the last of 7 sharp peaks (often with characteristic minimum pointing at sequence position of functionally important lysine residue) even if the digital version of the predictor cannot predict it.

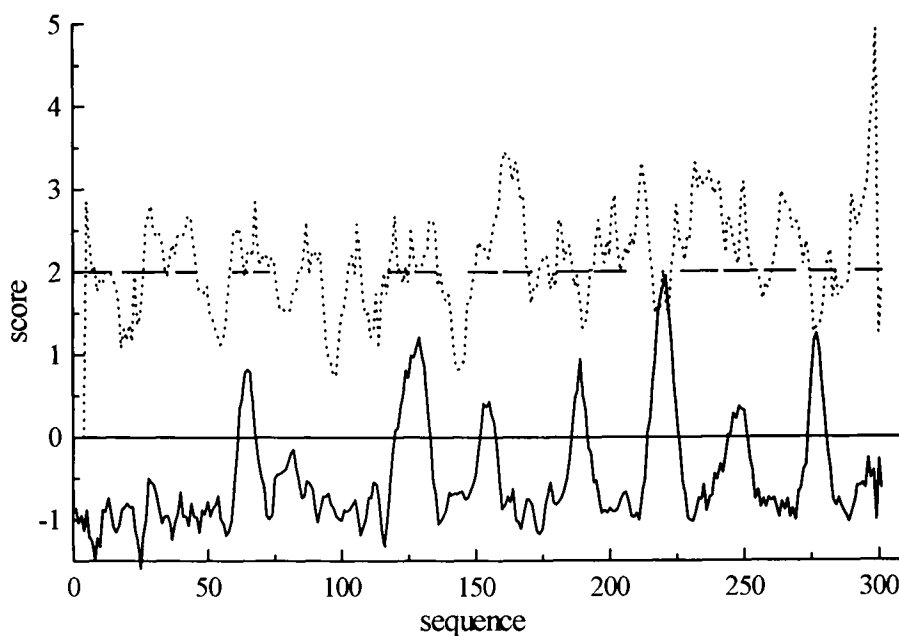


Figure 4: Score profiles for porin from *Rhodobacter capsulatus* are obtained by subtraction of turn preferences from helical preferences (full line) and as sum of β -sheet preferences and hydrophobic moment scores for assumed β -sheet conformation (dotted line). Kyte-Doolittle scale [17] is used to calculate preferences, while PRIFT scale [50] is used to calculate hydrophobic moments. Observed transmembrane strands are shown as bold horizontal bars at the score level 2.0.

3.10. Using prediction profiles with both α and β motifs

Unrealistic initial assumption that only α -helices exist as transmembrane polypeptide structure can be tested by using predictions for membrane or surface attached β -strands (Methods) as well. All tests in this section are done with decision constants fixed at zero. Previously unseen possibilities for β -strand formation in the membrane environment become apparent from profiles of summed β -preferences and β -hydrophobic moments (Figure 4 and 5). Dotted line in Figures 4 and 5 can be regarded as the score profile for potential formation of membrane-buried or membrane attached β -strands ('E' structure in the fourth column of Table 6). As before we used the Kyte-Doolittle scale for preference calculation and the PRIFT scale to find hydrophobic moments for assumed β -structure. Revealed potential for the β -structure formation in the membrane is quite robust with respect to the change in the choice of hydrophobicity scale for preference calculation, notwithstanding the complexity of the scores profile. Above mentioned combination of scales predicted correctly 79% of membrane-buried β -strand residues in 9 membrane β -class proteins (PORINS, Methods), 72% of such residues in three best known porins (porin from *R. capsulatus* OmpF and PhoE) and 76% of such residues in the *R. capsulatus* porin. When algorithm is allowed to make its own choice of decision constants these percentages raise to 87, 82 and 76 respectively. Only one membrane-embedded β -strand (the 15-th) is underpredicted in the *R. capsulatus* porin (Figure 4), but there are two pairs of strands that are fused in our prediction. For three best known porins 7 β -strands are underpredicted, 4 overpredicted and 7 pairs of strands are predicted fused.

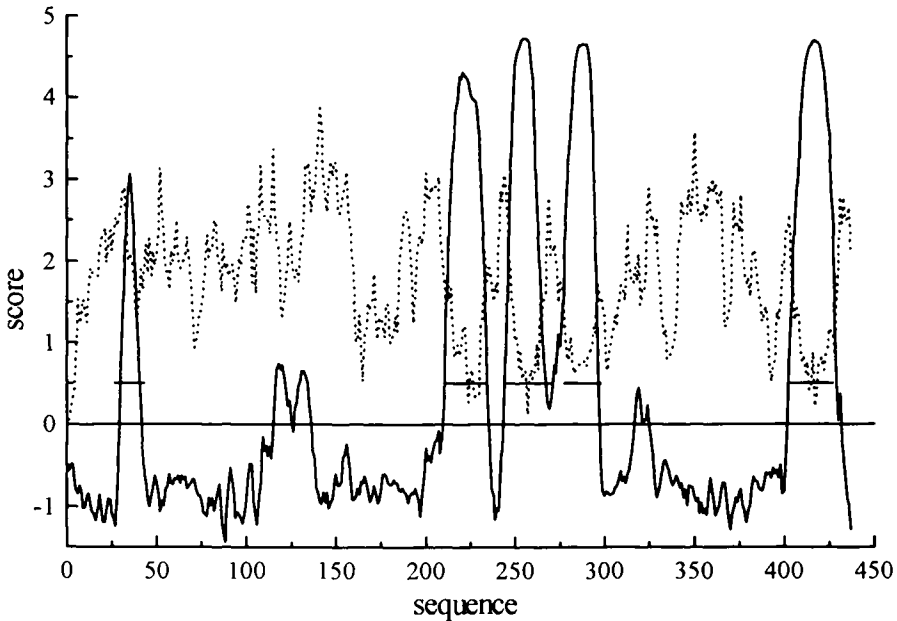


Figure 5: Nicotinic acetylcholine receptor ach1_xenla profiles for finding potential transmembrane α -helices and β -strands. Same conditions and same notation is used as for the Figure 4. Predicted transmembrane α -helices are shown as bold horizontal lines at the score level 0.5.

Underprediction of transmembrane helices, according to the SWISS-PROT reference standard, was very serious but variable, for proteins belonging to mitochondrial carrier family that are all expected to have six transmembrane helices [81-87]. The digitalization process in the algorithm, that decides whether given segment is in the TMH conformation or not, is the cause of instability in prediction performance for borderline cases. Preference and hydrophobic moment profiles contain considerably more information about the arrangement of potential transmembrane segments. Preference profiles (not shown) for the *adt2_yeast*, *adt1_bovin*, *adt_neucr*, *mcp_r*, *ucp_r*, *m2om_bovine*, and *txtp_r* agree that no more than 3 to 5 transmembrane helices can be predicted for each mitochondrial carrier protein and that second expected transmembrane helix can never be predicted by using our method. Accepted topological model for these proteins in the NBRF data base is three hydrophobic transmembrane α -helices for the brown fat uncoupling protein and phosphate carrier [88]. A β -strand that spans the membrane or three β -hairpins have been proposed for the adenine nucleotide translocator [89,90,83]. All mitochondrial carrier proteins have a tripartite structure, with three similar repeats about 100 residues each [82,91]. Our prediction profiles often better exhibit the tripartite symmetry for the profile of potential membrane attached or transmembrane β -strands than for predicted transmembrane α -helices (not shown).

The question of how many TMH segments are in the nicotinic acetylcholine receptor subunits has been going on for a number of years [92]. Earlier reviews [93] supported the four-TMH model. The possible existence of a scaffold of membrane associated β -strands supporting smaller number of transmembrane helices (may be only one) has been raised after low resolution electron microscopy studies [29]. One recent review [30] concludes that of four proposed TMH: M1, M2, M3 and M4 only M2 and M4 are the TMH while M1 and M3 most probably form β -structures. M2, M3 and M4 are α -helical according to Blanton and Cohen [92]. Our TMH predictor strongly predicts all of M1 to M4 segments as TMH segments in the *ach1_xenla* (Figure 5). High potential (dotted line) for the formation of membrane-buried β -strands is found in sequence domains 101-116, 138-159 and 341-364. Predicted percentage of α -helix transmembrane configuration (24%) is less than 34% [94] or 44.5% [95] suggested by circular dichroism experiments for the whole protein α -helix conformation, but similar to 25% suggested recently by hydrogen/deuterium exchange experiments [96]. Observed percentage of β -sheet residues (29% reported by Moore *et al.* [94], 34% reported by Chang *et al.*, [97]) is higher than predicted 96 residues (22%) in the potential membrane-embedded β -sheet conformation by our TMBS predictor. Observed percentage of β structures in the transmembrane domains alone (40% if β -turns are included according to Görme-Tschelnokow *et al.* [98]) is probably enough for the formation of six membrane buried β -strands in the presence of four transmembrane helices. Potential transmembrane sequence segments of α and β type, that are predicted by our algorithm, must be able to form novel combination of transmembrane regular structure.

Membrane import machinery protein *mas6_yeast* was predicted with a maximum of only two short transmembrane helices 101-116 and 201-215, instead of four expected, but with many potential amphipathic β -strands. High peak in β -amphipathicity just next to the LDL or IDI motif is found at the *mas6_yeast* residue # 69, *mcp_r* (from mitochondrial carrier family) residue # 83 and *ach1_xenla* residue # 350. Observed amphipathic β -structure of a leucine rich repeat peptide LRP32 also contains LDL motif [99]. This motif may be important

in protein-lipid interactions because the peptide LRP32 integrates into lipid bilayers, probably as oligomer, forming amphiphatic β -sheet and promoting ion conductances.

The *tonb_ecoli* protein may be the molecular machine which transduces protonmotive force into mechanical energy [100]. Its proposed transmembrane topology with two potential TMH and three potential transmembrane β -strands [100] is key to the understanding how it connects inner bacterial membrane to outer membrane receptor proteins. We predict only one TMH for residues 13 to 32, probably anchored in the inner bacterial membrane, and several β -strand segments mostly close to the C-terminal, which can interact with outer membrane due to unusually long rigid and highly charged domain which connects these two domains. Our proposed structure for TonB is similar to proposed structure for the TolA protein [101] from *Escherichia coli*, which is also thought to connect inner and outer membrane. Very long connecting domain II of TolA (residues 48 to 310) has been modelled as an α -helical tether. Our prediction of transmembrane segment 14-33 in the TolA agrees with expected span 14-35 [101]. No TMH is predicted by us in the domain II region. This domain is associated with high preference for extramembrane α -helical conformation, but with very low preference for our transmembrane 'H' conformation.

4. DISCUSSION

The observation that conformational preferences are specified by the contexts - local segment primary structure, amino acid attributes, the three-dimensional environment in protein and environmental media, has been discussed before [102-105]. Algorithms that do take into account context-dependence of preferences [106] generally perform better for secondary structure prediction. In this report simple mathematical representation of context dependence is obtained through preference functions that are analytical functions of the surrounding sequence hydrophobicity or of any other amino acid attribute. Furthermore, preference functions are used to predict secondary structure motifs. It has turned out that for integral membrane proteins preference functions are excellent predictors of transmembrane segments in helical conformation. In fact preference functions are much better predictors than the hydrophobicity scale chosen to extract these functions.

A case in point is the application of the Kyte-Doolittle hydrophobicity scale directly and indirectly through preference functions. For the best known membrane proteins direct application of the Kyte-Doolittle algorithm and of its improved versions [107] is inferior to the performance of our algorithm that was also used with Kyte-Doolittle hydrophobicity scale. For instance, helix B is not predicted as hydrophobic helix in subunits M and L of the photosynthetic reaction center but only as an amphiphilic membrane-spanning helix [107]. Helix F from bacteriorhodopsin could not be predicted even after change in the window size, but again only as an amphiphilic helix [107]. We did not use the hydrophobic moment calculations for predicting transmembrane helices, but only as a help in predicting potential membrane buried β -structures. We predict all of 11 transmembrane helices from subunits L, M, and H from both bacterial sources (*Rhodospseudomonas viridis* and *Rhodobacter sphaeroides*) without overpredicting membrane-spanning helices as happens when hydrophobic moment analysis is used in the predictor [18]. In 10 integral membrane proteins of known structure all observed transmembrane helices are predicted by us at their correct sequence location and none of nine long extramembrane helices are confused with transmembrane helices.

Transmembrane helical segments are predicted by us with a high accuracy in 168 integral membrane proteins. All of 168 tested membrane proteins are recognized as such, because at least one transmembrane segment is predicted in each protein. No TMH is predicted in porins.

There are several reasons why preference functions, based on a chosen hydrophobicity scale, are better predictors of transmembrane segments than that hydrophobicity scale. Helix formation in a suitable environment is an cooperative process when nearby residues in a sequence are not independent. In other words the preference for helix conformation of each residue strongly depends on hydrophobicities of its sequence neighbors (Figure 1). The sigmoidal shape of preference function dependence on average hydrophobicity, such as shown in the Figure 1, is found for all amino acid types (not shown). It is suggestive of an cooperative nonlinear process. This cooperative effect is most pronounced for transmembrane segments of integral membrane proteins.

For bitopic membrane proteins, having only one transmembrane segment, local sequence information should be enough to predict the sequence location of such a segment. The prediction accuracy of 97.5% reported for such segments (our result) is impressive only in the case when there is very little overprediction. In the case of bitopic membrane proteins we have found two different training procedures for extracting preference functions that result in high prediction accuracy without large overprediction. Obviously, such training procedures cannot be included in algorithms that use the same hydrophobicity scale, but do not use preference functions. One possible answer to the initial question is that preference functions are so much better than simple use of hydrophobicity scale, because preference functions are firmly connected with protein data base used for training and with secondary structure features present or expected in that data base. Therefore, another important advantage of preference functions is the possibility to enhance amino acid attributes or secondary structure preferences through training process that ends with extraction of preference functions. In our recent work [52] we demonstrated that enhancement of the Chou-Fasman type constant preferences for transmembrane configuration, leads also to high prediction accuracy for transmembrane segments, even if prediction model (two state model) and training procedure (without soluble proteins) was completely different. Evidence that transmembrane helices are autonomous folding domains [108] helps to clarify why many different methods of sequence analysis are good predictors of transmembrane segments that are potential TMH segments.

Inability to distinguish an α -helix from β -strand transmembrane structure is even more serious weakness of hydrophobicity analysis. To build any reasonable topological model for membrane protein we must know what is the secondary structure of its transmembrane segments. Such information cannot be the output of any other algorithm that uses hydrophobicity scale, without additional training that attempts to correlate amino acid attributes with conformational motifs in proteins of known structure. Residues known to prefer β -strand conformation in soluble proteins [109], are very frequent residues in known transmembrane segments [62,110]. It is possible that some membrane proteins with transmembrane helices have had predominantly β -structure before being incorporated in the membrane [111]. When algorithms, trained on soluble proteins, attempt to predict secondary structure of membrane proteins, transmembrane segments known to be helical are often broken or predicted as β -strands. Therefore, the training process that includes membrane proteins of known or partially known structure is absolutely essential for the recognition of transmembrane structural motifs.

The need for more extensive training procedure was recognized by neural network programmers, but they trained their algorithms only too well. Overtraining is more subtle, but equally serious problem, that can greatly diminish prediction usefulness. A case in point is Rost *et al.* neural network algorithm [9] whose performance is significantly decreased when tested on never-before-seen set of proteins (Results section). Since we did not use evolutionary information (alignments of similar proteins) our choice of 105 integral membrane proteins was unintentionally such that average number of possible homologues per one protein (as average weighted number of alignments that do take into account sequence lengths) was smaller in that group of proteins (14 per one protein) than in the set of 63 proteins selected by Rost *et al.* (23 per one protein) [9]. This would partly explain the decreased performance when 105 membrane proteins are tested with Rost *et al.* method whose accuracy depends on available evolutionary information [9].

There are several reasons why overtraining may have happened during Rost *et al.* procedure despite careful cross-validation procedure [9]. Firstly, the pairwise homology among chosen proteins was not always less than 30%, as documented before (Methods). In the original set of 69 membrane proteins used by Rost *et al.* there was a subset of 47 proteins that had less than 30% pairwise similarity with all other proteins from that subset and had on average only 13 homologues per each protein [9]. The prediction accuracy, as measured for that subset of proteins with the A_{TM} parameter (Methods), was only $A_{TM} = 0.665$ as compared with $A_{TM} = 0.736$ for all of 69 proteins. The remaining subset of 22 proteins (mainly opsins) with more than 30% pairwise similarity and with an average of 32 homologues per protein was predicted with much higher prediction accuracy of $A_{TM} = 0.814$. For membrane proteins, considerably less than 30% similarity in the sequence may be needed, when we want to exclude very similar folding motifs. Failure to exclude similar proteins will cause an artificial increase in prediction accuracy in the case when similar proteins are predicted with higher than average accuracy, no matter what prediction method is used. Secondly, multiple alignment procedure, as a part of the training and testing process, was specific for the chosen protein data base of 69 proteins [9]. It increased prediction accuracy for that data base, but it does not have to do so for a set of nonhomologous never-before-seen proteins that for instance are not associated with similarly large average number of homologues per each protein from that data base. Thirdly in the data set of only 69 membrane proteins the number of objects determining prediction accuracy is really quite small: not more than 20 to 30 transmembrane helices that are difficult to predict by using any prediction method. The prediction accuracy becomes quite high when such specific patterns are learned, either through direct training procedure or through the choice of filter parameters. Unfortunately, neural network parameters learned in the process become very specific for such patterns that may not repeat easily in proteins outside training data set. A known disadvantage of neural network algorithm is its inability to tell us what it learned, in this case how it become capable of correct prediction of transmembrane helices most difficult to predict.

Signal sequences are, as a rule, not predicted as transmembrane segments by the neural network algorithm [9]. In our data base of 168 integral membrane proteins there are 32 proteins with signal sequences at the N-terminal (labeled with letter 's', Methods). Rost *et al.* wrongly predict only 3 such proteins as having the transmembrane segment at the sequence location of known signal sequence (cyoa_ecoli, myp0_human and wapa_strmu) [9]. We predict all of 32 signal sequences except two as transmembrane helices. Overprediction happens because very high preference for transmembrane helix conformation is often

associated with signal sequences. Somewhat shorter length of signal sequences does not help, because many correct predictions of transmembrane helices are initially associated with predicted short segments (12 to 16 residues that have high preference for transmembrane helix). Filter modification with negative weight at protein N-terminal can easily eliminate most of false positive predictions of TMS at the location of known signal sequences [52]. We did not use such modifications in this work because it would lead to difficult to detect underpredictions of real TMH at the N-terminal, while overprediction of TMH is easily detected when it happens at the location of known signal sequence. One advantage of omitting filter modifications with respect to signal sequences is that potential signal sequences are predicted as TMH with the same high accuracy as all other TMH, but then additional information from experiments is needed to decide if potential TMH near N-terminal is indeed true TMH or signal sequence. Another advantage is that primary structures without transit polypeptide, or without N-terminal signal sequence next to first potential transmembrane segment can be tested with assurance that first TMH will not be underpredicted due to omission of the N-terminal segment. Underprediction of the whole first TMH, containing 35 amino acid residues, happens in the LHC-II sequence taken from the Nature article [13] or in the cb22_pea sequence without transit polypeptide, when Rost *et al.* method [9], optimized to eliminate signal sequences from consideration, is presented with such truncated versions of polypeptides.

Errors in the SWISS-PROT assignment of transmembrane segments will reduce the prediction performance for all prediction methods that use this data base as 'standard of truth'. Such errors can indeed happen. We discussed the case of tolq_ecoli protein from *Escherichia coli*, which has only one transmembrane segment according to SWISS-PROT version 29 assignment, but is strongly predicted by us with three transmembrane segments in helical conformation. The same topology of three transmembrane helices is currently accepted in the SWISS-PROT data base for very similar exbb_ecoli protein.

Fortunately, many different theoretical and experimental procedures were used in SWISS-PROT assignments for the proteins finally chosen by us, so that for the purpose of our weak training procedure this set of proteins can be considered as reference set, but probably not as the 'standard of truth'. Observed and predicted length distribution of transmembrane segments in protein data base (Figure 2) may indicate that considerable room is still left for improving the algorithm. However, average length of expected transmembrane segments in our test set of 168 membrane proteins (21.7 residues) is quite close to predicted average length (21.5 residues). In any event, the absence of length distribution for predicted transmembrane segments that is in-built in some of simpler algorithms using hydrophobicity scales is quite unrealistic.

The TMH predictor underpredicts some of expected transmembrane segments in voltage-gated channels [112] (cic1_cypca case was mentioned in the section 3.6). Closer analysis revealed that underpredicted TMS are highly charged S4 segments known to span the membrane with less than 10 residues [113]. Although often missed by the TMH predictor essential parts of channel machinery, such as S4 and P-segments of the *Shaker* potassium channel pore [114, 115], are clearly resolved by our preference profiles (in preparation).

The main goal of this work was accurate prediction of transmembrane helical structures, but we do realize that membrane proteins may exist that have both α -helices and β -strands as transmembrane structure. Preference function method is capable of predicting separately α -helical and β -strand conformation of segments that have potential to become

membrane buried. Known structures of β -class soluble proteins are used in order to extract β -sheet preferences and as a help in extracting turn preferences. The reason why we had to enlarge the data base of membrane proteins with soluble proteins of the β -class is very simple. Few porins of known structure were not enough to serve as the training data base for the extraction of β -strand preference functions. Therefore, as the best substitute we used soluble proteins of the β -class. It is not an disadvantage to use much more abundant information available in soluble proteins of known structure. The number of nonhomologous proteins used to train preference functions for one secondary structural motif, can serve as the rough estimate of what is the minimal number of proteins that must be used during training procedure by our method (30 to 40 integral membrane proteins and the same number of soluble proteins of the β -class).

We have used a very simple procedure to predict transmembrane β -strands in porins. As observed before [107,116] it is useful to take into account hydrophobic moment for assumed β structure when the goal is to predict such a structure. The standard training and testing procedure with the Kyte-Doolittle scale gives reasonably good results with porins and defensins in terms of predicting transmembrane β -strands, but overprediction of membrane β -structure happens in the photosynthetic reaction center subunits in the case when decision constants are fixed to zero values (not shown). Preliminary results with a choice of the Cid *et al.* [69] hydrophobicity scale are encouraging both in terms of increased accuracy in predicting TMBS and in terms of a low percentage of wrongly predicted TMH in soluble proteins (only 4 to 5% for our data sets of soluble proteins). At any rate, the prediction of β -strands, turn and undefined conformations as well as the calculation of hydrophobic moment profile for assumed α -helix and β -strand conformation helped to locate transmembrane helices and other potential membrane-embedded regular structures.

One application of our standard training and testing procedure is for the nicotinic acetylcholine receptor, where M1, M2, M3 and M4 segments are all strongly predicted as transmembrane helices, but in addition there are several sequence domains with a potential for membrane-embedded β -strands (Figure 5). Another application has been described in the case of mitochondrial carrier family proteins. In many proteins from this family that have a known tripartite structure we have seen such a structure revealed in great details through profile of summed β -moments and β -sheet preferences (not shown). Contrary to the proposed six-helix model for these proteins thought to be required to take account of the threefold repeat [82,83] tripartite symmetry does not require the presence of two transmembrane helices in each of three domains. A small change in the primary structure or even in polypeptide environment may be enough to transform one regular structure into another in one of three domains without significant change in the tripartite symmetry. Functional asymmetry of three domains is known to exist in these proteins and some experimental evidence already exists that movement of loops in and out of the membrane can regulate transport activity of the mitochondrial ADP/ATP carrier [117].

Our algorithm can give partial answer to the question what attributes are optimal predictors for specific folding motifs. Kyte-Doolittle type hydropathy values and Chou-Fasman type conformational preferences are two obvious answers to the question what amino acid attributes are good predictors for majority of transmembrane helices. Indeed, three such scales MODKD, KYTDO and CPREF (Table 4), are on the very top of the list of the best amino acid scales (Table 5). Performance parameters that punish overprediction (A_{TM} and Q_p) give advantage to hydropathy values. Modifications to the Kyte-Doolittle values in the MODKD

scale increase prediction accuracy by increasing Trp and decreasing Ala importance for the formation of TMH. Surrounding hydrophobicity scale for membrane proteins (PONG1) takes into account actual hydrophobic environment in the three-dimensional protein structures. It produces less of false-positive TMH predictions when tested through preference functions on soluble proteins (Table 8). It appears that this scale can be used when an alternative to Kyte-Doolittle scale [17] is sought, because very popular Engelman *et al.* scale [59] is associated with up to 25% of false-positive TMH predictions (Table 8). Optimal scale for identification of amphipathic helices (PRIFT) is obviously not optimal for the recognition of TMH. Solution hydrophobicity scales such as FAUPL are clearly inferior to protein derived scales such as PONG1, CHOTH or ROSEF. A good performance of scales that measure water-accessible surface area loss upon protein folding has been noticed before [32]. More interesting are relatively high A_{TM} scores for polarity scales GRANT and WOESE and for the antigenic determinant hydrophilicity scale HOPPW. It would be quite interesting, but outside the scope of this work, to see if some transmembrane helices, difficult to predict by hydrophobicity analysis, are well predicted by polarity or hydrophilicity attributes. Such job can be easily done by using the PREF suite of algorithms, version 3.0. Even scales with inferior performance, such as the Cid *et al.* scale [69], are potentially very useful when different folding motifs in the membrane are being sought: transmembrane β -strands instead of TMH.

The filter parameters of our optimal predictor for transmembrane helices were optimized by using the Kyte-Doolittle scale and a reference set of 63 integral membrane proteins having one or more of long transmembrane segments, for which experimental and theoretical analysis indicated an α -helix configuration. Optimization of parameters was done by trial and error procedure and certainly was not perfect. Automatic procedures for finding optimal parameters for the TMH predictor were recently developed within the framework of preference functions method [52]. We did not use such procedures due to their inherent shortcomings: the danger of overtraining the predictor is then increased and due to the size of optimization problem different order of parameter optimization can lead to different results. In any case, it is quite possible that some other scale of amino acid attributes could have been chosen initially in the optimization process to produce higher prediction accuracy than the KYTDO scale. The natural choice of scale associated with a chosen reference set of proteins is the scale of statistical preferences, such as the CPREF scale, that can be extracted from that data base of proteins.

To summarize, the practical advantages of using the PREF suite of algorithms are as follows:

- It is much less expensive in computer time than a neural network algorithm.
- It works with equal expected high accuracy in the case when very few or no homologues are known.
- It has the potential to identify those physical, chemical or protein-derived statistical properties that are the most important for segment folding into the TMH configuration.
- Well known Kyte-Doolittle scale [17] can be used throughout, except in the case when specific need exists to test other amino acid attributes.
- All stages of prediction process are associated with transparent rules that are objective, automatic and easily inspected.
- There is an automatic recognition of different folding types of integral membrane proteins and automatic choice of decision constants for each type which improves the prediction accuracy.

- Thirty to forty membrane proteins and same number of soluble proteins of known structure are sufficient to train the algorithm.
- Accurate prediction of transmembrane helical segments is superimposed on the prediction of all other secondary structure elements of interest.
- Peaks in the transmembrane helical preference of lesser height and width can be used for identification of primary structure segments of special interest such as signal sequences and pore-forming segments (in preparation).
- Membrane-embedded or surface-attached β -strands can also be recognized from the sum of prediction profiles for β -strand preferences and of hydrophobic moments for assumed β -strand conformation.

The negative aspects or disadvantages are as follows:

- Balanced training procedure is needed. Including many more extramembrane than transmembrane residues in the training data set is wrong not only because of unbalanced training procedure, but also because we know that undefined conformation is forced upon us for extramembrane residues due to our lack of knowledge.
- A high percentage of soluble proteins are falsely recognized as membrane proteins (from 12 to 17%).
- Only one conformation is predicted with high accuracy: transmembrane helix conformation. Predictions of other regular or irregular conformations are not associated with the same high accuracy.
- The monotopic membrane proteins [118], that cross only one bilayer but not two, such as the prostaglandin H2 synthase [119], and self-inserting membrane proteins or toxins [120], such as colicin A [121], diphtheria toxin [122], beetle δ -endotoxin [123] and annexin [124] are associated with poor prediction (not shown).

Several improvements to the proposed method can be envisaged.

- a) Multiple alignment was not used. It should improve prediction accuracy for a single tested protein when thirty to forty homologous proteins exist. As already shown before [125], the PREF method can use training data set of proteins specific for protein to be tested.
- b) The prevalence of positively charged residues in the interior loops [60,79] or 'positive inside rule' is shown to improve prediction accuracy of our algorithm [52], but was not used in the present work. The predictor can become informative about the direction of membrane crossing, especially in the case of plasma membrane proteins of bacterial origin, when 'positive inside rule' is taken into account.
- c) It is not known if mixed type α/β or $\alpha+\beta$ structure can exist as transmembrane structure and if so what combinations of α -helix segments and β -strand segments may join to form transmembrane structure. Extracting preference functions from large enough data base of porins and related proteins with β -strand transmembrane structure will soon be possible. Then, appropriate modification of PREF-SPLIT algorithm, along lines suggested in this report, will serve to predict sequence location of both transmembrane α -helices and transmembrane β -strands.

Availability of the prediction with preference functions. We have set up an automatic electronic mail server at the Internet address: predict@drava.etfos.hr. The server will return complete prediction results, such as given in Table 6, when provided with the sequence of your protein. For further information, send the word *help* to the server. Questions, comments and suggestions should be sent to juretic@mapmf.pmfst.hr or zucic@mia.os.carnet.hr.

ACKNOWLEDGEMENTS

We are grateful to Sandor Pongor from ICGB, Trieste, Italy and to Burkhard Rost from EMBL, Heidelberg, Germany for data bases of membrane and soluble proteins kindly provided for our use. Thanks are due to Vera Gamulin and Boris Lenhard from Rugjer Bošković Institute in Zagreb, Croatia, who helped us with Eisenberg and Rao and Argos analysis [20] that was carried out with the PCGENE software. This work was supported by the Croatian Ministry of Science and Technology grants 1-03-171 to D.J. and 1-07-159 to B.L. and N.T.

SUPPLEMENTARY MATERIAL AVAILABLE via INTERNET

Two data bases of soluble proteins of known structure used to find false positive prediction results (Table I and Table II). Gaussian parameters needed for evaluation of preference functions based on the Kyte-Doolittle hydrophathy scale [17] (Table III). Table with detailed prediction results for transmembrane helices in 168 integral membrane proteins (Table IV). Table with a detailed comparison of prediction results for 10 best known membrane proteins for our and three other algorithms (Table V). All these tables together with the FORTRAN 77 source code are available from the anonymous ftp server mia.os.carnet.hr in the /pub/pssp directory. The anonymous login is ftp and the e-mail address is accepted as password. The list of files with short descriptions is contained in the 00index.txt file.

REFERENCES

1. F. Eisenhaber, B. Persson and P. Argos, *Crit. Rev. Biochem. Mol. Biol.*, 30 (1995) 1.
2. P.Y. Chou and G.D. Fasman, *Biochemistry*, 13 (1974) 211.
3. J. Garnier, D.J. Osguthorpe and B. Robson, *J. Mol. Biol.*, 120 (1978) 97.
4. B.A. Wallace, M. Cascio and D.L. Mielke, *Proc. Natl. Acad. Sci. U.S.A.*, 83 (1986) 9423.
5. N. Qian and T.J. Sejnowski, *J. Mol. Biol.*, 202 (1988) 865.
6. D.G. Kneller, F.E. Cohen and R. Langridge, *J. Mol. Biol.*, 214 (1990) 171.
7. B. Rost and C. Sander, *J. Mol. Biol.*, 232 (1993) 584.
8. R. Lohmann, G. Schneider, D. Behrens and P. Wrede, *Protein Sci.*, 3 (1994) 1597.
9. B. Rost, R. Casadio, P. Fariselli and C. Sander, *Protein Sci.*, 4 (1995) 521.
10. M.S. Weiss, A. Kreuzsch, E. Schiltz, U. Nestel, W. Welte, J. Weckesser and G.E. Schulz, *FEBS Lett.*, 280 (1991) 379.
11. S.W. Cowan, T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R.A. Pauptit, J.N. Jansonius and J.P. Rosenbusch, *Nature*, 358 (1992) 727.
12. J. Deisenhofer, O. Epp, K. Miki, R. Huber and H. Michel, *Nature*, 318 (1985) 618.
13. W. Kühlbrandt, D.N. Wang and Y. Fujiyoshi, *Nature*, 367 (1994) 614.
14. S. Iwata, C. Ostermeier, B. Ludwig and H. Michel, *Nature*, 376 (1995) 660.
15. T. Tsukihara, H. Aoyama, E. Yamashita, T. Tomizaki, H. Yamaguchi, K. Shinzawa-Itōh, R. Nakashima, R. Yaono and S. Yoshikawa, *Science*, 272 (1996) 1136.
16. A. Bairoch and B. Boeckmann, *Nucl. Acids Res.*, 22 (1994) 3578.

17. J. Kyte and R.F. Doolittle, *J. Mol. Biol.*, 157 (1982) 105.
18. D. Eisenberg, E. Schwarz, M. Komaromy and R. Wall, *J. Mol. Biol.*, 179 (1984) 125.
19. P. Klein, M. Kanehisa and C. DeLisi, *Biochim. Biophys. Acta*, 815 (1985) 468.
20. J. Rao and P. Argos, *Biochim. Biophys. Acta.*, 869 (1986) 197.
21. J.A. Bangham, *Anal. Biochem.*, 174 (1988) 142.
22. J. Edelman, *J. Mol. Biol.*, 232 (1993) 165.
23. D.M. Engelman and T.A. Steitz, *Cell*, 23 (1981) 411.
24. S.H. White, *Annu. Rev. Biophys. Biomol. Struct.*, 23 (1994) 407.
25. J.P. Allen, G. Feher, T.O. Yeates, H. Komiya and D.C. Rees, *Proc. Natl. Acad. Sci. U.S.A.*, 84 (1987) 6162.
26. R. Henderson, J.M. Baldwin, T.A. Ceska, F. Zemlin, E. Beckmann and K.H. Downing, *J. Mol. Biol.*, 213 (1990) 899.
27. G. McDermott, S.M. Prince, A.A. Freer, A.M. Hawthornthwaite-Lawless, M.Z. Papiz, R.J. Cogdell and N.W. Isaacs, *Nature*, 374 (1995) 517.
28. M.S. Weiss and G.E. Schulz, *J. Mol. Biol.*, 227 (1992) 493.
29. U. Unwin, *J. Mol. Biol.*, 229 (1993) 1101.
30. F. Hucho, U. Görne-Tschelnokow and A. Strecker, *Trends Biochem. Sci.*, 19 (1994) 383.
31. S.W. Cowan and J.P. Rosenbusch, *Science*, 264 (1994) 914.
32. D. Juretić, B.K. Lee, N. Trinajstić and R.W. Williams, *Biopolymers*, 33 (1993) 255.
33. D.T. Jones, W.R. Taylor and J.M. Thornton, *Biochemistry*, 33 (1994) 3038.
34. C. Sander and R. Schneider, *Nucl. Acids Res.*, 22 (1994) 3597.
35. W. Kabsch and C. Sander, *Biopolymers*, 22 (1983) 2577.
36. J. Deisenhofer and H. Michel, *Science*, 245 (1989) 1463.
37. D.R. Madden, J.C. Gorga, J.L. Strominger and D.C. Wiley, *Cell*, 70 (1992) 1035.
38. M.A. Saper, P.J. Bjorkman and D.C. Wiley, *J. Mol. Biol.*, 219 (1991) 277.
39. B.K. Jap, *J. Mol. Biol.*, 205 (1989) 407.
40. S. Gerbl-Rieger, H. Engelhardt, J. Peters, M. Kehl, F. Lottspeich and W. Baumeister, *J. Struct. Biol.*, 108 (1992) 14.
41. F. Jähnig, in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., ed.) pp 707-717, Plenum Press, New York, NY, 1989.
42. G. Ried, R. Koebnik, I. Hindennach, B. Mutschler and U. Henning, *Mol. Gen. Genet.*, 243 (1994) 127.
43. V. De Pinto and F. Palmieri, *J. Bioenerg. Biomembr.*, 24 (1992) 21.
44. C.A. Mannella, M. Forte and M. Colombini, *J. Bioenerg. Biomembr.*, 24 (1992) 7.
45. C.P. Hill, J. Yee, M.E. Selsted and D. Eisenberg, *Science*, 251 (1991) 1481.
46. P. Bulet, S. Cociancich, M. Reuland, F. Sauber, R. Bischoff, G. Hegy, A. Van Dorsselaer, C. Hetru and J.A. Hoffmann, *Eur. J. Biochem.*, 209 (1992) 977.
47. P.K. Ponnuswamy and M.M. Gromiha, *Int. J. Peptide Protein Res.*, 42 (1993) 326.
48. D. Eisenberg, R.M. Weis and T.C. Terwillinger, *Proc. Natl. Acad. Sci. U.S.A.*, 81 (1984) 140.
49. A. Lupas, M. Van Dyke and J. Stock, *Science*, 252 (1991) 1162.
50. J.L. Cornette, K.B. Cease, H. Margalit, J.L. Spouge, J.A. Berzofsky and C. DeLisi, *J. Mol. Biol.*, 195 (1987) 659.
51. O.B. Ptitsyn, *J. Mol. Biol.*, 42 (1969) 501.

52. B. Lučić, N. Trinajstić and D. Juretić, in *From Chemical Topology to Three-Dimensional Geometry* (A.T. Balaban, ed.) pp 117-158, Plenum Press, New York, NY, 1997.
53. H.F. Lodish, *Trends Biochem. Sci.*, 13 (1988) 332.
54. M. Degli Esposti, M. Crimi and G. Venturoli, *Eur. J. Biochem.*, 190 (1990) 207.
55. B. Persson and P. Argos, *J. Mol. Biol.*, 237 (1994) 182.
56. G. von Heijne and C. Blomberg, *Eur. J. Biochem.*, 97 (1979) 175.
57. C. Chothia, *J. Mol. Biol.*, 105 (1976) 1.
58. G.D. Rose, A.R. Geselowitz, G.J. Lesser, R.H. Lee and M.H. Zehfus, *Science*, 229 (1985) 834.
59. D.M. Engelman, T.A. Steitz and A. Goldman, *Annu. Rev. Biophys. Biophys. Chem.*, 15 (1986) 321.
60. G. von Heijne, *J. Mol. Biol.*, 225 (1992) 487.
61. R. Grantham, *Science*, 185 (1974) 862.
62. C.M. Deber, C.J. Brandl, R.B. Deber, L.C. Hsu and X.K. Young, *Arch. Biochem. Biophys.*, 251 (1986) 68.
63. H.R. Guy, *Biophys. J.*, 47 (1985) 61.
64. C.R. Woese, D.H. Dugre, S.A. Dugre, M. Kondo and W.C. Saxinger, *Cold Spring Harbor Symp. Quant. Biol.*, 31 (1966) 723.
65. P.K. Ponnuswamy, M. Prabhakaran and P. Manavalan, *Biochim. Biophys. Acta*, 623 (1980) 301.
66. W.R. Krigbaum and A. Komoriya, *Biochim. Biophys. Acta*, 576 (1979) 204.
67. T.P. Hopp and K.R. Woods, *Proc. Natl. Acad. Sci. U.S.A.*, 78 (1981) 3824.
68. J. Janin, *Nature*, 277 (1979) 491.
69. H. Cid, M. Bunster, M. Canales and F. Gazitua, *Protein Eng.*, 5 (1992) 373.
70. S. Miyazawa and R.J. Jernigan, *Macromolecules*, 18 (1985) 534.
71. J. Garnier and B. Robson, in *Prediction of Protein Structure and the Principles of Protein Conformation* (G.D. Fasman, ed.) pp 417-465, Plenum Press, New York, NY, 1989.
72. J.-L. Fauchere and V. Pliška, *Eur. J. Med. Chem. - Chim. Ther.*, 18 (1983) 369.
73. M.A. Roseman, *J. Mol. Biol.*, 200 (1988) 513.
74. G. Casari and M. Sippl, *J. Mol. Biol.*, 224 (1992) 725.
75. L.K. Poulsen, A. Refn, S. Molin and P. Andersson, *Mol. Microbiol.*, 5 (1991) 1627.
76. D.N. Wang, V.E. Sarabia, R.A.F. Reithmeier and W. Kühlbrandt, *EMBO J.*, 13 (1994) 3230.
77. R.R. Kopito and H.F. Lodish, *Nature*, 316 (1985) 234.
78. G.D. Fasman and W.A. Gilbert, *Trends Biochem. Sci.*, 15 (1990) 89.
79. G. von Heijne, *EMBO J.*, 5 (1986) 3021.
80. T.M. Savarese and C.M. Fraser, *Biochem. J.*, 283 (1992) 1.
81. M. Klingenberg, *Trends Biochem. Sci.*, 15 (1990) 108.
82. J.E. Walker, *Curr. Opin. Struct. Biol.*, 2 (1992) 519.
83. M. Klingenberg, *J. Bioenerg. Biomembr.*, 25 (1993) 447.
84. M. Klingenberg, *Arch. Biochem. Biophys.*, 270 (1993) 1.
85. D.R. Nelson, J.E. Lawson, M. Klingenberg and M.G. Douglas, *J. Mol. Biol.*, 230 (1993) 1159.
86. B. Jank, B. Habermann, R.J. Schweyen and T.A. Link, *Trends Biochem. Sci.*, 18 (1993) 427.
87. F. Palmieri, *FEBS Lett.* 346 (1994) 48.

88. J.-L. Popot and C. de Vitry, *Annu. Rev. Biophys. Biophys. Chem.*, 19 (1990) 369.
89. W. Bogner, H. Aquila and M. Klingenberg, *Eur. J. Biochem.*, 161 (1986) 611.
90. H. Aquila, T.A. Link and M. Klingenberg, *FEBS Lett.*, 212 (1987) 1.
91. G. Brandolin, A. Le Saux, V. Trezeguet, G.J.M. Lauquin and P.V. Vignais, *J. Bioenerg. Biomembr.*, 25 (1993) 459.
92. M.P. Blanton and J.B. Cohen, *Biochemistry*, 33 (1994) 2859.
93. B. Traxler, D. Boyd and J. Beckwith, *J. Membr. Biol.*, 132 (1993) 1.
94. W.M. Moore, L.A. Holladay, D. Puett and R.N. Brady, *FEBS Lett* 45 (1974) 145.
95. G.D. Fasman, *Biopolymers*, 37 (1995) 339.
96. J.E. Baenziger and N. Méthot, *J. Biol. Chem.*, 270 (1995) 29129.
97. E.L. Chang, P. Yager, R.W. Williams and A.W. Dalziel, *Biophys. J.*, 41 (1983) 65a.
98. U. Görne-Tschelnokow, A. Strecker, C. Kaduk, D. Naumann and F. Hucho, *EMBO J.*, 13 (1994) 338.
99. D.D. Krantz, R. Zidovetzki, B.L. Kagan and S.L. Zipursky, *J. Biol. Chem.*, 266 (1991) 16801.
100. P.E. Klebba, J.M. Rutz, J. Liu and C.K. Murphy, *J. Bioenerg. Biomembr.*, 25 (1993) 603.
101. S.K. Levensgood, W.F. Beyer and R.E. Webster, *Proc. Natl. Acad. Sci. U.S.A.*, 88 (1991) 5939.
102. S.-C. Li and C.M. Deber, *Int. J. Peptide Protein Res.*, 40 (1992) 243.
103. G.E. Arnold, A.K. Dunker, S.J. Johns and R.J. Douthart, *Proteins*, 12 (1992) 382.
104. L. Zhong and W.C.Jr. Johnson, *Proc. Natl. Acad. Sci. U.S.A.*, 89 (1992) 4462.
105. H. Wako and T.L. Blundell, *J. Mol. Biol.*, 238 (1994) 693.
106. J.-F. Gibrat, J. Garnier and B. Robson, *J. Mol. Biol.*, 198 (1987) 425.
107. F. Jähnig, *Trends Biochem. Sci.*, 15 (1990) 93.
108. J.-L. Popot, *Curr. Opin. Struct. Biol.*, 3 (1993) 532.
109. P.Y. Chou and G.D. Fasman, *Advan. Enzymol.*, 47 (1978) 45.
110. C.M. Deber, A.R. Khan, Z. Li, C. Joensson and M. Glibowicka, *Proc. Natl. Acad. Sci. U.S.A.*, 90 (1993) 11648.
111. L.L. Randall and S.J.S. Hardy, *Science*, 243 (1989) 1156.
112. W. Catterall, *Annu. Rev. Biochem.*, 64 (1995) 493.
113. S.A.N. Goldstein, *Neuron*, 16 (1996) 717.
114. H.P. Larsson, O.S. Baker, D.S. Dhillon and E.Y. Isacoff, *Neuron*, 16 (1996) 387.
115. A. Gross and R. MacKinnon, *Neuron*, 16 (1996) 399.
116. M.M. Gromiha and P.K. Ponnuswamy, *Int. J. Peptide Protein Res.*, 42 (1993) 420.
117. E. Majima, K. Ikawa, M. Takeda, M. Hashimoto, Y. Shinohara and H. Terada, *J. Biol. Chem.*, 270 (1995) 29548.
118. M.L. Jennings, *Annu. Rev. Biochem.*, 58 (1989) 999.
119. D. Picot, P.J. Loll and M. Garavito, *Nature*, 367 (1994) 243.
120. J. Li, *Curr. Opin. Struct. Biol.*, 2 (1995) 545.
121. M.W. Parker, J.P.M. Postma, F. Pattus, A.D. Tucker and D. Tsernoglou, *J. Mol. Biol.*, 224 (1992) 639.
122. S. Choe, M.J. Bennett, G. Fujii, P.M.G. Curmi, K.A. Kantardjieff, R.J. Collier and D. Eisenberg, *Nature*, 357 (1992) 216.
123. J. Li, J. Carroll and D.J. Ellar, *Nature*, 353 (1991) 815.

124. R. Huber, R. Berendes, A. Burger, M. Schneider, A. Karshikov, H. Luecke, J. Romisch and E. Paques, *J. Mol. Biol.*, 223 (1992) 683.
125. D. Juretić, B. Lučić and N. Trinajstić, *Croat. Chem. Acta*, 66 (1993) 201.

Polycyclic Aromatic Hydrocarbon Carcinogenicity: Theoretical Modelling and Experimental Facts

László von Szentpály⁺ and Ratna Ghosh⁺

Chemistry Department, University of the West Indies, Mona Campus,
Kingston 7, Jamaica, West Indies

1. INTRODUCTION TO CHEMICAL CARCINOGENESIS

Cancer is a disease in which the cell proliferation control mechanisms are deregulated. The International Agency for Research on Cancer (IARC) defines a human carcinogen as any agent, the exposure to which increases the incidence of malignant neoplasia in man. At present, about sixty chemical compounds are classified as human carcinogens and a number of others are under heavy suspicion. However, there are many more chemicals which have definitely caused cancer in experimental animals.

How should the **carcinogenic potency** of a chemical, i.e., its ability to cause a particular degree of tumour incidence in a test group, be evaluated ?

⁺ Contact address during sabbatical leave, August 1997 - July 1998: Waldburgstrasse 207A,
D-70565 Stuttgart, Germany.

Abbreviations : PAH, polycyclic aromatic hydrocarbon; DE, diol epoxide; PAHDE, polycyclic aromatic hydrocarbon diol epoxide; PAHTC, polycyclic aromatic hydrocarbon triol carbocation; TC, triol carbocation; BaP, benzo[a]pyrene; BeP, benzo[e]pyrene; BA, benz[a]anthracene; DBA, dibenz[a,h]anthracene; BcPh, benzo[c]phenanthrene; Ch, chrysene; MCh, methylchrysene; MBA, 7-methyl benz[a]anthracene; DMBA, 7,12-dimethyl benz[a]anthracene; EBA, 7-ethyl benz[a]anthracene; DB(a,l)P, dibenzo[a,l]pyrene; MSCR, mechanism-based structure-carcinogenicity relationship; PMO, Perturbational molecular orbital method; dA, deoxyadenosine; dC, deoxycytosine; dG, deoxyguanosine; MOS, monooxygenase enzyme system; EH, epoxide hydrolase enzyme system; N2(G), exocyclic nitrogen of guanine; C⁺, electrophilic centre of PAHTC; K, intercalation constant; CD, circular dichroism; LD, linear dichroism.

Several approaches for identifying chemicals that present a carcinogenic hazard to man are known. The most direct evidence is derived from properly conducted epidemiological studies of human populations. The observation by Percival Pott in the eighteenth century of the high incidence of scrotal cancer among chimney sweeps, caused by exposure to soot, marks the beginning of the study of chemical carcinogenesis [1]. From studies of workers in various industries, a considerable number of carcinogens have been established, including cigarette smoke, asbestos, cadmium, radon, arsenic pesticides, etc. [2]. Such studies are extremely difficult to perform and evaluate, because of the large number of factors - occupational, environmental, familial, etc. - which may influence the development of cancer in the human population. Bias in the ascertainment of the relationship between exposure and cancer may arise due to difficulties to conduct controlled studies.

Another line of evidence involves the use of standardised experiments on animals [3]. If a chemical produces tumours in multiple tissues and/or multiple species after a relatively short time - of the order of 100 days - and the tumour leads to the death of the animals, then this chemical is considered a strong carcinogen [4,5]. Animal studies define precisely the route, frequency, duration and level of exposure. But how relevant are the test animals to humans? This type of assessment is still very costly, time consuming, potentially hazardous to the experimentalists, and has also been criticised for sacrificing a large number of animals for results which are only indirectly related to risks for humans.

Correct prediction of the genotoxic risks is especially needed because of the long periods of genotoxic events. Therefore a series of short term *in vitro* tests have been developed [6-9]. A breakthrough for such tests was the recognition that many carcinogens and mutagens need metabolic activation. Since the cells used *in vitro* studies largely lack the ability to metabolise xenobiotics, exogenous activating systems are added [6-8]; hereby intact hepatocytes are preferable to liver homogenates [9]. These tests are rapid and inexpensive, but so removed from man that it is difficult to relate the results to human carcinogenic hazards.

The fourth approach to risk assessment is basically theoretical and is the only one which is definitely harmless to the scientist. In general, theoretical models, concentrate on a few essential steps of a complex process and introduce physically justified simplifications that enable a study of the otherwise intractable phenomena. Mechanistic models of carcinogenesis are characterised as having biochemically interpretable parameters. Such

models are developed by combining mechanistic chemical and molecular biological evidence with theoretical methods and tools at various levels of sophistication. Mechanism-based structure-carcinogenicity relationship (MSCR) analyses are important products of such modelling and have been increasingly integrated into the process of ranking chemicals for carcinogenic potency [10-20]. The ranking is a factor in the low dose extrapolation model used in estimating human risks associated with exposure to that carcinogen. With the ambition to become a theoretical short-term test, MSCR analysis is a critical filter in preventing the synthesis and production of new hazardous chemicals.

Concerning the mechanism of action, there are two types of carcinogens: (i) genotoxic (or DNA-reactive), and (ii) epigenetic (or non-genotoxic). The primary biological activity of a genotoxic carcinogen is alteration of the information encoded in the DNA, mostly through covalent binding to it. Epigenetic carcinogens do not cause DNA damage directly, their potency is derived from other activities, such as prolonged stimulation of tissue, increasing the rate of spontaneous mutations, inhibition of intercellular communication, or forced cellular growth.

A **three-stage model of carcinogenesis** by genotoxic agents is used to recognise and rationalise the complex process of tumour formation [21,22]. In the **initiation stage**, the carcinogen effects an irreversible damage to a cell. A unifying model to explain the initial event has been proposed by Miller [23]: all genotoxic carcinogens are either electrophilic reactants, or must be converted metabolically into a reactive electrophilic form that can attack some nucleophilic site on DNA. There is strong evidence that covalent binding to DNA is decisive for the carcinogenic activity and correlations between the potency of carcinogens and their covalent binding index to DNA *in vivo* have been established [24-28]. The damage is permanently fixed as a mutation following cell replication resulting in "dormant" tumour cells [25-28]. The step between DNA damage and mutation depends on the repair of the damage, the mutagenicity of the adducts and the rate of cell division. Any influence which increases the cell division rate reduces the time available for repair and increases the probability for a fixation of the damage in the form of a mutation.

The subsequent **promotion stage** requires much longer times, is initially reversible, and does not involve further DNA damage [22,27,28]. This stage is not specifically influenced by the electronic properties of the primary genotoxic agents. The chemicals involved in this stage are promoter-type carcinogens, for example diterpene esters, which are non-initiators and non-

mutagens [22,29,30]. The final **progression stage** is characterised by uncontrolled growth and brings about the expression of a malignant and metastatic tumour [31]. The initial steps of metabolism and interaction of carcinogens with DNA are at a molecular level, and the tools of theoretical chemistry are appropriate to understand them. Models have been developed reflecting the experimental knowledge of the metabolism of potential carcinogens. One frequently asked question is “how much information regarding the carcinogenic potency can be derived from the electronic structure of the parent molecules and their isolated or putative metabolites?”

This chapter reviews the theoretical modelling of polycyclic aromatic hydrocarbons (PAH) and their activated metabolites in the light of the accumulated experimental evidence for their modes of genotoxic action. PAH's form a large class of molecules which are ubiquitous in human environment, i.e., urban air, car exhaust, cigarette smoke or barbecued food, and encompass an immense variety of structural types. It is no surprise that their structure-property relationships have been of continuous interest to theoreticians. In fact, PAH's have served as the testing field for many of the approximations used in MO and VB calculations [32-39].

2. PAH CARCINOGENICITY AND THEORETICAL MODELS

The carcinogenicity of tar and other PAH containing materials was first demonstrated in animal experiments in the early 1900. Dibenzo[a,h]anthracene (DBA), **synthesised** in 1929, and benzo[a]pyrene (BaP), isolated from pitch in 1930, were the first pure PAH's shown to be carcinogenic to animals. The first review article on PAH carcinogenicity in mice was published in 1932 [40]. The Iball “carcinogenicity index” calculated from tumour incidence and latent period has been introduced in 1939 [4,41]. It is defined as :

$$I = \frac{100 (\% \text{ carcinoma bearing mice})}{\text{average latent period in days}} \quad (1)$$

The history of the quantum theoretical approach to the elucidation of PAH carcinogenesis reaches back to the late 1930's. Schmidt [42,43] and Swartholm [44] attempted to develop a parallel between calculated reactivities at certain regions of PAH's, which were later termed the K and

L regions (Fig. 1), and carcinogenic potency. The first well-known model of carcinogenesis is due to Pullmans and Daudels in France [45-51]. Reaction parameters connected to different regions of the parent hydrocarbon, such as the K, L and M regions (Fig. 1) were introduced, and activation by additional metabolic processes was expected at M-regions. Epoxidation in the K-region was assumed to compete with deactivating reactions at the L-region, thus Boyland's idea [52] that PAH epoxides are activated metabolites involved in carcinogenesis has been incorporated into the Pullman K, L, M-theory [48-51].

On the experimental side, it took more than twenty years to directly demonstrate the metabolic formation of PAH epoxides [53,54]. Although K-region epoxides appeared to meet the criteria for being the reactive electrophilic agents postulated by Miller [23], it was found that the PAH-nucleoside adducts did not arise from K-region epoxides [55]. The problem was experimentally resolved by the identification by Sims *et al* [56,57] of the vicinal bay-region diol epoxide of BaP and the demonstration that the *anti*- isomer of this metabolite reacts with DNA to give the products as obtained *in vivo* [58]. By 1976, experimental studies on "model carcinogens", such as BaP and BA have decisively contributed to the present knowledge of chemistry underlying PAH-induced carcinogenesis [56-59].

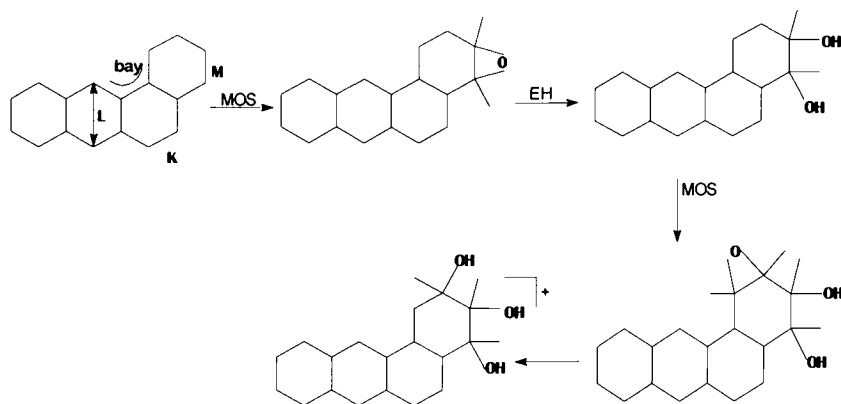


Figure. 1. Reactivity regions and metabolic activation leading to ultimate carcinogens shown on benzo[a]anthracene. (Aromaticity and H atoms are not depicted.)

The parent PAH's are largely inactive and do not cause DNA damage directly. They are made excretable by metabolism, which in principle can either activate or detoxify chemical carcinogens. The metabolism has been found to be very complex; it nevertheless displays common characteristics for the activation of the PAH themselves, [56-59] and their N-heteroaromatic relatives of the benzacridine and dibenzacridine family [60]. The initial activating step is an epoxidation of the M-region by the microsomal NADPH-dependent monooxygenase system (MOS), containing the enzyme cytochrome P-450c. The MOS activity from different tissues or cells toward different regions of a PAH varies greatly. This "regioselectivity" is one of the major factors that determine the susceptibility of animals and humans to carcinogenic effects. On the other hand, the MOS effectively oxidises compounds with very different structures; thus it is a very flexible metabolic system and possesses broad and overlapping substrate selectivities. The M-region epoxide can be deactivated nonenzymatically to phenols or form glutathione conjugates [57]. Further competing deactivating or less activating reactions may occur at other regions or sites and yield other epoxides, phenols or quinones.

In the second activating step the M-region epoxide (arene oxide) is transformed to a *trans*-dihydro diol by the epoxide hydrolase (EH) enzyme system [61]. The level of EH in the metabolising enzyme complex is important for the relative amount of the dihydro diol PAH metabolite(s) [61]. The stereospecificities of the MOS and EH determine the *trans*-configuration and the ratio of (+) and (-) enantiomers of the dihydro diol metabolites [62]. Borgen *et al* have shown that the M-region dihydro diol of BaP is metabolised to a reactive intermediate that binds more efficiently to DNA [63]. Sims *et al* demonstrated that a second enzymatic epoxidation by the MOS can lead to a bay-region dihydro diol epoxide (PAHDE) [56,57]. The likelihood of this epoxidation depends again on the regioselectivity of the MOS and, additionally, the conformation of the OH-groups in the M-region. Thus, the presence of a methyl, methoxy, fluorine or any other substituent at the *peri* position adjacent to the M-region induces a diaxial OH conformation in the diol and thereby reduces the metabolism to PAHDE [64]. Bay-region DE's exist in two diastereomeric forms, *syn* or *anti*, and each isomer occurs as a pair of (+) and (-) enantiomers. In addition, the *trans* OH-groups can assume diaxial or diequatorial conformations. In the *syn*-PAHDE (or PAHDE-1) the benzylic hydroxyl group and the epoxide oxygen are on the same face of the molecule, whereas in the *anti*-diastereomer (PAHDE-2) they are on opposite faces.

2.1. The bay-region theory

The discovery of the role played by PAHDE's in the chemical initiation of cancer is a landmark in developing our understanding of the mechanisms of carcinogenesis [56-59]. It has also enhanced the development of short-term tests by pointing out the necessity for exogenous activating enzyme systems [6-9] and led to the new theoretical approaches, *viz.*, the "bay-region theory" [10-15] and the MCS model [16,18,19]. Within two years of the identification of the bay-region of the diol epoxides of benzo[a]pyrene (BaPDE) by Sims and co-workers [56,57] and only a year after the publication of Dewar and Dougherty's book on "The PMO Theory of Organic Chemistry" [36], Jerina, Daly and Lehr [10] used perturbational MO (PMO) delocalisation energy, ΔE_{deloc} to describe the opening of the PAHDE's three membered oxirane ring and the formation of a PAH triol carbocation (PAHTC). In PMO theory, the extension of the π -system to include the exocyclic atom, b (Fig.1) is enhanced by

$$\Delta E_{\text{deloc}} = (2 - 2 |c_{\text{ob}}|) \beta \quad (2)$$

c_{ob} is the nonbonding MO (NBMO) coefficient on the carbon atom where the oxirane ring is opened. It is calculated according to the Longuet-Higgins zero sum rule for odd alternant aromatic hydrocarbons [36,65]. Incidentally, the first authors to try such NBMO coefficients of exocyclic atoms of odd-alternant PAH derivatives were Dipple, Lawley and Brookes [66] in 1968.

The bay-region theory predicted that vicinal diol epoxides formed adjacent to a bay-region should be more reactive and more potent carcinogens than other isomeric diol epoxides possible for the hydrocarbon. The model has been a useful guide to experimentalists in predicting the structures of ultimate carcinogens for a number of PAH's [67]. Considering the complexity of chemical carcinogenesis, it is surprising that the rank correlations between ΔE_{deloc} and the Badger index of carcinogenic potency [5] are not really satisfactory [10,16,51b,67]. Osborne has pointed out some "false positives" obtained from using ΔE_{deloc} : noncarcinogenic polyacenes and polyphenes are calculated to have larger ΔE_{deloc} than the most potent dibenzopyrenes [68]. The PMO-type bay-region theory [10,67] cannot explain why benzo[e]pyrene (BeP) and BA are just marginally carcinogenic whereas BaP is a strong carcinogen. In fact, the ΔE_{deloc} values [10] for the triol carbocations BePTC (0.714 β) and BATC (0.766 β) amount to as much

as 90.0% and 96.5 % of that of BaPTC (0.794β). Three shortcomings need to be mentioned :

(1) The bay-region model does not determine the likelihood of the M-region epoxidation, nor that of parallel less activating, or deactivating reactions. As pointed out by Pullman [51b] this appears to be a step backward from the earlier K, L, M-region theory. Even the strongest supporters of the bay-region theory have agreed that the concept of the bay-region alone is insufficient [10,69]. Pullman has suggested to adopt the complementary L-region idea [51b]. Models have been subsequently developed that account for reactions competing with the first activating epoxidation [12,16,19], one of which will be discussed below.

(2) Even more important is the fact that the formation of the triol carbocations (PAHTC) has not been correctly calculated. Any treatment based on a simple Hückel-MO or PMO calculations for odd AH ions neglect the effect of the differently charged carbon atoms and hence, must be in error. The ionic charge distributed over the aromatic system affects the electronegativity of carbon atoms in specific ways and this has a profound effect on the π -energy. Breakdowns of both the PMO and HMO approximations with ionic reaction intermediates are documented in the work of Dewar and Thompson [36,70], Streitwieser *et al* [35,71] and Szentpály [39]. The reactivity patterns with radical and ionic reaction intermediates of PAH are different [34-39]. It has been pointed out by Dewar [36] that the PMO method works better for radical than ions, and adequate modifications of the PMO method have been developed for ionic intermediates [16,38,39].

(3) The prominent stereochemical and shape selectivity observed with PAHDE's calls for an inclusion of the targeted parts of DNA into the modelling.

These three topics will be treated in the given sequence in the subsequent sections.

2.2. The MCS model

A first successful attempt to deal with two of the problems just outlined has been the MCS model of initiation of cancer [16,19]. The model quantifies three important factors influencing carcinogenicity.

M: the metabolic factor relates the probability for M-region epoxidation to that of the competing reactions.

C: stands for carbocation formation during the attack of the ultimate carcinogen on the DNA target sites.

S: a size and solubility factor to which shape and stereospecificity have to be added in a revised form.

2.2.1. Metabolic factor

As mentioned before, the Pullmans have designated the metabolically most reactive bond as the M region of the PAH [48-51].

Since the experimental establishment of the pertinent metabolic pathway, several authors have emphasised the importance of the metabolic factor assessed by the probability for the first activating enzymatic reaction [11-13,16,18,19]. Of course, caution is required in applying theoretical reactivity indices to enzymatic epoxidations. However, the monooxygenase enzyme system effectively metabolises a broad range of xenobiotic compounds and is characterised by broad and overlapping substrate specificities [59,72]. As a specific reaction at a given region in a group of related compounds is investigated, the rate of epoxidation is assumed to be determined by the reactivity of the PAH substrate rather than its fit into the receptor site.

The MCS model introduces a simple, but nonetheless effective index M describing the probability of M-region activation as compared to detoxification pathways. A nonconcerted mechanism for the enzymatic epoxidation has been assumed on the basis of experimental [73] and theoretical [74] evidence. The ease of epoxidation is negatively correlated to N_m the smaller of the two Dewar reactivity numbers [36] in the M-region.

$$N_m = 2 (c_{o,m-1} + c_{o,m+1}) \quad (3)$$

$c_{o,m-1}$ and $c_{o,m+1}$ are the NBMO coefficients at atoms $m-1$ and $m+1$ of the odd AH resulting from severing the π -system at atom. The NBMO coefficients are obtained by a pencil and paper procedure given by Longuet-Higgins [65] and Dewar [36]. Detoxification has been observed at different centres and regions. Several metabolites of low biological activity have been described for the "model carcinogens" such as BaP, BA, 7MBA, DMBA and DBA [59,75-77]. Phenylation at the most reactive position 6 of BaP is a major deactivation path [59,75]. L-regions have been postulated to be involved in deactivating reactions [48-51]. L-region epiperoxides of carcinogens were reported as noncarcinogenic¹ [78]. Some reactions

¹ However, a mutagenic effect of DMBA-epiperoxide has been observed in short-term tests [79].

competing against the bay-region DE formation have been found to involve K-regions [59a]. Detoxification, for example to water soluble conjugates and other diol epoxides occurs also during the later steps of the metabolism [57,59]. In order to find a common denominator for the manifold of possible reactions, the single most reactive position, say d of the parent PAH, has been chosen to provide a measure for the rate of deactivation processes. The position d is characterised by the smallest Dewar reactivity number, N_d on the parent PAH. This approach may be seen as providing a normalisation for the M-region reactivity with respect to the overall reactivity of the PAH. N_m and N_d have been combined to a single metabolic factor M [16].

$$M = (N_m - N_d)^2 \quad (4)$$

In order to differentiate between radical and cationic pathways of nonconcerted epoxidation, tests using radical and electrophilic PPP superdelocalisabilities have been carried out for the metabolic factor of methylated and N-heteroaromatic polycycles [19].

2.2.2. Carbocation formation

A simple modification of Hückel or PMO delocalisation energies ΔE_{deloc} was necessary in order to avoid the pitfalls of calculating involuntarily, the formation of a PAH triol radical while discussing that of a PAH triol carbocation (PAHTC).

Two major contributions to the π -energy change in the ring-cleavage reaction leading to a PAHTC need to be included in the model. In a simplified Pariser-Parr-Pople (PPP) method with Hückel orbitals as the basis [33], the PAHTC is an arylmethyl cation, with “resonance” energy

$$E_R = E_D + E_B + E_C \quad (5)$$

additional to that of the aromatic system of the PAHDE. E_D stands for the delocalisation energy gained by extending the π system to include the exocyclic atom b (Fig. 1). The PPP value for ED has been approximated [80] by

$$E_D = (1.50 - 1.03 |c_{ob}|) \beta \quad (6)$$

and is closely related to Dewar’s and Jerina’s

$$\Delta E_{\text{deloc}} = (2 - 2 |c_{\text{ob}}|) \beta$$

Since both E_D and ΔE_{deloc} are used in linear regressions, it is immaterial whether one of them, or simply $|c_{\text{ob}}|$ is taken. E_B , the bond-bond interaction energy is a small stabilising contribution depending mainly on the number of rings, and will not be pursued here. The other important contribution, the charge dispersal energy, E_C however enters with opposite signs into the resonance energies of arylmethyl ions and radicals. According to the PPP method [32,33,80],

$$E_C = -1/2 \sum_r \sum_s c_{\text{or}}^2 c_{\text{os}}^2 (\gamma_{rs} - \gamma_{rs}) \quad (7)$$

γ_{rs} are two-centre two-electron repulsion energies. For aromatic ions, the on-site electron pair densities, i.e., the diagonal elements of the spinless second order density matrix are lower than those of any classical structure. For radicals, however, the pair densities are increased relative to those in the corresponding classical structures. Thus the resonance energy of the ion exceeds that of the radical by $2|E_C|$ [33]. HMO and PMO do not differentiate between the formation of an arylmethyl radical and its carbocation. Empirically, radicals are better described by these methods; this has been related to the constant Coulson charge order $Q_\pi = 1$ for arylmethyl radicals as opposed to the variable π -charge order $Q_\pi \leq 1$ on arylmethyl cations [16,39]. The actual PPP values of E_C have been correlated to an excellent accuracy to their PMO- ω counterparts [16].

$$E_{C,\omega} = \omega \beta \sum_r (1 - q_r) q_r = \omega \beta (1 - \sum_r c_{\text{or}}^4) \quad (8)$$

Here, $q_r = c_{\text{or}}^2$ is the net π -charge on atom r of the carbocation, c_{or} the NBMO coefficient and ω an empirical parameter. For twenty arylmethyl cations, we found an excellent linear correlation ($r = 0.980$) between E_C of the PPP method and $\sum_r c_{\text{or}}^4$ of PMO- ω [16]:

$$E_C = -41.71 + 56.64 \sum_r c_{\text{or}}^4 \quad (9)$$

in kcal mol⁻¹ units.

2.2.3. Size factor

An examination of the PAH carcinogenicity data indicates that at least four rings are a necessary but not sufficient criterion for carcinogenic activity. The existence of an optimum size for carcinogenicity has been suggested, and approximately twenty to twenty-four carbon atoms have been inferred from plots of number of atoms vs. Iball indices [41b,81]. The size criterion has an empirical character, since the optimum size is influenced by many factors. The ability of the parent PAH to reach the site of activation, and that of the metabolites to reach the cell nucleus, along with intercalation properties and substrate-receptor specificities, appear to be inherently size and shape dependent and constitute factors possibly related to carcinogenic potency. The size was included as the first two terms of a power series expansion on the number, n of carbon atoms [18]. The statistically valid fit to the Iball indices obtained by a multilinear regression allowed the calculation of an optimum size of about twenty-one carbon atoms for a series of PAH [18], thus confirming earlier analyses [16,41,51a,81].

2.2.4. Performance and limitations

The MCS model is highly successful in ranking and rationalising the carcinogenic potencies of representative sets of bay-region and non-bay-region PAH molecules [16] as well as their methylated and N-heterocyclic derivatives [19]. According to the analysis of variances by a multivariate least squares technique the carbocation formation index $E_D + E_C$ is the most significant variable followed by the metabolic factor M , while the size factor trails far behind in significance. The overall fit to the experimental Iball indices of twenty-six PAH molecules is characterised by a standard error, $SE = \pm 6.8$ and a multiple correlation coefficient $r = 0.961$ [16]. The accuracy obtained could and should not be any better, considering that the confidence limit of the Iball index itself is $\pm 10^1$. Omitting E_C from the list of variables is equivalent to using Jerina's ΔE_{deloc} together with M and the size factor as the three variables. In doing this, the correlation decays drastically to $SE = \pm 14.1$ and $r = 0.824$, while the total F-value drops from 87.5 to a mere 15.5 by a factor of 5.6 [16]. If $E_D + E_C$ is the single variable, the linear correlation coefficient is still $r = 0.732$ for the same set of twenty-

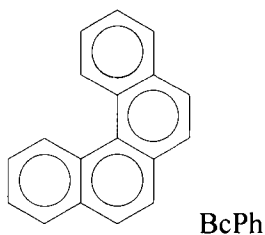
¹Several predictions have been made, and it was suggested to repeat or re-evaluate the experiments on picene [16]. Subsequent investigations triggered by our calculation of $I = 17$ for picene have confirmed the theoretical prediction and attributed a definitive carcinogenicity to this supposedly noncarcinogenic molecule [82].

six PAH molecules, whereas ΔE_{deloc} as the only variable is giving a much poorer value, $r = 0.450$.

However, the vibrant interest in SCR - both MSCR [10-20] and those involving artificial intelligence [83] - has subsided somewhat since the late 1980's, the reasons being the following :

(a) the experimental evidence for the “stereospecificity”, i.e., the strong influence of the absolute configurations of diol epoxides on carcinogenic activity came into focus [59,84], and it became mandatory to investigate its origins. It seemed unlikely that the stereospecificity can be comprehended by studying the electronic structure and reactivity of the metabolites per se without explicitly including the target site(s) in DNA. This will be discussed in Section 3.

(b) Exceptionally high carcinogenicity and mutagenicity has been established for “fjord-region” PAHDE metabolites [77,85-87], but they have defied attempts to rationalise by the models just described. Benzo[c]phenanthrene (BcPh) is a prototype by being the smallest PAH showing a fjord-region, i.e., a deep bay-region:



Notwithstanding its slight nonplanarity, the positional reactivities in electrophilic aromatic substitution correlate very well with Hückel localisation energies and there is no evidence that puckering of the rings has any effect on the reaction rates of BcPh [88], which was classified as a weak carcinogen [41b]. It came as a big surprise that both diastereomeric (*syn*- and *anti*-) BcPh dihydro diol epoxides (BcPhDE) are about one hundred-fold more active tumour initiators than the parent BcPH [77,85]. This is puzzling as the diastereomers have different, but relatively low hydrolysis rate constants [86].

In addition to *syn*- and *anti*-bcPhDE's four more “fjord region” diol epoxides have been synthesised and investigated, *viz.*, *syn* and *anti*-benzo[c]chrysene DE's (BcChDE) as well as *syn*- and *anti*-benzo[g]chrysene DE's (BgChDE) [87]. All six “fjord-region” diol

epoxides combine very *high* carcinogenic and mutagenic activity with **low** chemical reactivity. They are all more stable in physiological buffer at 37°C than the corresponding bay-region DE's, i.e., PhDE and ChDE [87], or the "model diol-epoxides" BaPDE. However, the mutagenicity of BcChDE is twelve times that of the highly mutagenic BaPDE. Contrasting to the stereoselectivity of BaPDE, the difference in mutagenic potency between the diastereomers is small for all of the fjord-region diol epoxides [87]. Obviously, such exceptional contrasts between carcinogenic and mutagenic potency on the one hand, and relative hydrolysis rates on the other cannot be explained without invoking a "shape-dependence" in the substrate-receptor interactions.

(c) The significant increase in carcinogenic potency due to methyl substitution at the aromatic end of the bay-region cannot be fully elucidated by electronic structure calculations [15,19,20,89,90]. Examples for the "bay-region methyl effect" are notable in comparing the Iball Carcinogenicity indices of BA, 12-methyl BA (12 MBA), 7-methyl BA (7MBA), 7,12-dimethyl BA (7,12 DMBA), chrysene (Ch) and 5-methyl Ch (5MCh) : I(BA) = 7 ; I(12MBA) = 25 ; I(7MBA) = 45 ; I(DMBA) = 125 ; I(Ch) = 5 ; I(5MCh) = 50 [41,91]. The increase in the chrysene pair is particularly striking by comparison to the negligible effect of the 5-methyl group on the hydrolysis rate constants of ChDE's. As could be expected from both PM arguments [20] and *ab initio* calculations [90], the acid-catalysed rate constant, k_{H^+} of *syn*-5MChDE is $k_{H^+} = 40 \text{ M}^{-1}\text{s}^{-1}$ while that of the non-methylated *syn*-ChDE is only marginally lower, $k_{H^+} = 36 \text{ M}^{-1}\text{s}^{-1}$ [92-94]. Another puzzling set of experimental facts is that 5MCh is about equipotent with the strong carcinogen BaP [91,93], while the hydrolysis rate constants of both *anti*- and *syn*-5MChDE are an order of magnitude lower than those of the corresponding BaPDE diastereomers [94]. The methyl group in the bay-region of 5MChDE actually enhances the extent of covalent binding to DNA, despite its presumed steric interference with bonding [59d,94]. The factors related to electronic structure and reactivity are therefore unable to account for the enhanced carcinogenic potency of 5MCh and its metabolites. On the other hand, the hydrolysis reaction, initially assumed to be a model for carcinogenic action [95], yielded a set of rate constants that correlates well with our $\sum_i c_{oi}^4$ in the charge-dispersal energy index E_C (eqns. 8 and 9), as will be shown in Section 6.

All of these observations evidence that **intrinsic activity**, i.e., the activity once the carcinogen has reached its site(s) of action, is **not** governed by the

molecular electronic structure alone. Such a conclusion contrasts the paradigmata of the SCR analyses as performed in the mid-eighties. One point which has become clear is that the ultimate carcinogen and DNA must be considered together in an environment mimicking the cellular scenario. There is a synergy or complimentary combination of various factors which affect multiple parallel and/or consecutive steps of chemical carcinogenesis. To this end, an overview of DNA-carcinogen binding and DNA-mediated hydrolysis of PAHDE's is presented in the next two sections.

3. DNA BINDING OF CARCINOGENIC HYDROCARBON METABOLITES

It is evident from the previous sections that carcinogenicity (and mutagenicity) are composite properties which are influenced by metabolic activation of PAH to PAHDE and PAHTC and the subsequent binding to cellular macromolecules like DNA. The mutation induction may reflect different isomers of PAHDE's. This in turn may lead to a greater probability of error during repair or replication of sequences containing adducts of the more active diol epoxides.

Of the two diastereomeric forms of BaPDE, it is the (+) *anti* isomer which preferentially forms covalent bonds between its benzylic C-10 and the exocyclic amino group (N-2) of deoxyguanosine (dG), primarily *via trans* epoxide ring opening [96]. Its absolute configuration was assigned with the help of exciton chirality dichroism method [58]. The minor adducts are obtained by binding to 6-NH₂ of deoxyadenosine (dA) as well as 7-N and 6-O of guanine [97-99]. However, these are not detected in more complex systems, possibly due to their labile nature or to the ease of recognition by DNA repair enzymes [99]. *Syn*-BaPDE-DNA adducts form in smaller quantities, mainly by reaction with N2(G) and also with adenine and cytosine [100]. In all tumourigenicity assays involving mice or mammalian tissues including those from humans, the (+) *anti*-BaPDE shows generally greater activity than the *syn*-isomer. The (-) enantiomer of *anti*-BaPDE is found to be a less potent tumour initiator than the (+) isomer [101]. The facts though correlating with the DNA binding capacity of these two enantiomers *in vitro* [102], cannot be entirely explained merely by differences in extent, strength and specificity of DNA binding or in the persistence of different DNA adducts [103]. Thus the extent of binding of the (+) *anti* isomer is 94 % at N2(G), 4 % to other positions on G and 2 % to

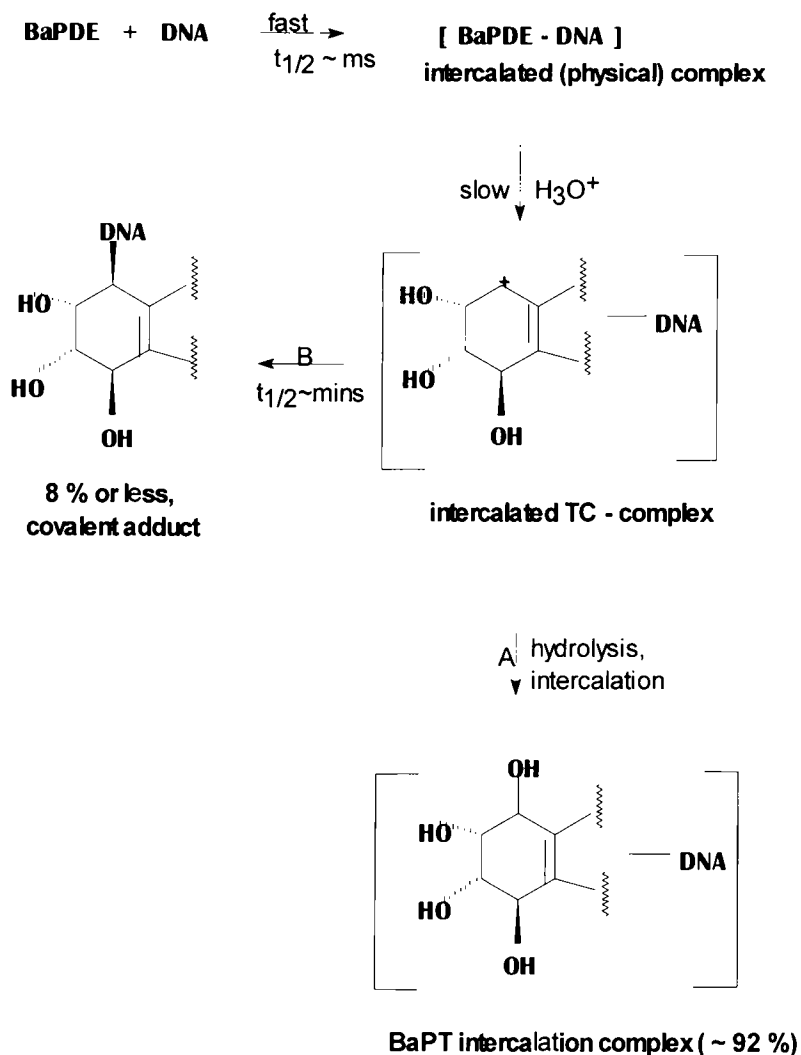
N(6)A. On the other hand, for the (-) *anti* isomer, it is only 59 % to N2(G), 21 % to O6(G), 18 % to N6(A) and 2 % to unspecified position on G. The relative binding to DNA of (+) *anti* and (-) *anti* is 10 : 1. It is worth mentioning that all the adducts formed initially may not survive the analytical procedure comprising multi-step extraction of DNA, high temperature enzymatic hydrolysis and subsequent chromatographic analysis. Processing of covalent adducts by cells, i.e., recognition and repair by appropriate enzymes and even the fidelity of replication may depend on conformations of these lesions, which in turn depend on the presence of hydroxyl groups at the 7- and 8-positions [104]. These hydroxyl groups in the potent carcinogen (+) *anti*-BaPDE prefer diequatorial orientation [105]. That the 6-fluoro substitution diminishes the tumourigenic potency is ascribed to the conformational transition from diequatorial to diaxial orientation of diol hydroxyl groups [106]. The fallacy of this argument and an alternative interpretation is provided in section 5.

In the late seventies a large body of experimental (mainly uv absorption, fluorescence and electric linear/circular dichroism spectroscopy) and theoretical (model building by hand or computer) data were already gathered regarding the geometry and conformations of various possible DNA-BaPDE adducts, with special attention to the nature of receptor sites. These data tally with a mechanism involving initial rapid intercalation (completed in < 5 ms) of DE between base pairs of DNA, followed by a slow protonation leading to the biologically significant, BaPDE-DNA covalent adduct (chart I). This, however, accounts for about 8 % of BaPDE, the major product being intercalated tetraol complex by route A [107]. The overall rate of reaction is influenced by pH, temperature, ionic strength and solvent, whereas the ratio of A/B is independent of those variables. Although the covalent binding of BaPDE is preferred at dG sites, the formation of physical intercalation complexes has been shown to be favoured at alternating dA-dT sequences [107].

Evidently the conformations of diol epoxide and/or its triol carbocation (TC) and also the secondary structure of DNA are the determinants of the binding process. To elucidate the mechanism of carcinogen-DNA interaction, however, some questions need to be addressed :

- (i) How significant is the role of intercalation ?
- (ii) Is there a significant base sequence specificity in the adduct formation ?
- (iii) Do the structure and the dynamics of PAHDE-DNA adducts affect the DNA conformation and topology ?

Chart I



Theoretically, the problem of stereoselectivity of the PAHDE-DNA adducts has been approached through computer graphics modelling [108], determination of preferred adduct orientations through energy minimisation

procedures [109], conformational space search using semi-empirical potential energy calculations [110], molecular modelling of intercalated complexes of *anti*-BaPDE enantiomers into double-stranded dC.pdG [111], and also modelling intercalated bay-region PAH metabolites (DE's and TC's) by molecular mechanics and *ab initio* calculations [112].

The experimental studies of the conformation in solution of *anti*-BaPDE adducts with DNA have been performed by optical methods [107,113] or two-dimensional NMR spectroscopy [114]. The main product, derived through bonding of the (+) enantiomer with the exocyclic N2(G) has the pyrene moiety located in the minor groove of the helix; it is variously described as having a wedge-like geometry and being intercalated in a bent region of DNA double helix [115] or dynamically disordered structure with two major orientation functions [116]. Other products obtained mainly from the (-) enantiomer can assume different conformations, some with pyrene in the enlarged minor groove and some with the long axis of the pyrene ring being nearly perpendicular to the DNA axis and presumed to be intercalated.

In terms of spectroscopic and fluorescence characteristics two major types of BaPDE adducts have been identified in native DNA or double-stranded polynucleotides. The "type I" or "site I" adducts show a negative linear dichroism (LD) spectrum and a 10-11 nm red shift in uv absorption maximum relative to BaPT in water (314, 328, 343 nm). The "site II" adducts display a positive LD and a small (2-3 nm) red shift in the uv region. In denatured DNA a third type of adduct, with λ_{\max} at 351 nm, is observed. While the type I adducts dominate in DNA modified with weaker carcinogens like (+) *syn*-BaPDE or (-) *anti*-BaPDE and resemble those of classical intercalation (physical) complexes of DNA, the type II adducts are more abundant in DNA modified with potent carcinogen (+) *anti*-BaPDE, where the externally bound conformation has the aromatic moiety residing in the minor groove with its long axis being inclined at an angle less than 35° to the DNA helix. In case of site I adducts the average angle between major transition moments in BaPDE and the overall helix axis *s* of DNA is reported in the range 61-79° in agreement with the quasi-intercalative description. Some LD studies, notably those of Hogan *et al* [115], and Ericksson *et al* [116], did not quite concur with the adduct location in an undisturbed minor groove.

Interestingly, the computations of Broyde and Hingerty in 1984-85 [117] attest to the possibility of two kinds of conformations for the adducts of (+) *anti*-BaPDE and dC.pdG. The carcinogen-base stacked conformation is very

close in energy (a difference of ~ 2 kcal mol⁻¹) to the other conformations where the backbone largely retains B-character and the carcinogen is placed outside the helix. Interconversion of the external to intercalated conformation (in which the nucleoside is slightly kinked, but not denatured) requires merely rotations of the guanine glycosidic torsion from *anti* to high *anti* and the C3' - O3' torsion from *trans* to *gauche* minus.

When the conformations of the aromatic chromophores of racemic BaPDE's were followed by the kinetic flow dichroism technique as a function of reaction time, the *anti*-diastereomer was observed to change conformation from an intercalative to an apparently external binding site, whereas the *syn* DE molecules did not appear to undergo any appreciable reorientation during or after the covalent binding reaction. The association constant for the latter (at 23°C, ionic strength ~ 0.005) is also smaller (5200 M⁻¹) than that for the *anti* compound (12 200M⁻¹). The fluorescence decay experiments, when duly corrected for the presence of free BaPT in solution, show an obvious heterogeneity of fluorescence lifetime and quencher accessibilities.

The emergent molecular view can be summed up as follows: the adducts of less potent (+) *syn*- and (-) *anti*-BaPDE's predominantly exist in quasi-intercalative conformations with relatively intact base stacking. On the other hand, (+) *anti*-BaPDE adducts exist in a locally disordered and flexible DNA structure, are labile and spend at least part of the time outside the DNA helix in the minor groove where it may get protonated more readily.

In a series of 1-alkylbenzo[a]pyrenes varying in size from methyl to *tert*-butyl, no significant differences were found in the extents of activation to carcinogenic metabolites, although biological activity and extent of intercalative binding to DNA were shown to be inversely related to the size of the 1-alkyl group [118]. These findings obviously imply that physical association of DE is a critical factor in the covalent attachment of carcinogens to DNA despite lingering confusion about their interrelationship. The earlier data of Pulkrabek *et al* [119] demonstrated that the extent of binding of racemic *anti*-BaPDE to DNA is directly correlated to the dG density, thus indicating random binding. However, the excimer fluorescence studies of poly (dG-dC).poly(dG-dC) modified with (+) *anti*-BaPDE [120] suggest a non-random covalent attachment favouring dG's close to already modified guanines. Other factors, such as the base sequence in the proximity of the target base, and the degree of denaturation can also influence the extent and the specificity of covalent binding [121]. For example, a dG within a poly dG tract is more vulnerable to attack by *anti*-

BaPDE than a dG with non-dG neighbours and so is DNA located at replication forks [122]. The fate of the final adduct may be dependent on the neighbouring bases in case of *in vitro* DNA cleavage experiments [123].

Since DNA and alternating poly(dG-dC).poly(dG-dC) are similar as substrates, they both give rise to type II adducts (major) with racemic as well as (+) *anti*-BaPDE. Distinct stereoselectivities are shown by double-stranded poly(dG-dC) vs. poly(dG).poly(dC) as suggested by the opposite signs of CD spectral signals of the pyrene moiety. The G-C polymers are much more reactive than the A-T polymers towards this metabolite. Salt titration of the *anti*-BaPDE modified B form of poly(dG-dC).poly(dG-dC) indicates that the externally bound entity becomes intercalated under high salt condition. *Syn*-BaPDE supposedly formed quasi-intercalative adducts with adenine [124]. However, LD measurements of BaPDE-modified poly(dA-dT).poly (dA-dT) give no signal from BaPDE chromophore spectral region, suggestive of disordered binding sites [125]. These generalisations are borne out by the studies of several non-alternating and alternating copolymer duplexes modified by BaPDE [126]. Intercalation complex formation is rather low ($K = 5\ 000\ M^{-1}$) in the case of dG-dC polymers and highest ($K = 21\ 000\ M^{-1}$) for dA-dT alternating sequences. Native DNA and also (dG-dT).(dC-dA) exhibit intermediate values of K . The preference of BaPDE for intercalation between neighbouring (dA-dT) alternating sequences may be due to their high flexibility and relatively low stability. In contrast, the (dA.dT) nonalternating copolymer is stiffer and may inhibit such an insertion between base pairs as well as effective overlap of the purines and the BaP ring system. The fraction of covalent binding is significantly higher in all dG duplexes, but does not necessarily correlate with the intercalative association constant.

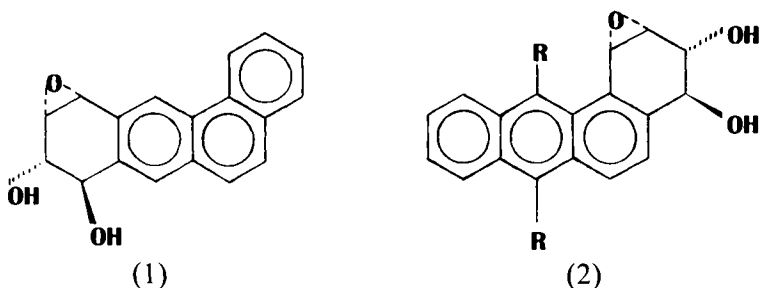
It is imperative to examine the effect of carcinogen-DNA adducts on the topology as well as conformation of DNA, since these properties are believed to influence the specific DNA interactions. From the spectroscopic studies (uv, CD, LD) it can be inferred that both B and Z forms of polynucleotides can covalently react with BaPDE with a very low affinity for the Z form [124, 127-128]. In fact, both diastereomers favour and preserve B-like conformations around the adduct site even at 4.5M NaCl. This might introduce flexibility in the poly (dG-dC). poly (dG-dC) structure manifested by a reduced LD signal [129]. Thus *anti*-BaPDE adduct formation may affect the behaviour of DNA selectively not only at the

binding site, but also at other points in the DNA sequence¹. Several researchers have found that the covalent binding of BaPDE to SV 40 DNA [131] and supercoiled ϕ X 174 DNA [132] bring about the unwinding of supercoiled DNA. The result is similar to, but less effective than that caused by the classical intercalator, ethidium bromide. Agarwal *et al* [133] observed that BaPDE causes a rapid positive supercoiling of relaxed circular pBR 322 DNA. Gel electrophoresis and kinetic flow LD experiments carried out by Geacintov *et al* [134] evidence a large initial unwinding of the supercoiled DNA on mixing with the (+) *anti*- or the (-) *anti*-BaPDE's, consistent with initial physical complex formation by intercalation. The subsequent slower decrease is attributed to rewinding of DNA, ostensibly due to removal of intercalated DE. It may lead to a final state of unwinding and an increase in superhelicity.

Similar studies on other carcinogenic hydrocarbon metabolites essentially corroborate the tentative picture presented above. However, many interesting details and conflicting features of bay-region hindered and fjord region DE's call for newer perspective in describing their interactions. The relatively low carcinogenicity of benzo[e]pyrene, BeP as compared to the strongly potent isomeric BaP has long been an intriguing problem. Conformationally BePDE differs from BaPDE in having the diol hydroxyl groups in one of the bay-regions with quasi diaxial orientation. This has been invoked to explain its reduced biological activity. However, axial hydroxyl groups are not necessarily detrimental to the chemical reactivity of a DE. For example, the OH groups of *syn*-BaPDE, an extremely reactive compound, are shown to be partially in diaxial conformation [135]. It is also noteworthy that the oxirane ring opening in BePTC formation and thus BeP carcinogenicity are greatly reduced according to the MCS model [16]. This is due to the relatively low charge dispersal energy, E_C even though the delocalisation energy, ΔE_{deloc} espoused by the bay-region theory [10,67] is only slightly smaller than that of BaPDE. The spectroscopic studies have revealed a heterogeneity of binding sites for the BePDE-DNA adducts, a stronger "site I" (i.e. intercalation site) attachment [136] and discernible effects of DNA concentration on the absorption and the ELD properties. Presumably stereochemical and shape factors reinforce the electronic parameters (*vide* section 5).

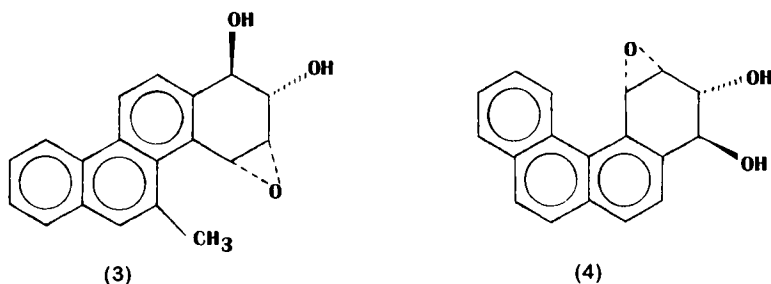
¹ For extensive treatment of long range effects of carcinogens on DNA, see Ladik and Förner [130].

In the case of benz[*a*]anthracene (BA), a weak carcinogen, non-benzo-ring-*anti*-diol epoxide isomer (1) seems to be the dominant metabolite, with only a small percentage of DNA adduct arising from the bay-region epoxide isomer (2), if R=H.



Both bind preferentially at the N2(G) site, but with varying reactivity (i.e., (2) > (1)). In the 7-methyl (7MBA) and 7,12-dimethyl (DMBA) derivatives the substitution in the meso region promotes formation of bay-region diol epoxides, which show greater activity, as expected from PMO-theory [20,36]. These PAH-DE's react with DNA by a similar mechanism as postulated for BaPDE's, producing a mixture of site I and site II adducts. The relatively high fractions of covalent binding and site II adduct formation by BADE appear to be paradoxical [107]. With DMBA both *anti*- and *syn*-diol epoxide isomers react extensively with DNA and a higher proportion of adenosine adducts is obtained, in addition to the usual N2(G) adducts. In recent years the absolute configurations of various DMBA-DNA adducts in hamster embryo cell cultures, in rat mammary epithelial cells as also in human mammary MCF-7 cell-mediated V 79 mutation assays have been delineated by using the sensitive ³²P post-labelling techniques [137] with the objective of evaluating the role of stereoselectivity in the biological activity of DMBA. The two *anti*-DMBADE-DNA adducts (bonded to dG and dA) comprise more than 90 % of the total adducts at all doses. The third major product arises from bonding of *syn*-DMBADE to deoxyadenosine. Such stereoselective activation of bay-region diols of hydrocarbons with methyl-hindered bay-regions is found to be a recurrent feature in many potent PAH carcinogens. The photoelectron spectral data and the results of *ab initio* SCF-MO calculations have been employed as descriptors for

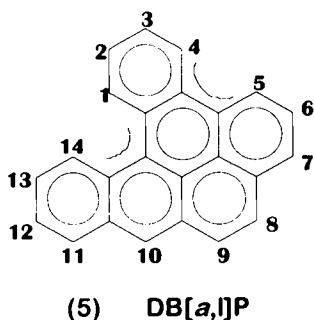
electronic influences on physical and covalent binding of BA and alkyl-BA metabolites to DNA [138]. The decrease of the first ionisation potentials from BADE to MBADE \approx EBADE (7-ethyl BADE) parallel to the increase in physical association constants, merely lends support to the conventional wisdom that alkylation of BADE in the 7- and 12- positions enhances reactivity.



A subset of PAHDE's with a substituent opposite the epoxide group, either a methyl group as in 5-methylchrysene diol epoxide (3), (5MChDE) or a benzene ring, as in benzo[*c*]phenanthrene diol epoxide (4), (BcPhDE) is characterised by a deep bay or a "fjord" region. These diol-epoxides perform better as DNA aralkylating agents and carcinogens than the isomers with unsubstituted bay-region. In fact, Jerina's research group [139] has reported preparation and chemical characterisation of sixteen principal adducts formed *via* reactions *in vitro* of four configurational isomers of 3,4-diol 1,2-epoxides derived from BcPh-*trans*-3,4-dihydrodiol, with deoxyguanylic and deoxyadenylic acids. Unlike BaPDE, BcPhDE's react, mainly through *trans* attack, with dA as well as dG residues in DNA. The actual proportions of dA and dG adducts and also the mode of attack (*cis vs. trans*) are strongly influenced by the absolute configuration of the diol epoxide. The technique of electrofluorescence polarisation spectroscopy (EFPS) also enables the investigation of these adduct structures [140]. For DNA or poly(dA-dT) treated with each stereoisomer of *anti*-BcPhDE a mixture of quasi-intercalated adenine adducts and externally bound guanine adducts are obtained. The angle between the transition moment of phenanthrene nucleus and the DNA helix axis ($55-61^\circ$) is always higher by $2-4^\circ$ for poly(dA-dT). Ostensibly phenanthrene nuclei in BcPhDE-dA adducts are tilted more than those of BcPhDE-dG adducts are tilted more than those of BcPhDE-dG adducts. The percentage reaction at adenine (A/A + G) is 80-90 % for the (+) *syn*-isomer and 40-60 % for the other three

isomers. The results are identical for another set of fjord region DE's, *viz.*, benzo[*c*]chrysene-DE's.

Adducts of DNA with chrysene (Ch), and its methyl analogues (5MCh, 6 MCh), gave values of $\sim 53^\circ$, indicating the major adduct to be external, probably located in the minor groove. It is further substantiated by the NMR study of solution conformation of the (-) *trans-anti*-5MCh-dG adduct opposite dC in a DNA duplex [141]. Chrysenes are unique in another way, being metabolised to shorter lived triol epoxides, in addition to the usual diol epoxides. The mutagenic and cell forming activities of *anti*-triol epoxide is greater than those of its *syn* isomer while both of them far exceed the biological potency and chemical reactivity shown by the corresponding *anti*- and *syn*-diol epoxides. The phenolic hydroxyl group in 9 position in chrysene-1,2-diol 3,4-epoxide is supposed to resonance stabilise the bay-region carbocation [142].



Metabolic activation of the very potent carcinogen, dibenzo[*a,l*]pyrene, *i.e.* DB[*a,l*]P is found to be stereoselective in human mammary carcinoma MCF-7 cell cultures [143a]. The fjord region 11,12-diol 13,14-epoxides are generated in preference to the bay-region DE's. The structure of DB[*a,l*]P may make 1, 2, 3, 4 positions less favourable for oxidation. Analysis of the resulting DNA adducts by 35S-post-labelling, immobilised boronate chromatography and HPLC brings out a product profile consisting of six different adducts. The major part is formed by reaction with deoxyadenosine of two (+) *syn*- and one (-) *anti*-DB[*a,l*]PDE's. The high carcinogenicity of DB[*a,l*]P must somehow be connected with the high extent of binding to dA in DNA of MCF-7 cells mediated by the metabolic

formation of a single regioisomeric pair of diol epoxides (cf. BaP) [143b]. Hecht and his co-workers have succeeded in synthesising these diastereomeric fjord region epoxides of DB[a,l]P and have also shown by NMR that they are in conformational equilibria, with the hydroxyl groups preferentially in pseudo-diequatorial orientation, as found in other fjord-region diol epoxides of BcPh and 5MCh [144].

This brief overview of some parts of the vast literature on PAH carcinogenesis, focussing substantially on benzo[a]pyrene diol epoxides allows certain inferences to be made, at the same time exposing many lacuna in understanding and theoretical modelling.

The carcinogenic potency of PAH's is definitely related to the metabolic formation of particular stereoisomeric diol epoxides in the bay or fjord region, and can be generally measured by the extent of covalent binding of such reactive PAH metabolites to DNA. However, both are merely necessary, **but not sufficient**, criteria of the tumourigenic activity of PAHDE or the parent PAH. The PAH intermediates with different levels of biological activity give rise to DNA adducts displaying different kinds of interactions with and distortions of the DNA molecule.

The question concerning the sequence of molecular events leading to covalent binding between DNA and the pertinent PAHDE still begs an unequivocal answer. One widely accepted hypothesis, supported by a wealth of experimental evidence and theoretical computations, puts forth physical association *via* intercalation as the essential prerequisite for covalent bonding leading to tumour induction. Intercalation certainly imposes limitations on size and shape of inserting molecule. There is, however, no direct proportionality between the magnitude of physical intercalation (expressed as intercalation constant, *K*) and the level of covalent binding when the ionic strength and salt conditions are varied in a reaction mixture [127]. The *K* values, of course, are affected by the size and orientation of the aromatic ring system *vis-à-vis* the DNA helix. Furthermore, while covalent binding of several PAHDE's is found at dG sites, physical intercalation occurs preferentially at dA-dT sequences. It is quite likely that the PAHDE molecules may diffuse from one binding site to another on the DNA molecule so that the initial physical intercalation with dA-dT sites may give way to the subsequent reaction with dG [59e,145]. Obviously, the significant parameters to be considered will include not only the chemical reactivity of the PAH reactive intermediates, but also steric factors, such as proper orientation and sufficiently close approach of the reacting electrophile to the base target site. The stereoselective covalent

bond formation of a particular PAHDE, say (+) *anti*-BaPDE and guanine in double stranded DNA, may be critically dependent on exactly how and in which time scale the physically bound BaPDE molecule can move in relation to the DNA molecule and whether the dynamic binding modes (e.g., intercalation *vs.* external association) would be seriously influenced by the nucleotide sequence and the substitution in different molecular regions of PAH.

4. HYDROLYSIS AND PAH CARCINOGENICITY

For a PAHDE to be carcinogenic it must have a reasonable lifetime to permit selective, specific and persisted interaction(s) with DNA. Activation near or at the target site is certainly the most effective means of generating potent carcinogens at a sufficiently high level. Still there are ample chances of detoxification either by its covalent binding to inhibitors or hydrolysis to inactive tetraols.

The encounter of cellular DNA and DE's take place in an aqueous environment. The dominant reaction pathway of BaPDE, for example, in such cases is the hydrolysis to tetraols, accelerated with reference to solvolysis in buffered media; only about 10 % of BaPDE is embodied in covalent adduct formation with DNA. The common denominator between the competing activation and detoxification pathways appears to be the formation and stabilisation of the reactive electrophile, the triol carbocation (PAHTC). This would involve the acid-catalysed cleavage of a C-O bond in the oxirane ring which is characterised by a high proton affinity and a moderately high C-O bond breaking energy after protonation. Aromatic rings in the immediate neighbourhood facilitate both protonation of oxirane¹ and ring opening.

Several questions arise regarding interrelationships between multiple reaction mechanisms of PAHDE's in the presence of nucleic acids. As in case of DNA binding, the solvolysis of BaPDE's, with or without DNA has been most thoroughly investigated over the last two decades. Hence, the ensuing discussions are focused on DNA-mediated hydrolysis of BaPDE, mentioning few other PAHDE's whenever relevant.

¹ The molecular origin of the 20 fold increase in mutagenicity of K-region arene imines over that of the corresponding arene oxides has been rationalised by the increased proton affinity of the former [146].

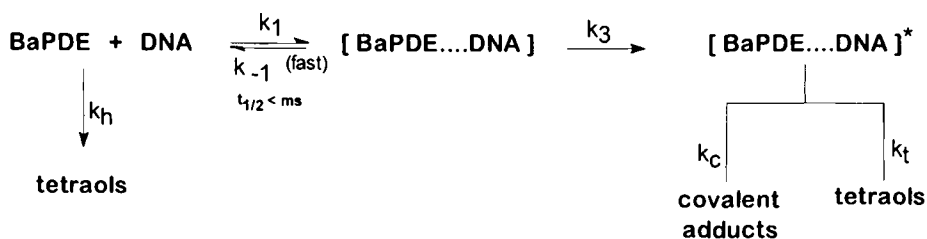
When BaPDE reacts with DNA in aqueous solutions, quite a few pathways are possible :

- (i) noncovalent intercalative complex formation with DNA;
- (ii) enhanced hydrolysis of BaPDE bound to DNA;
- (iii) acceleration of hydrolysis rate *via* general acid catalysis by DNA;
- (iv) covalent binding to DNA;
- (v) hydrolysis of free BaPDE molecules in solution.

A series of systematic studies on hydrolysis *per se* were undertaken by Sayer, Whalen, Islam, Jerina and others [147-150]. The thrust of their research has been to identify the stereoelectronic, conformational and other incidental factors affecting the kinetics and product distribution in the solvolysis of various PAHDE's under diverse reaction conditions. The aim is, of course, elucidation of reaction mechanisms of PAHDE's in biological systems. Geacintov *et al* have conducted detailed investigations of reactions of BaPDE's with aqueous solutions of double stranded DNA as a function of temperature, pH and NaCl concentration at different levels of DNA contents [151].

A kinetic scheme is constructed to take into consideration different parameters which may influence noncovalent binding of BaPDE to DNA, the hydrolysis and covalent binding of the complexed DE molecules [151]. It is tacitly assumed that decay of free BaPDE molecules by hydrolysis is negligible i.e., $k_h \ll k_3, k_1 [\text{DNA}]$.

Chart II



Furthermore, with $k_1, k_2 \gg k_3, k_h$ in this dynamic equilibrium system, BaPDE molecules will move rapidly between the binding sites and the outside solution. If the DNA concentration (5×10^{-5} - 10^{-3} M) is far in excess of that of BaPDE ($< 10^{-5}$ M) and the pH of the medium is below 9.5, more

than 95 % of DE bound to DNA is hydrolysed. The time frame is at least four orders of magnitude greater than that for noncovalent complex formation; the latter has no bearing upon the extent of covalent binding under the given conditions. The activation energy in the presence of DNA is significantly reduced to 36.4 ± 3.8 kJ/mols, from a value of 59.4 ± 2.9 kJ/mol in DNA-free sodium cacodylate (5 mM) buffer at an ionic strength of 0.1M NaClO₄ [151]. Obviously this is one of the reasons behind 180-fold rate enhancement. Possibly the micro-environment of BaPDE associated with DNA may favour the protonation of epoxide oxygen, thereby catalysing the formation of triol carbocation [152].

On the other hand, McLeod *et al* [153], have observed that both DNA-catalysed hydrolysis of (+) *anti*-BaPDE and its covalent binding to DNA are strongly dependent on DNA concentration as well as base composition, namely % (G + C) in purified native DNA and synthetic polynucleotides. The association constant, K for noncovalent intercalative complex formation does not exhibit a clear correlation to base sequence or base composition. The copolymer poly (dI-dC) which differs from poly (dG-dC) and natural DNA by lacking the guanine exocyclic amino group, is comparatively inefficient at catalysing hydrolysis. This is suggestive of a prominent role of the N2(G) in the catalytic process which appears to leave B-DNA structural features largely undistorted. Interestingly, this has already been established as the primary target for covalent binding of *anti*-BaPDE [154].

If, however, the catalytic mechanism would involve a protonated phosphodiester group, as suggested by Michaud *et al* [152], the ability of nucleotides to catalyse PAHDE hydrolysis is not anticipated to increase with increasing proportion of (G+C). The two sets of experimental findings can indeed be reconciled by postulating a two step reaction mechanism in which the first step has a base composition dependence and the second one is the acid catalysis by guanosine 5'-monophosphate (GMPH⁻). That it has a 60 - 80 times greater catalytic effect than that of H₂PO₄⁻ despite their similar pK_a values, can be traced to some sort of association between GMPH⁻ and BaPDE diastereomers [155].

In pursuing the fate of diol epoxides in spontaneous and acid catalysed solvolysis reactions (in dioxane-water, 0.1 M NaClO₄). Islam *et al* [156] have conclusively proved that the reactive intermediate in the rate-limiting step of spontaneous hydrolysis is a triol carbocation (TC). Trapping by the strong nucleophiles, aside and N-acetylcysteine anions rules out the possibility of a zwitterionic intermediate. While this is a certainty for *syn*-

BaPDE hydrolysed at $\text{pH} > 5.5$, there is no such clear evidence for the spontaneous hydrolysis of *anti*-BaPDE, occurring mainly at $\text{pH} > 7.0$. Hydronium ion catalysed hydrolysis at lower pH of both isomers are shown to be true general acid catalysis. Thus, in a rate-limiting step, especially for the *syn*-BaPDE's the carbocation may be generated through proton transfer from H_3O^+ or any acid with pK_a 5-8 which is synchronised with or followed by C-O bond breaking. If the resulting OH ion would react at a rate faster than its diffusion, the capture of the carbocation might escape detection and the observed mode of hydration - about 60 % *cis* in spontaneous reaction and only ca. 10 % *cis* in H_3O^+ -catalysed hydrolysis [147a] - of (+) *anti*-BaPDE could also be comprehended.

The 5MChDE's make an interesting case history. Metabolic activation of 5MCh containing two dissimilar bay-regions can produce two sets of diastereomeric diol epoxides: *anti*- and *syn*-DE I in which epoxy group lies in the "hindered" bay-region, i.e., flanked by methyl; *anti*- and *syn*-DE II where the epoxide is in the other bay-region. The carcinogenic potential of *anti*-MCh DE I is greater than that of the *anti*-MChDE II ($>$ *anti*-ChDE); DE I forms more DNA adducts than DE II in mouse skin *in vivo* and also binds more extensively to calf thymus DNA *in vitro* [157]. Similarly, *syn*-MChDE I is more effective than *syn*-MChDE II.

However, the trends in solvolytic reactivity, as measured by the hydrolysis (mainly pH-independent) rate constant, k_o (s^{-1}) or half life, $t_{1/2}$ in aqueous 1mM cacodylate buffer, do not agree with the above [92, 94]. It is interesting to compare the relative extent of binding of DE's to native DNA - there is hardly any difference between the ratio for the *anti* pair and that for the *syn* I and II enantiomers. Regardless of whether the DE is an *anti*- or *syn*- isomer, the ability to bind to DNA is magnified when the methyl group and the epoxide ring are located in the same bay-region.

Amplification of hydrolysis rate by DNA as well as the relative acceleration of rate of hydrolysis in presence of native and denatured DNA follow the same order : *anti*-DE I $>$ *anti*-DE II $>$ *anti*-ChDE $>$ *syn*-DE I $>$ *syn*-DE II. But qualitatively, the product profiles as obtained by HPLC are quite different. The importance of the secondary structure of DNA has been observed by McLeod *et al* in their examination of covalent binding of BaPDE and reverse BaPDE to DNA [153(a)]. Similar catalytic effects have been reported for *anti*-BaPDE hydrolysis : native DNA $>$ denatured DNA $>$ mononucleotides $>$ buffer [158]. All these findings suggest that physical interactions between PAHDE and DNA are important in determining the relative degree of covalent binding of these DE's *in vitro*. One can discern a

correlation with Chen's studies of physical binding of pyrene with deoxyribonucleotides, denatured DNA and native DNA [159]. The binding constant of pyrene decreases as follows: purine bases > pyrimidines, whereas double stranded DNA > single stranded DNA > mononucleotides. The general pattern, though not mimicked by solvolysis, has also been postulated by other studies on BaPDE's [151,152,155,160] and reverse BaPDE [153a].

It is well documented that enhanced hydrolysis of carcinogenic PAHDE's in the presence of DNA is acid-catalysed, involves formation of PAHTC in the rate-determining step and leads to both covalent adducts and tetraols. The clue to the mechanism is only partially available from the kinetic data in DNA-free buffered media which are also expressed in terms of pH-independent and acid-catalysed hydrolyses. Comparing several kinetic and mechanistic characteristics of the hydrolysis reactions with and without DNA, one can surmise that the more carcinogenic PAHDE's (*viz.*, *anti*-BaPDE, *anti*-BcPhDE, *anti*-ChDE etc.) undergoes acid-catalysed hydrolysis more rapidly whereas the *syn* isomers, the less active ones, possess higher k_0 values for spontaneous (pH-independent) hydrolysis. In other words, the *anti*-form would be hydrolysed in an acid catalysed reaction, likely to be prevalent in the acidic domains near the DNA surface. If formed in close vicinity to DNA, the carbocation (TC) may be expected to react with the nucleophilic sites on DNA. By implication, the acid-catalysed reaction is regarded as the optimal route for the binding to DNA while the pH-independent hydrolysis can be taken as a detoxification pathway.

The theoretical basis for such a rationale has been laid in the recent work of Pack *et al* [161,162]. Using the Poisson-Boltzmann approximation the pH-contour maps on and near the surface of B-DNA (poly(dG).poly(dC)) have been constructed under simulated conditions of 45 mM tris buffer with 3mM Mg^{2+} at pH 7.5. Three domains of high H^+ concentration ($>10\mu M$) are predicted: one is spread over the minor groove and two are localised in the major groove near N7(G) and C5(C) for a G.C base pair [114,163]. The reduction in pH by two units would translate into one hundred fold increase in TC production compared to the bulk rate. This is manifested in the accelerated rate of DNA-mediated hydrolysis. Elaborating on the two state model of Islam *et al* [149] in which the DE is either free or statically bound, Pack and Wong [163(a)] concluded that the catalysis by DNA is primarily an electrostatic effect of acidic domains in the surface grooves of the nucleic acid. While such computations were found satisfactory for *anti*-BaPDE hydrolysis, they could not adequately reproduce

the observed rate constants for *syn*-BaPDE. The [DNA] dependence of the kinetic behaviour of the latter might be construed as a signal for the physical binding of DE to DNA. Energy minimisation calculations of the conformation of representative intercalation complexes [114,163b] followed by pH mapping can model the measured rate profile quite well; but the issue of “site I” vs. “site II” adducts is not completely resolved. The calculations indicate that for *syn*-BaPDE intercalated site I dominates the hydrolysis rate for most of the DNA concentration range (up to 1.2 mM) owing to relatively lower pH (5.25), i.e., higher $[H_3O^+]$ near the epoxide oxygen of intercalated BaPDE. In contrast, this remains an open question for *anti*-BaPDE, although the fraction of intercalated, i.e., “site I” bound molecules is reported to be larger than that for the *syn*-isomer [149,152]. The acid-catalysed rate constants for both isomers are supposed to be unaffected by intercalation. Notably, Gupta *et al* [164] from an entirely different perspective arrived at the conclusion that the favourable stacking interactions between the aryl group of *anti*-BaPDE and the base at the transition state is responsible for the catalytic effectiveness (i.e., higher solvolysis rate) in nucleotide-catalysed hydrolysis. Another pertinent result from calculations of reaction paths for BaPDE's is concerned about energetic feasibility of proton transfer from H_2O and H_3O^+ [162]. The H_3O^+ -*anti*-BaPDE complex is found to be more stable by 22.5 KJ mol⁻¹ than the protonated *syn*-DE moiety, and the barrier to proton transfer is also less in the former case. But calculations have also indicated that the protonated diol epoxide is not a stable entity and rearranges immediately to the benzylic carbocation [165]. Such reasoning may divulge the link between the carcinogenic potency of the *anti*-isomer and its acid-catalysed hydrolysis in the acidic region of the DNA surface.

5. MOLECULAR MODELLING OF INTERCALATED PAH TRIOL CARBOCATIONS

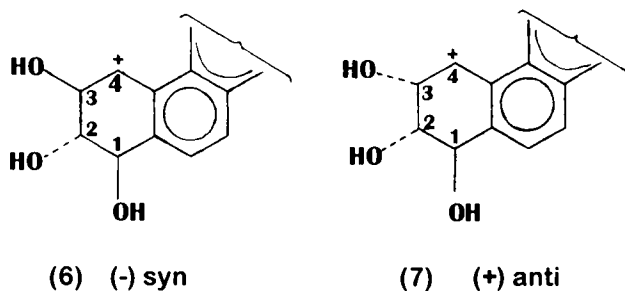
Many factors, e.g., molecular shape and size, conformational features of the metabolites, steric accessibility of the target centres, ease of formation and stability of the PAHDE and/or PAHTC, influence the ability of a PAH to express its potential as a tumour initiator. This complexity is a challenge to our search for a combination of a few highly significant factors to be grouped together in a mechanism-based model. Until recently, molecular modelling of physical complexes with DNA has been exclusively targeting

the PAHDE-DNA intercalation. However, the formation [59] and presence [156] of PAHTC (TC) intermediates calls for the inclusion of DNA-TC complexes into a comprehensive modelling.

It is of interest to investigate whether a TC intercalation between base pairs could accelerate the reaction(s) leading to covalently bonded PAHDE - DNA adducts. Surprisingly, DNA-TC complexes have become the object of molecular modelling only very recently [112]. The aim of the studies has been to find an explanation for the enantiomeric stereoselectivity and shape dependence of PAH carcinogenicity in terms of steric and energetic compatibility of the bay-region TC's with the B-DNA structure.

5.1. *Ab initio* calculations on PAHTC conformations

In analogy to the *syn*- and *anti*-diastereomers of PAHDE, there are two metabolically possible diastereomeric forms of PAH triol carbocations, *syn* and *anti*, each with two conformers differing by the geometry of the hydroxylated ring.



According to MNDO calculations of Adams and Kaminsky [14], the energy differences between the conformers are rather small. Thus additional correlated *ab initio* calculations should be performed to get reliable results. However, it is still prohibitive to fully optimise the geometry of triol carbocations of carcinogenic PAH's at a high level of *ab initio* theory, and no such calculation has been published till date. Phenanthrene, a noncarcinogen, is the smallest bay-region PAH. The conformational equilibrium of PhTC is expected to be similar to that of the carcinogenic

bay-region PAHTC's, since the geometry of the saturated part is likely to be determined by its environment.

The conformational energies of the half-chair conformers of metabolically allowed diastereomers of phenanthrene triol carbocation have been obtained by semiempirical and *ab initio* calculations at the fully optimised Hartree-Fock HF/6-31G and HF/6-31G* levels, and checked by single-point, full second order Moller-Plesset MP2/6-31G* calculations at the HF/6-31G* geometry [112a]. The full geometry optimisation of each conformer of phenanthrene triol carbocation at the direct HF/6-31G* level requires approximately three weeks of CPU time, and each single point correlated calculation by the direct MP2/6-31G* method takes twenty-four CPU hours using Gaussian 90 [166] on the Convex C3440 mini-supercomputer of the University of the West Indies [112].

Table 1 presents the relative energies of the conformers of PhTC isomers obtained from AM1, MNDO and *ab initio* calculations. According to the *ab initio* results, the conformational energy differences for the *syn* and *anti* structures are sufficiently small to allow for a coexistence of both conformers at room temperature. At the correlated MP2/6-31G*//HF/6-31G* level of theory, the eee and eea conformers of phenanthrene triol carbocation are predicted to correspond to the global energy minima *in vacuo* for the *syn* and *anti* conformations, respectively. These conformations have two intramolecular H-bonds, while the diaxial forms, aaa and aae, have only one. In the latter forms one of the three hydroxyl groups (2-OH in *syn* and 1-OH in *anti* diastereomers) is *trans* to the others. The energy of H-bonds is known to be very sensitive to the level of approximation used, and, remarkably, those which do not describe H-bonding sufficiently well, namely, MNDO and HF/6-31G, predict the higher stability of the diaxial forms (see ref.[14] and Table 1).

Table 1. Conformation energies (kJ/mol) of different forms of phenanthrene triol carbocation [112a]

Conformer ^a	Level of approximation				
	AM1	MNDO	HF/6-31G	HF/6-31G*	MP2/6-31G*//6-31G*
<u>Syn</u>					
eee	0.0	3.1	0.1	0.0	0.0
aaa	20.8	0.0	0.0	9.6	7.3
<u>Anti</u>					
eea	0.0	9.4	6.5	3.5	0.0
aae	2.0	0.0	0.0	0.0	0.6

^a The orientation of the 1-, 2- and 3- OH groups : a - pseudo-axial, e - pseudo-equatorial.

Hence, the conformational preference for the eee and eea forms *in vacuo* appears to be a consequence of the formation of an additional intramolecular H-bond. On the other hand, in the diaxial conformations the hydrogen atoms of the 1-OH and 2-OH groups are available for intermolecular H-bonds. Therefore the OH group orientations are expected to depend on the molecular environment. In particular, the stability of diaxial conformers should be significantly increased in a molecular environment where all possibilities for H-bonding can be utilised.

Even after trapping an intermediate bay-region PAHTC [156] there is little prospect to experimentally determine the conformation of reactive TC intermediates. The theoretical predictions can however, be checked indirectly by calculating the relative stabilities of the diaxial and diequatorial forms of PAHDE's, for which experimental estimates are available. NMR proton spin coupling experiments indicate equilibria between the two forms, with the diequatorial favoured for *anti*-DE's and

diaxial more likely to occur with *syn*-DE's [167]. Very recent calculations by Schwerdtfeger and Szentpály concur with experiment in showing that pseudo-diequatorial *anti*-PAHDE's are up to 30 kJ mol⁻¹ more stable *in vacuo* than their pseudo-diaxial counterparts [168].

As *syn*-DE's are normally less potent than the *anti*-enantiomers (the fjord-region and methylated bay-region DE's being exceptions), it has been inferred that a shift of equilibrium towards diaxial DE's in general reduces the carcinogenic potency. Such conclusions seem somewhat premature in view of the influence of the molecular environment on the hydroxyl group orientations. The possibility of a directed conformational change, *viz.*, *eea* → *aae* by specific hydrogen bonding to proton acceptors, e.g. N3 atoms of dG, has to be considered for DE's and reactive TC intermediates alike.

The *ab initio* calculations indicate that the bay-region PAHTC conformers with pseudo-diaxial orientation of the 1-OH and 2-OH group (structures 6 and 7) are energetically acceptable at room temperature and can be populated by directed conformational change, even if the pseudo-diequatorial conformers were favoured in aqueous solution [112]. This provides a quantum-theoretical rationale for the results of the AMBER force field calculations [112] presented below.

5.2. AMBER modelling of intercalated PAHTC-DNA complexes

The equilibrium conformations of physical complexes of selected PAHTC's with a dinucleotide fragment of B-DNA have been calculated by the all-atom AMBER empirical force field method [112]. The B-DNA has been represented by the dG₂.dC₂ dinucleotide fragment. No phosphate groups were placed at the ends of the strands, which instead carried terminal O3' and O5' hydroxyl groups.

Using AMBER, the geometry of the dinucleotide and its complexes with PAHTC's and PAHDE's is obtained by potential energy minimisation in the space of the independent bond and torsional angles, while the covalent bond lengths are fixed at their standard values [169]. The potential energy function is represented by bond angle bending, torsional distortion, van der Waals and electrostatic interactions, and hydrogen bonding. Atomic point-charges of the dinucleotide are taken from [169]. Since the model does not explicitly include solvent molecules, a dielectric function $\epsilon(R) = R/1A$ is applied [169]. The geometry of nucleic bases is fixed and corresponds to the X-ray structure. The hydrogen atoms of hydroxyl and methyl groups are allowed to rotate around the C-O and C-C bonds, respectively; the other hydrogens are fixed at their standard distance from the carbon skeleton. The

bond lengths and atomic point-charges of PAHTC's are obtained from fully optimised AM1 calculations [170]. The charge of hydrogen atoms attached to sp^2 -carbons is incorporated into the charge of the latter, whereas the other hydrogens are explicitly treated.

The B-DNA structure given by Arnott and Hukins [171] is taken as initial conformation for the DNA fragment as this structure is close to the native DNA at physiological conditions. According to electric linear dichroism measurements [154], there are several possibilities for the mutual geometrical arrangement of the components of the physical complex. Hence, different initial orientations and conformations of PAH metabolites were considered with the aim to find the most stable structure of the intercalated complex.

Figure 2 presents the absolute configuration of twenty bay-region PAHTC's. The overlap of the p -orbitals of the reactive centres, C^+ and $N2(G)$ is assessed by the two-dimensional distance R_{2D} , a projection of their separation on a plane parallel to the base-pairs. This can be used as an index for grading their readiness for bonding (see Table 2).

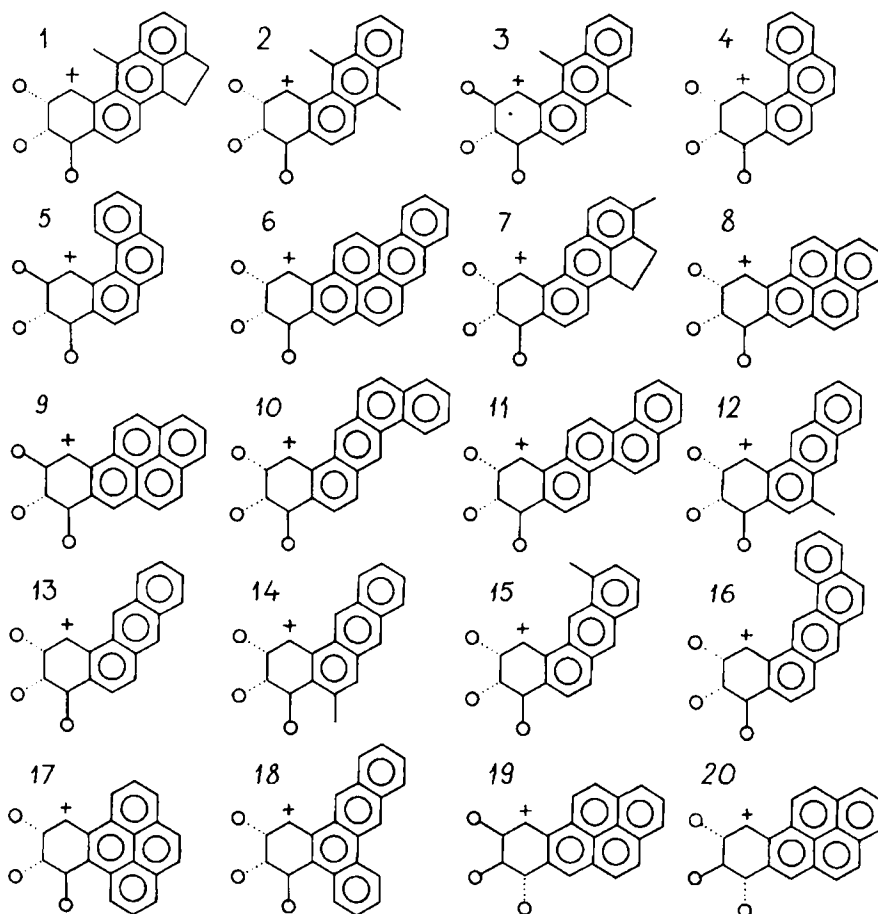


Figure 2. Structures of selected bay- region PAH triol carbocations. (Hydrogen atoms are not depicted, OH groups are shown as o.)

The equilibrium geometry of the carbon skeleton of TC's does not change much upon complexation, while the hydroxyl groups alter *substantially* their conformations and are involved in specific intermolecular interactions. The AMBER structures of the physical complexes of the dinucleotide with TC's are characterised by the following common features :

- (i) The axially oriented hydroxyl groups OH(1) and OH (2) (schemes 6 and 7) form H-bonds with the N(3) atoms of the adjacent guanine residues;
- (ii) The reactive exocyclic atom, C^+ is located close to the target N2(G).

Table 2. Distance R_{2D} , a measure for the preorganisation of intercalated PAHTC and dG₂.dC₂ dinucleotide

Structure ^a	1	2	3	4	5	6	7	8	9	10
$R_{2D}/\text{\AA}$ ^b	0.30	0.21	0.14	0.61	0.37	0.95	0.88	1.02	1.08	0.99
Structure	11	12	13	14	15	16	17	18	19	20
$R_{2D}/\text{\AA}$	0.80	0.84	0.92	1.82	1.69	1.38	0.90	1.11	1.26	0.89

a See Figure 2.

b Two-dimensional distance between reactive centres C⁺ and N2(G).

Thus two hydrogen-bonds between TC's and the DNA fragment determine their mutual orientation and stabilise the complex. In addition, these H-bonds are stereospecific and caused by the stereochemical compatibility of the N(3) atoms in successive guanine residues and the axial hydroxyl groups in PAHTC's.

The most potent ultimate carcinogens (1-5) display the smallest R_{2D} values and thus occupy positions of maximum juxtaposition of the atoms involved in the alkylation. The other molecules inflict noticeable conformational changes in DNA and deviate from the ideal position mentioned above. The structural features of the complexes with inactive structures 14-20 are least favourable for covalent bonding between the reactive centres. Due to the conformational restrictions imposed by the neighbouring bases, native B-DNA is more rigid than a dinucleotide fragment. As the conformational excitation of the fragment increases, in the case of inactive metabolites, the PAH-DNA incompatibility is expected to increase up to impeding the intercalation into native DNA.

The most interesting examples for PAH metabolite-DNA steric incompatibility are the enantiomers 19 and 20 of the carcinogenic metabolites 8 and 9 of benzo[a]pyrene, respectively. All of these four structures are electronically highly stabilised [16] and form stable complexes with the DNA fragment, but with 19 and 20 the fragment

becomes left-stranded instead of right-stranded [112a]. Although such considerable conformational change is allowed in a flexible DNA fragment, it seems most unlikely in native DNA.

Figure 3 summarises equilibrium geometric features of the intercalated complexes. In all of the highly potent PAHTC structures 1-5 the A area next to the bay-region is occupied either by a methyl group or a benzene ring. This group or ring is canted away from the C2' methylene group in the deoxyribose residue of the cytosine strand towards the guanines.

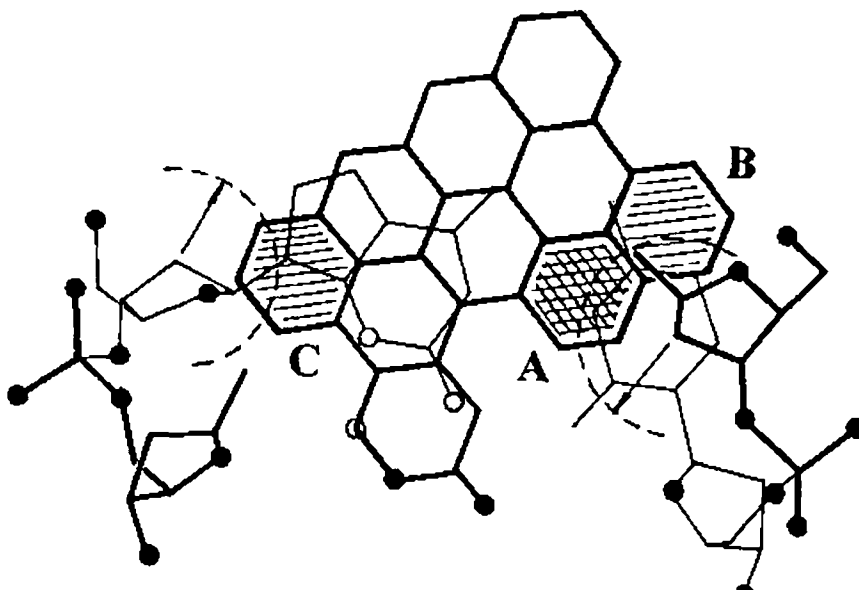


Figure 3. Shape selectivity of a generalised (+) - *anti*- PAHTC for dG₂.dC₂. Heteroatoms and groups shown as circles (● top, ○ bottom). Adjuncts in region A enhance, whereas those in regions B and C reduce carcinogenicity. Van der Waals radius of C2'-methylene groups indicated by dotted circles.

The strong enhancement of carcinogenicity upon methyl substitution at the aromatic flank of the bay-region could not be explained by electronic structure calculations [89,90]. According to the AMBER modelling, the “bay-region methyl effect” is a consequence of the improved fit between the triol carbocation and the DNA binding site. The molecular modelling points to a close parallelism between the effects of a methyl substituent and

an additional condensed benzene ring in the A area (Figure 3). In fact, similar heretofore unexplained enhancements of potency have been reported for other substituents in the same position, *viz.*, fluorine atom, hydroxyl group or even benzene ring [59,172]. The observation that the enhancement is insensitive to the chemical nature of the adjunct shows that steric requirements dominate over the differences in electronic structure. Two triol carbocation stereoisomers, derived from the (+)-*anti* and (-)-*syn* forms of such diol epoxides, conform to the spatial requirements with respect to the corresponding guanine residue equally well (see Figure 3 and Table 2). This explains the experimental fact that both diastereomers *viz.* 2 and 3, of DMBA and 4 and 5 of BcPh are exceptionally potent [77,85,173].

An increase in the size of the aromatic system promotes the stability of the carbocation and enhances carcinogenicity except in cases where the shape of the PAH precludes its intercalation [16,18,19]. Figure 3 shows two possible cases of such PAH-DNA shape incompatibility. Adjuncts in the B area are unfavourable because of their van der Waals contact with the C2' methylene group of deoxyribose residue on the cytosine strand. This leads to canting of the PAH metabolite away from the juxtaposition of reactive centres, C⁺ and N2(G). Structures 15 and 16 are examples for such an unfavourable shift. Another situation of repulsion is found in structures 14, 17 and 18. Adjuncts in region C create a van der Waals contact with the C2' methylene group of the deoxyribose residue of the strand of guanines; consequently the active centre of the PAHTC is pushed away from the target N2(G) toward the strand of cytosines.

Thus, we have two synergistic explanations for the reduced carcinogenicity of compounds with a methyl or methoxy group in *peri* position to the M-region, say 5MBA (14) :

- (i) The diaxial conformation of the M-region dihydro diol reduces the probability for a vicinal epoxidation by MOS [173].
- (ii) Even if a bay-region DE has been formed, the *peri* position falls in the region C, and bulky substituents exhibit a van der Waals contact with a C2' methylene group as described above. The reduced carcinogenicity is then due to the loss of preorganisation in the TC-DNA complex [112]. The marginal carcinogenicity of BeP (17) and dibenz[*a,c*]anthracene (18) can be comprehended by the same reasoning.

The structural features of the physical complexes of PAH triol carbocations with the DNA fragment explain the enantiomeric stereoselectivity and shape-dependence in carcinogenicity. The ability to adopt a particular conformation and orientation in the physical complex

formation with DNA seems to be a necessary condition for highly potent metabolites [59e,112].

The biologically active conformer of the *anti*-stereoisomer of triol carbocations is predicted to be *aae*, regardless of the conformation of the preceding diol epoxides. Consequently, the pseudo-diequatorial DE conformation should not be seen as a necessary condition for high potency. Models based on diequatorial PAHDE intercalation have led to conclusions which are at variance with experiments. Thus, the (+) *syn*-diol epoxide of benzo[*a*]pyrene should be carcinogenic according to ref. [109], in contrast to observations [59,77,84]. The Szentpály-Shamovsky model [112] avoids such contradictions by relating the tumourigenicity to the diaxial conformer of the triol carbocations as the true ultimate carcinogens. The biologically active conformation of TC's formed from (+) *anti* or (-) *syn*-enantiomers of bay-region PAHDE's is energetically acceptable. To sum up, the stereo and shape selectivity are mainly determined by the capability of TC's to adopt the biologically active orientation near the target centre in the physical complex with DNA.

6. CONCLUSION

The insight gained from explicit modelling by molecular mechanics and *ab initio* computations of TC-DNA physical binding emphasises the importance of shape selectivity and also the role of hydrogen bonding in shape selection. Evidently, structural descriptor(s) highlighting features of conformation, shape and possibly alignment conducive to van der Waals contacts should be included in an improved version of the MCS-type model. A multivariate linear regression on the most recent compilation of carcinogenicity indices [91] of thirty representative carcinogens (including DB[*a,l*]P, DMBA, MBA and 5MCh) gives very good correlation, $r = 0.97$ with the carbocation stabilisation parameter, $\sum c_{or}^4$ and a structural descriptor, F for the presence of adjuncts in the areas A, B or C (Figure 3) as the most significant variables. This is neither an artefact, nor a coincidence, but has a solid basis in the interpretations of mechanism founded on experimental findings and theoretical calculations.

In the conceptual framework of chemistry, stabilisation of the transition state is the single most important molecular event preceding a reaction which is the covalent binding to DNA in the context of initiation of PAH carcinogenesis. The PAHTC is established beyond doubt as the reactive

intermediate. Stabilisation of this TC would be related to the process of bringing together the reactive centres, namely C⁺ and N of the relevant base site in DNA, in the optimal alignment in **Transition State**, thereby compensating for the unfavourable entropic change through binding. In a way, the dG₂.dC₂ dinucleotide may be viewed as a “super-solvent” for potent (i.e., properly organised) PAH-TC's for stabilising this reaction intermediate by hydrogen bonds and electrostatic interactions.

It is no wonder, then, the acid-catalysed hydrolysis rate constant, k_{H^+} for some *syn*-PAHDE's also exhibits a very good correlation with the same parameter $\sum c_{or}^4$ as mentioned above. Incidentally, the details of these statistical analyses will be published elsewhere in the near future.

At this point of time we expect a renaissance of MSCR. Such relationships are obtainable by identifying discrete molecular functionalities which are relevant to the carcinogenic activity and by ranking the chemicals for potency by a model depicting these functionalities and their sequence. The essentials of such a model can be described in a nutshell as follows: DNA or more specifically dG₂.dC₂ catalyses PAHDE deactivation to tetraols. However, in up to 10% of the events, the catalyst itself is attacked by reactive TC intermediates. This does not only lead to “poisoning” of DNA as a catalyst, but may further destroy the whole system by deregulating the cell proliferation control mechanisms.

Diverse physico-chemical and biological parameters representing a range of possible mechanisms of interactions between DNA and several possible important metabolites modulate the carcinogenic potency of a particular PAH. The “trick” is to find out the right combination in the right context of the defined problem. We are still a long way from arriving at a complete theory of PAH carcinogenesis encompassing recognition and repair of DNA damage, but at least a comprehensive picture of the first irreversible steps of cancer initiation seems to be emerging.

Acknowledgements

The assistance of Dr. Werner Marx of Max Planck Institut für Festkörperforschung, Stuttgart, Prof. Frans Oesch and Mrs. H. Holl of University of Mains and Ms. M. Sinha of University of Cincinnati has been invaluable in searching for and collecting material through the labyrinthine expanses of literature on PAH carcinogenicity. Prof. W.C. Herndon and Mrs. Y. Shang of University of Texas at El Paso have helped by providing a copy of their recent compilation of carcinogenicity indices. The stimulating discussions with Prof. Oesch, Dr. A. Seidel, Prof. W.C. Herndon,

Prof. C. Párkányi and Dr. I.L. Shamovsky are gratefully acknowledged. The authors would like to thank Dr. I.L. Shamovsky and Dr. P. Schwerdtfeger for their computational contributions. The authors appreciate very much the wonderful rendering of the text into prescribed camera-ready format by Devon Gardner.

REFERENCES

1. P.Pott, Reprinted in Natl. Cancer Inst., Monogr., **10** (1963) 7.
2. R. Doll, R.Peto, J. Natl. Cancer Inst., **66** (1981) 1192.
3. IARC Monographs on the Evaluation of the Carcinogenesis Risk of Chemicals to Humans, Vols. 1-50, Intl. Agency for Research on Cancer, Lyon, 1972 onward.
4. J. Iball, Am. J. Cancer, **35** (1939) 188.
5. G.M. Badger, Brit. J. Cancer, **2** (1948) 309.
6. B.N. Ames, J. McCann and E. Yamasaki, Mut. Res., **31** (1975) 347.
7. Critical Evaluation of Mutagenicity Tests, R. Bass, V. Glocklin, P. Grosdanoff, D. Herschler, D. Müller and D. Neubert (eds.) MMV Medizin Verlag München, München, 1984.
8. K.H. Norpoth and R.G. Garner (eds.), Short-Term Tests for Detecting Carcinogens, Springer, Berlin, 1980.
9. D. Utesch, H.R. Glatt and F. Oesch, Cancer Res., **47** (1987) 1509.
- 10(a). M. Jerina and J.W. Daly, in Drug Metabolism - from Microbe to Man, D.V. Parke and R.L. Smith (eds.), Taylor and Francis, London 1976, p. 13.
- 10(b). D.M. Jerina and R.E. Lehr, in Microsomes and Drug Oxidation, V. Ullrich, I. Roots, A. Hildebrandt, R.W. Eastabrook and A.H. Conney (eds.), Pergamon Press, Oxford, 1977, p. 709.
- 11(a). G.H. Loew, J. Phillips, J. Wong, L. Hjelmeland and G. Pack, Cancer Biochem. Biophys., **2** (1978) 113.
- 11(b). G.H. Loew, B.S. Sudhidra and J.E. Ferrell, Chem. Biol. Interact., **26** (1979) 75.
- 11(c). J.C. Ferrell and G.M. Loew, J. Am. Chem. Soc., **101** (1979) 1385.
12. R.S. Umans, M. Koruda and D.J. Sardella, Mol. Pharmacol., **16** (1979) 633.
13. J.C. Arcos, Y.-T. Woo, D.Y. Lai and M.F. Argus, Chemical Induction of Cancer, Vols. III A, B and C, Academic Press, Orlando, FL, 1982-1988.
14. S.M. Adams and L.S. Kaminsky, Mol. Pharmacol., **22** (1982) 459.

15. J.P. Lowe and B.D. Silverman, *Acc. Chem. Res.*, **17** (1984) 332.
- 16(a). L.v. Szentpály, *J. Am. Chem. Soc.*, **106** (1984) 6021.
- 16(b). L.v. Szentpály, in *Molecular Basis of Cancer, Part A*, R. Rein (ed.) A.Liss, New York, NY, 1985, p. 327.
- 16(c). L.v. Szentpály, *Int. J. Quantum Chem., Quantum Biol. Symp.*, **12** (1985) 287.
17. W.K. Lutz, *Arch. Toxicol. Suppl.*, **7** (1984) 194.
18. W.C. Herndon and L.v. Szentpály, *J. Mol. Struct. (Theochem)*, **148** (1986) 141.
19. L.v. Szentpály and C. Párkányi, *J. Mol. Struct. (Theochem)*, **151** (1987) 245.
20. J.P. Lowe and B.D. Silverman, *J. Mol. Struct.(Theochem.)*, **179** (1988) 47.
- 21(a). I. Berenblum, in *Carcinogenesis as a Biological Problem*, American Elsevier, New York, NY, 1974.
- 21(b). I. Berenblum, in *Cancer*, F.F. Becker (ed.), Plenum Press, New York, NY, 1975, p. 323.
- 22(a). T.J. Slaga, A. Sivak and R.K. Boutwell (eds.), *Carcinogenesis*, Raven Press, New York, NY, 1978, Vol. 2.
- 22(b). T.J. Slaga (ed.), *Mechanisms of Tumor Promotion*, CRC Press, Boca Raton, FL, 1984, Vol.2.
23. J.A. Miller, *Cancer Res.*, **30** (1970) 559.
24. P.Brookes and P.D. Lawley, *Nature*, **202** (1964) 781.
- 25(a). W.K. Lutz, *Mut. Res.*, **65** (1979) 289.
- 25(b). W.K. Lutz, *Adv. Exp. Med. Biol.*, **136B** (1982) 1349.
26. D.H. Phillips, P.L.Grover and P.Sims, *Intl. J. Cancer*, (1979) 201.
27. I.B. Weinstein, in *Cancer and Chemotherapy*, S. Crooke and A. Prestyako (eds.), Academic Press, New York, NY, 1980, Vol. 1, p. 169.
- 28(a). T.J. Slaga, S.M. Fischer, L.L. Triplett and S. Nesnow, *J. Am. Coll. Toxicol.*, **1** (1982) 83.
- 28(b). T.J. Slaga, S.M. Fischer, C.E. Weekes, A.J.P. Klein- Szántó and J. Reiners, *J. Cellular Biochem.*, **18** (1982) 99.
29. B.L. van Duuren, *Progr. Exp. Tumour Res.*, **11** (1969) 31.
30. E. Hecker, *Methods Cancer Res.*, **6** (1971) 439.
31. L. Foulds, *Neoplastic Development*, Vol. 1, Academic Press, New York, NY, 1969.
32. R. Pariser and R.G. Parr, *J. Chem. Phys.*, **21** (1953) 466; 767.

- 33(a). J.A. Pople, *Trans. Faraday Soc.*, **49** (1953) 1375.
- 33(b). A. Brickstock and J. Pople, *Trans. Faraday Soc.*, **50** (1954) 901.
34. K. Fukui, T. Yonezawa and C. Nagata, *Bull. Chem. Soc. Jpn.*, **27** (1954) 423.
35. A. Streitwieser, *Molecular Orbital Theory for Organic Chemists*, J. Wiley, NY, New York, 1961.
36. M.J.S. Dewar and R.C. Dougherty, *The PMO Theory of Organic Chemistry*, Plenum Press, New York, NY, 1975.
37. W.C. Herndon, *Isr. J. Chem.*, **20** (1980) 270.
38. L.v. Szentpály and W.C. Herndon, in L.B. Ebert (ed.), *Polynuclear Aromatic Compounds*, Adv. Chem. Ser. 217, Am. Chem. Soc., Washington, DC, 1988.
39. L.v. Szentpály, *J. Mol. Struct. (Theochem.)*, **187** (1989) 139.
40. J.W. Cook, I. Heiger, E.L. Kennaway and W.V. Mayneard, *Proc. Roy. Soc. (London)*, **B111** (1932) 455.
41. For a discussion see :
- (a) J.C. Arcos, M.F. Argus and G. Wolf, in *Chemical Induction of Cancer*, Vol. 1, Academic Press, New York, NY, 1968, p. 414.
- (b) J.C. Arcos and M.F. Argus, *Adv. Cancer Res.*, **11** (1968) 305.
42. O. Schmidt, *Z. Physik. Chem.*, **39** (1938) 59; **42** (1939) 83; **43** (1939) 185; **44** (1939) 193.
43. O. Schmidt, *Naturwissensch.*, **29** (1941) 146.
44. N.V. Svartholm, *Arkiv Kemi Minerol. Geol.*, **A15**, No.13 (1942) 1.
45. R. Daudel, *Rev. Sci.*, **84** (1946) 37.
46. A. Pullman, *Ann. Chim.*, **2** (1947) 5.
47. P. Daudel and R. Daudel, *Biol. Med.*, **39** (1950) 201.
48. A. Pullman, *Bull. Soc. Chim. France*, **21** (1954) 595.
49. A. Pullman and B. Pullman, *Adv. Cancer Res.*, **3** (1955) 117.
50. A. Pullman and B. Pullman, in *Cancerisation par les Substances Chimique et Structure Moléculaire*, Masson, Paris, 1955.
51. For reviews : (a) W.C. Herndon, *Intl. J. Quantum Chem., Quantum Biol. Sympos.*, **1** (1974) 123.
- (b) B. Pullman, *Int. J. Quantum Chem.*, **16** (1979) 669.
52. E. Boyland, *Biochem. Soc. Symp.*, **5** (1950) 40.
53. P.L. Grover, A. Hower and P. Sims, *FEBS Lett.*, **18** (1971) 76.
54. J.K. Selkirk, E. Huberman and C. Heidelberger, *Biochem. Biophys. Res. Commun.*, **43** (1971) 1010.

- 55(a). W.M. Baird and P. Brookes, *Cancer Res.*, **33** (1973) 2378.
- 55(b). W.M. Baird, A. Dipple, P.L. Grover, P. Sims and P. Brookes, *Cancer Res.*, **33** (1973) 2386.
56. P. Sims, P.L. Grover, A. Swaisland, K. Pal and A. Hewer, *Nature*, **252** (1974) 326.
57. P. Sims and P.L. Grover, *Adv. Cancer Res.*, **20** (1974) 165.
58. H.W.S. King, M.R. Osborne, F.A. Beland, R.G. Harvey and P. Brookes, *Proc. Natl. Acad. Sci. (Wash.)*, **73** (1976) 2679.
59. For reviews see :
- (a) P. Sims and P.L. Grover, in H.V. Gelboin and P.O.P. Ts'o (eds.), *Polycyclic Hydrocarbons and Cancer*, Vol. 3, Academic Press, New York, NY, 1981, p. 117.
- (b) R.G. Harvey, *Acc. Chem. Res.*, **14** (1981) 218.
- (c) D.R. Thakker, H. Yagi, W. Levin, A.W. Wood, A.H. Conney and D.M. Jerina, in *Bioactivation of Foreign Compounds*, M.W. Anders (ed.), Academic Press, New York, NY, 1985, p. 177.
- (d) *Polycyclic Hydrocarbons and Carcinogenesis*, R.G. Harvey (ed.), ACS Symp. Series, **283**, Washington, DC, 1985.
- (e) R.G. Harvey, *Acc. Chem. Res.*, **21** (1988) 66.
60. W. Levin, A.W. Wood, R.L. Chang, S. Kumar, H. Yagi, D.M. Jerina, R.E. Lehr and A.H. Conney, *Cancer Res.*, **43** (1983) 4625; see also *ibid.*, **43** (1983) 1656.
- 61(a). F. Oesch, *Xenobiotica*, **3** (1972) 305.
- 61(b). T.M. Guenther and F. Oesch, in ref. 59a, p. 182.
62. S.K. Yang, D.W. McCourt, J.C. Leutz and H.V. Gelboin, *Science* **196** (1977) 1199.
63. A. Borgen, H. Darvey, N. Castagnoli, T.T. Crocker, R.E. Rasmussen and I.Y. Yang, *J. Med. Chem.*, **16** (1973) 502.
64. A. Dipple, in *Chemical Carcinogens*, C.E. Searle (ed.) ACS, Washington, 1976, p. 245.
65. H.S. Longuet-Higgins, *J. Chem. Phys.*, **18** (1950) 265, 275, 283.
66. A. Dipple, P.D. Lawley and P. Brookes, *Eur. J. Cancer*, **4** (1969) 493.
67. M. Nordquist, D.R. Thakker, H. Yagi, R.E. Lehr, A.W. Wood, W. Levin, A.H. Conney and D.M. Jerina, in *Molecular Basis of Environmental Toxicity*, R.S. Bhatnagar (ed.), Ann Arbor Sci. Publ., Ann Arbor, 1980, p. 329.
68. M.R. Osborne, *Cancer Res.*, **39** (1979) 4760.
69. I.A. Smith and P.G. Seybold, *Int. J. Quantum Chem., Quantum Biol. Symp.*, **5** (1978) 311.

70. M.J.S. Dewar and C.C. Thompson, *J. Am. Chem. Soc.*, **87** (1965) 4414.
- 71(a). A. Streitwieser, W.C. Langworthy and J.I. Brauman, *J. Am. Chem. Soc.*, **85** (1963) 1757, 1761.
- 71(b). A. Streitwieser, H.A. Hammond, R.H. Jagow, R.M. Williams, R.G. Jesaitis, C.J. Chang and R. Wolf, *J. Am. Chem. Soc.*, **92** (1970) 5141.
- 71(c). A. Streitwieser, P.C. Mowery, R.G. Jesaitis and A. Lewis, *J. Am. Chem. Soc.*, **92** (1970) 6529.
72. F. Oesch, in *Chemical Carcinogens*, C. Nicolini (ed.), Plenum Press, New York, NY, 1982, p. 1.
- 73(a). R.E. Keay and G.A. Hamilton, *J. Am. Chem. Soc.*, **98** (1976) 6578.
- 73(b). R.P. Hanzlik and G.O. Shearer, *Biochem. Pharmacol.*, **27** (1978) 1441.
- 73(c). Y. Sawaki, H. Kato and Y. Ogata, *J. Am. Chem. Soc.*, **103** (1981) 3832.
- 73(d). H. Mimoun, L. Saussine, E. Daire, M. Postel, J. Fisher and R. Weiss, *J. Am. Chem. Soc.*, **105** (1983) 3101.
- 73(e). M. Hawkins and L. Andrews, *J. Am. Chem. Soc.*, **105** (1983) 2527.
- 74(a). A.T. Pudzianowski and G.M. Loew, *Int. J. Quantum Chem.*, **23** (1983) 1257.
- 74(b). K. Korzekwa, W. Trager, M. Gouterman, D. Sprangler and G.H. Loew, *J. Am. Chem. Soc.*, **107** (1985) 4273.
- 75(a). B. Pullman, P.O.P. Ts'o and H.V. Gelboin (eds.), *Carcinogenesis*, D. Reidel Publ., Dordrecht, 1980.
- 75(b). P. Sims, *ibid.*, p. 33.
- 75(c). J. Deutsch, P. Okano and H.V. Gelboin, *ibid.*, p. 125.
- 75(d). J. Capdevila, Y. Saeki, R.A. Prough and R.W. Eastbrook, *ibid.*, p. 113.
76. J.E. Tomaszewski, D.M. Jerina and J.W. Daly, *Biochemistry*, **14** (1975) 2024.
77. A.H. Conney, *Cancer Res.*, **42** (1982) 4875.
78. L.C. Cusachs and R.H. Steele, *Int. J. Quantum Chem.*, **1 Suppl.** (1967) 175.
79. M.-H. Tu, D. Perry and C. Chen, *Mol. Pharmacol.*, **28** (1975) 455.
80. S.F. Mason, *J. Chem. Soc.*, (1958) 808.
81. J.C. Arcos and M. Arcos, *Progr. Drug Res.*, **4** (1962) 407.
82. K.L. Platt, E. Pfeiffer, P. Petrovic, H. Friesel, D. Beermann, E. Hecker and F. Oesch, *Carcinogenesis* (London), **11** (1990) 1721.
- 83(a). G. Klopman, *J. Am. Chem. Soc.*, **106** (1984) 7315.

- 83(b). G. Klopman, K. Namboodiri and A.N. Kalos, in *Molecular Basis of Cancer, Part A*, R. Rein (ed.), A. Liss, New York, NY, 1985, p. 287.
84. W. Levin, A.W. Wood, R.L. Chang, T.J. Slaga, H. Yagi, D.M. Jerina and A.H. Conney, *Cancer Res.*, **37** (1977) 2721.
85. W. Levin, A.W. Wood, R.L. Chang, Y. Ittah, M. Croisy-Delcey, H. Yagi, D.M. Jerina and A.H. Conney, *Cancer Res.*, **40** (1980) 3910.
86. J.M. Sayer, H. Yagi, M. Croisy-Delcey and D.M. Jerina, *J. Am. Chem. Soc.*, **103** (1981) 4970.
87. H.R. Glatt, A. Piee, K. Pauly, Th. Steinbrecher, R. Schrode, F. Oesch and A. Seidel, *Cancer Res.*, **51** (1991) 1659.
88. M.J. Le Guen and R. Taylor, *J.C.S. Perkin II*, (1974) 1274.
- 89(a). G.M. Loew, M. Poulsen, J. Ferrell and D. Chaet, *Chem. Biol. Interact.*, **31** (1980) 319.
- 89(b). M.T. Poulsen and G.H. Loew, *Cancer Biochem. Biophys.*, **5** (1981) 81.
90. P.D. Silverman, *Chem. Biol. Interact.*, **53** (1985) 313.
91. Yuemei Shang and W.C. Herndon, *Carcinogenicities of Aromatic and Alkylaromatic Hydrocarbons*, Dept. of Chemistry, UTEP, El Paso, TX 79968, USA, 1995.
92. A.A. Melikian, J.M. Leszesynska, S. Amin, S.S. Hecht, D. Hoffmann, J. Pataki and R.G. Harvey, *Cancer Res.*, **45** (1985) 1990.
93. S.S. Hecht, W.E. Bondinelli and D. Hoffmann, *J. Natl. Cancer Inst.*, **53** (1974) 1121.
94. M.H. Kim, N.E. Geacintov, M. Page, J. Pataki and R.G. Harvey, *Carcinogenesis*, **6** (1985) 121; **7** (1986) 41.
95. D.M. Jerina, J.M. Sayer, D.R. Thakker, H. Yagi, W. Levin, A.W. Wood and A.H. Conney, in ref. 75, p. 1.
96. A.M. Jeffrey, K. Jennette, S. Blobstein, I.B. Weinstein, F.A. Beland, R.G. Harvey, H. Kasai, I. Miura and K. Nakanishi, *J. Am. Chem.Soc.*, **98** (1976) 5714.
97. K.M. Straub, T.Meehan, A.L. Burlingame and M. Calvin, *Proc. Natl. Acad. Sci., U.S.A.*, **74** (1977) 5285.
- 98(a). M.R. Osborne, R.G. Harvey and P. Brookes, *Chem.-Biol. Interact.*, **20** (1978)123.
- 98(b). A.M. Jeffrey, K. Grseskowiak, I.B. Weinstein, K. Nakanishi, P. Roller and R.G. Harvey, *Science*, **206** (1979) 1309.
99. M.R. Osborne, S. Jacobs, R.G. Harvey and P. Brookes, *Carcinogenesis*, **2** (1981) 553.

- 100(a). W.M. Baird and L. Diamond, *Biochem. Biophys. Res. Commun.*, **77** (1977) 162.
- 100(b). For a review: A.M. Jeffrey, in ref. 59d, p. 187.
- 101(a). M.K. Buening, P.G. Wislocki, W. Levin, H. Yagi, D.R. Thakker, H. Akagi, M. Koreeda, D.M. Jerina and A.H. Conney, *Proc. Natl. Acad. Sci. USA*, **75** (1978) 5358.
- 101(b). T.J. Slaga, W.J. Brachen, G. Gleason, W. Levin, H. Yagi, D.M. Jerina and A.H. Conney, *Cancer Res.*, **39** (1979) 67.
102. T. Meehan and K. Straub, *Nature*, **277** (1979) 410.
- 103(a). J.C. Pelling and T.J. Slaga, *Carcinogenesis*, **3** (1982) 1135.
- 103(b). J.C. Pelling, T.J. Slaga and J. DiGiovanni, *Cancer Res.*, **44** (1984) 1081.
- 103(c). B. Jernström, P. Lycksell, A. Gråslund and B. Nordén, *Carcinogenesis*, **5** (1984) 1129.
104. R.G. Harvey, *Biochemistry*, **25** (1986) 3290.
105. R.E. Lehr, S. Kumar, W. Levin, A.W. Wood, R.L. Chang, A.H. Conney, H. Yagi, J.M. Sayer and D.M. Jerina, in ref. 59d, p. 64.
106. R. Chang, A. Wood, A.H. Conney, H. Yagi, J.M. Sayer, D.R. Thakker, D.M. Jerina and W. Levin, *Proc. Natl. Acad. Sci. USA*, **84** (1987) 8633.
107. See ref. 59e and refs. 36-43, 46-47 therein.
108. F.A. Beland, *Chem. Biol. Interactions*, **22** (1978) 329.
- 109(a). K.J. Miller, E.R. Taylor and J. Dommen, in ref. 59d, p. 239.
- 109(b). A.J. Hopfinger, in *Molecular Basis of Cancer, Part A*, R. Rein (ed.), A.R. Liss, New York, NY, 1985, p. 277.
110. B.E. Hingerty and S. Broyde, *Biopolymers*, **24** (1985) 2279.
- 111(a). A. Subbiah, S.A. Islam and G. Neidle, *Carcinogenesis*, **4** (1983) 211.
- 111(b). L.H. Pearl and S. Neidle, *FEBS Lett.*, **209** (1986) 269.
- 112(a). L.v. Szentpály and I.L. Shamovsky, *Mol. Pharmacology*, **47** (1995) 624.
- 112(b). L.v. Szentpály and I.L. Shamovsky, *Intl. J. Quantum Chem., Quantum Biol. Symp.*, **22** (1995) 191.
- 113(a). N.E. Geacintov, H. Yoshida, V. Ibanez, S.A. Jacobs and R.G. Harvey, *Biochem. Biophys. Res. Commun.*, **122** (1984) 33.
- 113(b). P. Lu, H. Jeong, R. Jankowiak, G.J. Small, S.K. Kim, M. Cosman and N.E. Geacintov, *Chem. Res. Toxicol.*, **4** (1991) 58.
- 113(c). I. Pontén, S.K. Kim, A. Gråslund, B. Nordén and B. Jernström, *Carcinogenesis*, **15** (1995) 2207.

- 113(d). B. Jernström and A. Gräslund, *Biophys. Chem.*, **49** (1994) 185.
- 113(e). S.A. Windsor, M.H. Tinker, M.R. Osborne and A. Seidel, *Carcinogenesis*, **17** (1996) 605.
- 114(a). M. Cosman, C. de los Santos, R. Fiala, B.E. Hingerty, S.B. Singh, V. Ibanez, L.A. Margulis, D. Live, N.E. Geacintov, S. Broyde and D. Patel, *Proc. Natl. Acad. Sci. USA*, **89** (1992) 1914.
- 114(b). C. de los Santos, M. Cosman, B.E. Hingerty, V. Ibanez, L.A. Margulis, N.E. Geacintov, S. Broyde and D.J. Patel, *Biochemistry*, **31** (1992) 5245.
- 114(c). M.A. Fountain and T.R. Krugh, *ibid*, **34** (1995) 3152.
115. M.E. Hogan, N. Dattagupta and J.P. Whitlock, *J. Biol. Chem.*, **256** (1981) 4504.
116. M. Ericksson, B. Nordén, B. Jernström and A. Gräslund, *Biochemistry*, **27** (1988) 1213.
117. S. Broyde and B. Hingerty, in *Molecular Basis of Cancer, Part A*, R. Rein (ed.), A. Liss, New York, NY, 1985, p. 153.
- 118(a). R.G. Harvey, M. Osborne, J.R. Connell, S. Venitt, C. Crofton-Sleigh, P. Brookes, J. Pataki and J. DiGiovanni, *The Role of Chemicals and Radiation in the Etiology of Cancer; Vol. 10*, E. Huberman and S.H. Barr (eds.), Raven Press, New York, NY, 1985, p. 449.
- 118(b). D.E. Paulus, A.S. Prakash, R.G. Harvey, M. Abramovitch and P.R. LeBreton, *Polynuclear Aromatic Hydrocarbons, Vol. 9*, M. Cooke, A.J. Dennis (eds.), Battelle, Columbus, OH, 1986, p. 745.
119. P. Pulkrabek, S. Leffler, I.B. Weinstein and D. Gruenberger, *Biochemistry*, **16** (1977) 3127.
120. M. Eriksson, B. Nordén, B. Jernström, A. Gräslund and P.O. Lycksell, *J. Chem. Soc. Chem. Comm.*, (1988) 211.
121. T.C. Boles and M.E. Hogan, *Proc. Natl. Acad. Sci. USA*, **81** (1984) 5623; *Biochemistry*, **25** (1986) 3039.
122. S. Paules, M. Cordeiro-Stone, M.J. Mass, M.C. Poirier, S.H. Yuspa and D.G. Kaufman, *Proc. Natl. Acad. Sci. USA*, **85** (1988) 2176.
123. V.V. Lobanekov, M. Plumb, G.H. Goodwin and P.L. Grover, *Carcinogenesis*, **7** (1986) 1689.
- 124(a). F.M. Chen, *Biochemistry*, **24** (1985) 5045.
- 124(b). F.M. Chen, *J. Biomol. Struct. Dyn.*, **4** (1986) 401.
125. D. Zinger, N.E. Geacintov and R.G. Harvey, *Biophys. Chem.*, **27** (1987) 131.

126. N.E. Geacintov, Report(1985), DOE/EV/04959-7, NTIS, from Energy Res. Abstr., **11** (3), (1986) Abstr. No. 6320.
127. K. Moussaoni, N.E. Geacintov and R.G. Harvey, *Biophys. Chem.*, **22** (1985) 285.
128. I. Zeger, P.O. Lycksell, A. Gräslund, M. Ericksson, B. Nordén and B. Jernström, *Carcinogenesis*, **8** (1987) 899.
129. M. Ericksson, B. Nordén, B. Jernström and A. Gräslund, *Nucleosides Nucleotides*, **7** (1988) 717.
130. J. Ladik and W. Förner, *The Beginnings of Cancer in the Cell*, Springer-Verlag, Berlin, 1994, Ch. 5.3, p. 112.
- 131(a). N.R. Drinkwater, J.A. Miller, E.C. Miller and N.C. Yang, *Cancer Res.*, **38** (1978) 3247.
- 131(b). T. Meehan, H. Gamper and J.F. Becker, *J. Biol. Chem.*, **257** (1982) 10479.
132. H. Yoshida, C.E. Swenberg and N.E. Geacintov, *Biochemistry*, **26** (1987) 1351.
133. K.L. Agarwal, T.P. Hrinyo and N.C. Yang, *Biochem. Biophys. Res. Commun.*, **114** (1983) 14.
134. N.E. Geacintov, S.E. Carberry, C.E. Swenberg and R.G. Harvey, Poster at 79th Annual Meeting, Amer. Assn. for Cancer Research, New Orleans, LA, May 1988.
- 135(a). F.A. Beland and R.G. Harvey, *J.C.S. Chem. Commun.*, (1976) 84.
- 135(b). H. Yagi, O. Hernandez and D.M. Jerina, *J. Am. Chem. Soc.*, **97** (1975) 6881.
- 136(a). N.E. Geacintov, A.G. Gagliano, V. Ibanez and R.G. Harvey, *Carcinogenesis*, **3** (1982) 247.
- 136(b). A.G. Gagliano, N.E. Geacintov, V. Ibanez, R.G. Harvey and H.M. Lee, *Carcinogenesis*, **3** (1982) 969.
137. H.H.S. Lau, S.L. Coffing, H.M. Lee, R.G. Harvey and W.M. Baird, *Chem. Res. Toxicology*, **8** (1995) 970 and refs. 12-14, 22-27 therein.
138. S.M. Fetzer, C.-R. Huang, R.G. Harvey and P.R. LeBreton, *J. Phys. Chem.*, **97** (1993) 2385.
139. S.K. Agarwal, J.M. Sayer, H.J.C. Yeh, L.K. Pannell, B.D. Hilton, M.A. Pigott, A. Dipple, H. Yagi and D.M. Jerina, *J. Am. Chem. Soc.*, **109** (1987) 2497.
140. S.A. Windsor, M.H. Tinker, M.R. Osborne and A. Seidel, *Carcinogenesis*, **17** (1996) 605.
141. M. Cosman, R. Xu, B.E. Hingerty, S. Amin, R.G. Harvey, N.E. Geacintov, S. Broyde and D.J. Patel, *Biochemistry*, **34** (1995) 6247.

- 142(a). H. Glatt, A. Seidel, W. Bochnitschek, H. Marquardt, R.M. Hodgson, P.L. Grover and F. Oesch, *Cancer Res.*, **46** (1986) 4556.
- 142(b). D.R. Phillips, H. Glatt, A. Seidel, W. Bochnitschek, F. Oesch and P.L. Grover, *Carcinogenesis*, **7** (1986) 1739.
- 143(a). S.L. Ralston, H.H.S. Lau, A. Seidel, A. Luch, K.L. Platt and W.M. Baird, *Cancer Res.*, **54** (1994) 887;
- 143(b). S.L. Ralston, A. Seidel, A. Luch, K.L. Platt and W.M. Baird, *Carcinogenesis*, **16** (1995) 2899.
144. J. Krseminski, J.-M. Lin, S. Amin and S.S. Hecht, *Chem. Res. Toxicol.*, **7** (1994) 125.
145. N.E. Geacintov, *Carcinogenesis*, **7** (1986) 759;
idem, *Comments Mol. Cell. Biophys.*, **4** (1986) 17.
146. L.v. Szentpály and P. Schwerdtfeger, in *Recent Advances of Chemistry and Molecular Biology in Cancer Research*, Q. Dai, M.-A. Armour and Q. Sheng (eds.) Science Press, Beijing and Springer-Verlag, Berlin, 1993, p. 219.
- 147(a). D.L. Whalen, J.A. Montemarano, D.R. Thakker, H. Yagi and D.M. Jerina, *J. Am. Chem. Soc.*, **99** (1977) 5522.
- 147(b). D.L. Whalen, A.M. Ross, H. Yagi, J.M. Karle and D.M. Jerina, *J. Am. Chem. Soc.*, **103** (1978) 5218.
- 147(c). D.L. Whalen, A.M. Ross, J.A. Montemarano, D.R. Thakker, H. Yagi and D.M. Jerina, *J. Am. Chem. Soc.*, **101** (1979) 5086.
- 148(a). J.M. Sayer, H. Yagi, M. Croisy-Delcey and D.M. Jerina, *J. Am. Chem. Soc.*, **103** (1981) 4970.
- 148(b). J.M. Sayer, D.L. Whalen, S.L. Friedman, A. Paik, H. Yagi, K.P. Vyas and D.M. Jerina, *J. Am. Chem. Soc.*, **106** (1984) 226.
- 149 N.B. Islam, D.L. Whalen, H. Yagi and D.M. Jerina, *J. Am. Chem. Soc.*, **109** (1987) 2108.
150. J.M. Sayer, R.E. Lehr, S. Kumar, H. Yagi, H.J.C. Yeh, G.M. Holder, C.C. Duke, J.V. Silverton, C. Gibson and D.M. Jerina, *J. Am. Chem. Soc.*, **112** (1990) 1177.
- 151(a). H. Hibshoosh, V. Ibanez, M.J. Benjamin and R.G. Harvey, *Biophys. Chem.*, **20** (1984) 121 and refs. 19-22 and 24-25 therein.
- 151(b). M.H. Kim, N.E. Geacintov, M. Pope, R.G. Harvey, *Biochemistry*, **23** (1984) 5433.
152. D.P. Michaud, S.C. Gupta, D.L. Whalen, J.M. Sayer and D.M. Jerina, *Chem. Biol. Interactions*, **44** (1983) 41.
- 153(a). M.C. McLeod, B.K. Mansfield and J.K. Selkirk, *Carcinogenesis*, **3** (1982) 1031.

- 153(b). M.C. McLeod and K.L. Zachary, *Carcinogenesis*, **6** (1985) 147.
- 153(c). M.C. McLeod and K. Zachary, *Chem. Biol. Interactions*, **54** (1985) 45.
154. A. Gräslund and B. Jernström, *Quart. Rev. Biophysics*, **22** (1989) 1.
155. S.C. Gupta, T.M. Pohl, S.L. Friedman, D.L. Whalen, H. Yagi and D.M. Jerina, *J. Am. Chem. Soc.*, **104** (1982) 3101.
156. N.B. Islam, S.C. Gupta, H. Yagi, D.M. Jerina and D.H. Whalen, *J. Am. Chem. Soc.*, **112** (1990) 6363.
157. A.A. Melikian, E.J. LaVoie, S.S. Hecht and D. Hoffmann, *Cancer Res.*, **42** (1982) 1239; *Carcinogenesis*, **4** (1983) 843.
158. N.E. Geacintov, H. Yoshida, V. Ibanez and R.G. Harvey, *Biochemistry*, **21** (1982) 1864.
159. F.M. Chen, *Anal. Biochem.*, **130** (1983) 346.
160. T. Meehan and D.M. Bond, *Proc. Natl. Acad. Sci. USA*, **81** (1984) 2635.
- 161(a). G. Lamm and G.R. Pack, *Proc. Natl. Acad. Sci. USA*, **87** (1990) 9033.
- 161(b). G.R. Pack, G.A. Garrett, L. Wong and G. Lamm, *Biophysical Journal*, **65** (1993) 1363.
- 161(c). G. Lamm, L. Wong and G.R. Pack, *Biopolymers*, **34** (1994) 227.
162. L. Wong and G.R. Pack, *Intl. J. Quantum Chem.: Quantum Biol. Symp.*, **19** (1992) 1.
- 163(a). G.R. Pack and L. Wong, *Chem Phys*. **204**(1996)279.
- 163(b). G. Lamm, L. Wong and G.R. Pack, *J. Am. Chem. Soc.* **118** (1996) 3325.
164. S.C. Gupta, N.B. Islam, D.L. Whalen, H. Yagi and D.M. Jerina, *J. Org. Chem.*, **52** (1987) 3812.
165. P. George, C.W. Bock and J.P. Gluster, *J. Chem. Phys.*, **94** (1990) 8161.
166. M. Frisch, J.M. Head-Gordon, J.A. Pople *et al*, *Gaussian 90* (revision J) Gaussian Inc., Pittsburgh, PA, 1990.
167. J.M. Sayer, H. Yagi, J.V. Silverton, S.L. Friedman, D.L. Whalen and D.M. Jerina, *J. Am. Chem. Soc.*, **104** (1982) 1972.
168. P. Schwedtfeger and L.v. Szentpály, unpublished results.
169. S.J. Weiner, P.A. Kollman, D.A. Case, K.C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, *J. Am. Chem. Soc.*, **106** (1984) 765.

170. M.J.S. Dewar, E.G. Soebisch, E.F. Healey and J.J.P. Stewart, *J. Am. Chem. Soc.*, **107** (1985) 3902.
171. S. Arnott and D.W.L. Hukins, *Biophys. Biochem. Res. Commun.*, **47** (1972) 1504.
172. S.H. Hecht, S. Amin, A.A. Melikian, E.J. Lavoie and D. Hoffmann, in ref. 59d, p.85.
173. S.K. Yang, M.W. Chou and P.P. Fu in ref. 75, p. 143.

Cycloaddition Reactions Involving Heterocyclic Compounds as Synthons in the Preparation of Valuable Organic Compounds. An Effective Combination of a Computational Study and Synthetic Applications of Heterocycle Transformations

Branko S. Jursic

Department of Chemistry, University of New Orleans, New Orleans
Louisiana 70148, USA

1. INTRODUCTION

There are numerous examples that make use of heterocycles in preparative organic chemistry but it is only rarely that aromatic heterocycles are used as mask synthons for the preparation of valuable organic compounds as part of the masked building blocks. Usually, the heterocycle constitution is preserved and, as an intact unit, it is used as a building block for the complex organic molecule that has this heterocycle unit. It is, therefore, very interesting to present computational, as well as some experimental results, in which heterocyclic compounds are used as starting materials or are used as a source for the preparation of other heterocyclic aromatic compounds *via* ring opening and ring closure. These chemical transformations are relatively simple, involve traditional synthetic procedures and are consequently much more economical. For the preparation of valuable organic compounds, resources that include aromatic heterocycles are also beneficial and may not be as obvious to the traditional synthetic organic chemist as good starting materials. By using computational tools, the synthetic chemist can be aided in accomplishing the necessary chemical transformation on the aromatic heterocycle, which is to transform the heterocycle into a valuable starting material insofar as its reactivity and selectivity toward certain chemical transformations are concerned. By using semiempirical, *ab initio*, and Density Functional Theory (DFT) methods, we wish to demonstrate that this approach could be an integral part of complicated synthetic pathways for preparation of valuable chemicals. The success of this approach, of course, depends upon a strong foundation in organic synthetic methods, computational techniques, and additionally, the creativity of the chemist. The purpose of this chapter is to offer an understanding of this combinatory approach; namely, the incorporation of heterocyclic aromatic compounds as synthons into preparative synthetic schemes.

There are certainly better synthetic schemes than the ones presented in this chapter. However, the ones used reflect the influence of computational results in the planning of the preparations of many organic compounds performed in our laboratories. They are used to demonstrate the validity of the computational (theoretical) and traditional synthetic chemistry to better understand and more easily achieve the final goal: to prepare a worthy organic compound.

2. COMPUTATIONAL METHODOLOGY

All semiempirical calculations were performed on a DEC 7620 computer. Chem-3D Plus on a Macintosh IIfx was used as a graphical interface for drawing and visualizing all structures and for preparing input files for MOPAC [1]. The transition state structures were localized, optimized, and verified as explained in our previous work [2]. All Density Functional Theory (DFT) computational studies were performed with B3LYP [3] and 6-31G(d) [4] basis set as incorporated into GAUSSIAN [5] computational package.

3. DIELS-ALDER REACTIONS WITH FIVE-MEMBERED HETEROCYCLES WITH ONE HETEROATOM

Simple heterocyclic compounds such as furan, pyrrole, and thiophene can be structural building blocks in the synthesis of a wide variety of organic compounds through their incorporation into the skeleton of the proposed organic molecule and the subsequent elimination of the heteroatom [6]. One of the simplest ways to achieve this is through Diels-Alder reactions with the previously mentioned heterocycles as dienes [7]. After the cycloaddition (Diels-Alder reaction) is performed, the heterocyclic ring of the bicycloadduct can be easily opened and transformed to another desirable functionality. For instance, furan can be opened to a 1,4-dicarbonyl compound and then closed to form a cyclopentenone ring that is a cyclic part of prostaglandins [8]. The thiophene sulfur can be oxidized and then eliminated to generate a stereospecific double bond [9]. Similar transformations can be envisioned for the cycloadduct with pyrrole [10]. Unfortunately, there are not many heterocyclic aromatic compounds which undergo Diels-Alder reactions as dienes. The main reason for the low reactivity is their high aromaticity (delocalization of π electrons, i.e., aromatic rather than double and single bonds).

3. 1. Furan, pyrrole, and thiophene as dienophiles in reaction with acetylene, ethylene, and cyclopentadiene

Aromaticity is one of the most controversial concepts in chemistry [11]. Encountering problems in directly measuring physical properties that are believed to exist, one of those properties being aromaticity, is a common dilemma in chemistry. Nevertheless, the concept of aromaticity and anti-aromaticity plays a very important role in teaching organic chemistry and in the explanation of many organic reactions, the Diels-Alder reaction being one such example [12]. Organic

chemists associate aromaticity with a special chemical reactivity: unsaturated compounds that would rather undergo a substitution reaction than an addition reaction. In terms of molecular orbital terminology, the cyclic structures that have particularly stable arrangements of occupied π -orbitals are referred to as aromatic. Here we would like to refer to aromaticity in terms of occupied π -orbitals that produce uniform chemical systems *via* the uniformity of ring's bond orders. The concept of bond order generally relates to the valence multiplicity between atoms in a molecule. If we take the carbon-carbon bonds in ethane, ethylene, and acetylene which have bond orders one, two, and three, respectively, it means that bond orders of fractions like 1.5, 1.4, 1.3, etc., represent delocalization. As long as the bond orders are uniform, the more delocalized the cycle is, the more aromatic it will be. We can, therefore, measure uniformity of cyclic systems through bond orders that are between 1 and 2.

Various definitions of bond order based on quantum mechanical theories have been proposed and correlations with bond distances have been suggested [13]. The magnetic properties of these heterocycles were also computed as a measure of their aromaticity [14]. Depending on the computational method used, the bond orders for ethane, ethene, and ethyne will not be 1, 2, and 3 respectively, but should be close to these values. Consequently, bond orders computed with the same computational methods must be considered for determining bond order uniformity of cyclic chemical systems. Nevertheless, every computational method should compute full uniformity of the bonds in the cycle for highly aromatic rings such as benzene. According to bond order calculations, benzene has an ideal aromaticity because all carbon-carbon bonds have the same bond order (1.4168). The deviation is zero and the cyclic system is fully uniform; it is, consequently, aromatic. The most uniform of these three heterocyclic compounds is pyrrole. Its deviation from uniformity was 0.81232 or 0.162464 per cyclic bond of the heterocycle, while the deviations for furan were 1.29081 and 0.258162 respectively, making it have the least uniform heterocycle (Table 1). The

Table 1. Bond orders computed by AM1 semiempirical method, their average and deviation from uniformity

Bond type	BO	BOD	BO	BOD	BO	BOD
	Furan (X=O)		Thiophene (X=S)		Pyrrole (X=N)	
X ₁ -C ₂	1.10362	0.24396	1.18005	0.18702	1.18138	0.17007
C ₂ -C ₃	1.67028	0.32269	1.61202	0.24494	1.55453	0.20307
C ₃ -C ₄	1.19010	0.15748	1.25124	0.11583	1.28543	0.06602
C ₄ -C ₅	1.67029	0.32270	1.61203	0.24495	1.55454	0.20308
C ₅ -X ₁	1.10363	0.24395	1.18005	0.18703	1.18138	0.17006
Σ	6.73794	1.29081	6.83540	0.97979	6.75727	0.81232

BO = bond order; BOD = deviation from average bond orders of a bond in the cycle. Average bond orders are 1.3479 for furan, 1.3671 for thiophene, 1.3515 for pyrrole, and 1.4168 for benzene

aromaticity of thiophene and pyrrole is very similar. Thus, there is some experimental evidence that would select pyrrole, followed by thiophene as most aromatic of these three heterocycles [15]. One may assume that these compounds are unlikely to participate in reactions that are directly involved in discrepancies in their uniformity, or at least the activation barrier for such a reaction should be substantial.

Let us now explore the uniformity of the thiophene ring as the middle range of uniformity of the three studied heterocycles (Table 1). By employing a general knowledge of chemistry, it is conceivable that a substituent could be put on the sulfur of thiophene which could substantially decrease the uniformity of the thiophene ring. If we methylate the thiophene sulfur [16] or oxidize it to a sulfoxide or sulfone [17], then the uniformity of the thiophene [18] ring is destroyed or at least diminished. Reactions that involve aromatic rings should now be feasible. To prove this assumption using the theory of bond order uniformity, we have computed bond order uniformities for those compounds (Table 2). There were a

Table 2. Bond orders computed by AM1 semiempirical method, their average and deviation from uniformity for some S-derivated thiophenes

Bond type	I		II		III	
	BO	BOD	BO	BOD	BO	BOD
S1-C2	0.96773	0.36027	0.80847	0.46199	0.60321	0.59734
C2-C3	1.83375	0.50575	1.83691	0.56648	1.89076	0.69021
C3-C4	1.03701	0.29099	1.06148	0.20897	1.01481	0.18574
C4-C5	1.83367	0.50567	1.83692	0.56647	1.89077	0.69022
C5-S1	0.96786	0.36014	0.80846	0.46199	0.60320	0.59734
Σ	6.64001	2.02282	6.35224	2.26587	6.00274	2.76085

I = S-methylthiophenium cation; II = thiophene 1-oxide; III = thiophene 1,1-dioxide; BO = bond order; BOD = deviation from average bond orders of a bond in the cycle. Average bond orders are 1.32800 for S-methylthiophenium cation, 1.27045 for thiophene 1-oxide, 1.2005484 for thiophene 1,1-dioxide, and 1.3671 for thiophene.

few excellent observations that were obtained from the bond order picture for derivatives of thiophene. Although relative in nature, bond orders were smaller than 1 for the C-S bond going from the S-methylthiophenium ion to thiophene 1,1-dioxide. This information indicated a considerable "pulling" of electron density from the thiophene ring on the S-substituent. The second most interesting observation was that localization of single and double bonds were following the same direction with maximal deviation in uniformity if the thiophene ring observed for thiophene 1,1-dioxide. These observations clearly demonstrated that those three thiophenes were engaged in chemical reactions as electron acceptors and that thiophene 1,1-dioxide was probably the closest to an ideal diene for Diels-Alder reactions that have localized single and double bonds. It is hard to compare the S-methylthiophenium cation with two neutral oxides because the charge might have a profound influence on the Diels-Alder reaction due to coulombic interactions between reactions. Nevertheless, we can state that thiophene 1,1-dioxide should

be more reactive in Diels-Alder reactions than all other heterocycles studied up to this point, with the exception of the *S*-methylthiophenium cation.

Using a traditional organic chemistry approach, one can conclude that uniformity (aromaticity) of thiophene must be diminished or even destroyed by introducing substituents on the sulfur of the thiophene atom. It is then of interest to explore the influence that a substituent has on the uniformity of the thiophene ring. The three substituents that were used, H₂N, H₃C, and F₃C, should cover electronic characteristics ranging from a strongly electron donating to an electron withdrawing capability in chemical systems. The AM1 computed bond orders and deviations from uniformity are presented in Table 3. Keeping in mind that the bond

Table 3. Bond orders computed by AM1 semiempirical method, their average and deviation from uniformity for 2,5-diaminothiophene, 2,5-dimethylthiophene, and 2,5-bis(trifluoromethyl)thiophene

Bond type	I		II		III	
	BO	BOD	BO	BOD	BO	BOD
S ₁ -C ₂	1.11840	0.20077	1.15777	0.19027	1.19573	0.14923
C ₂ -C ₃	1.55761	0.23844	1.58230	0.23426	1.49887	0.15391
C ₃ -C ₄	1.24384	0.07533	1.26005	0.08799	1.33561	0.00935
C ₄ -C ₅	1.55699	0.23782	1.58233	0.23429	1.49889	0.15393
C ₅ -S ₁	1.11902	0.20015	1.15773	0.19031	1.19572	0.14924
Σ	6.59586	0.95251	6.74018	0.93712	6.72482	0.61566

BO = bond order; BOD = deviation from average bond orders of an bond in the cycle; I = 2,5-diaminothiophene; II = 2,5-dimethylthiophene; III = 2,5-bis(trifluoromethyl)-thiophene. Average bond orders are 1.31917 for 2,5-diaminothiophene, 1.34804 for 2,5-dimethylthiophene, and 1.34496 for 2,5-bis(trifluoromethyl)thiophene in comparison with 1.3671 for thiophene.

order deviation from cycle uniformity for thiophene was 0.97979 (Table 1), it was somewhat surprising that by placing those three substituents on the thiophene ring, a more uniform cyclic bond order distribution was observed with 2,5-bis(trifluoromethyl)thiophene as the most uniform of three substituted thiophenes (Table 3). One could infer that methyl groups would increase electron density on the aromatic rings, while trifluoromethyl groups would do the opposite. Those differences must be also reflected by the sum of the ring bond orders, which was not the case. These values were almost identical (6.74 for 2,5-dimethylthiophene and 6.72 for 2,5-bis(trifluoromethyl)thiophene); however, distribution of bond orders was much more uniform in the case of 2,5-bis(trifluoromethyl)thiophene (Table 3). This indicated that 2,5-bis(trifluoromethyl)thiophene had the highest aromaticity and consequently, highest activation barrier for Diels-Alder reactions in which it was involved as a diene.

The FMO energies and bond orders for the three dienophiles that will be used throughout this computational study are presented in Table 4. Clearly, the bond orders computed with the AM1 method are what one would expect on the basis of

their structures. Furthermore, FMO energies suggest that cyclopropene should be the most reactive dienophile for the Diels-Alder reaction.

Table 4. Frontier molecular orbital energies (eV) and bond order for dienophiles used in this study.

Dienophile	HOMO	LUMO	BO ₁	BO ₂
acetylene	-11.49954	2.05318	2.964365	
ethylene	-10.55144	1.43781	2.001883	
cyclopropene	-9.81859	1.04221	1.956942	0.981596

BO₁ = bond order for the CC multiple bond in the dienophile; BO₂ = bond order for the CC single bond in cyclopropene.

Organic chemists use Frontier Molecular Orbital (FMO) [19] energies, symmetry, and coefficients to judge reactivity and stereoselectivity outcomes for many reactions, but a special usefulness of this approach was noted for pericyclic reactions [20]. According to the FMO theory, the reactant pair that has the lowest energy difference between their frontier orbitals should be the most reactive one. Let us apply this widely tested approach to challenge our conclusions obtained on the basis of the uniformity of bond orders in the ring of hetero-aromatic compounds. Frontier orbital energies for our five-membered aromatic compounds are presented in Table 5. In normal Diels-Alder reactions, the most

Table 5. Frontier molecular orbital energy gaps (eV) between reactants computed by AM1 semiempirical methods

Compound	HOMO	LUMO	ΔE_I	ΔE_{II}	ΔE_{III}	ΔE_{IV}	ΔE_V	ΔE_{VI}
Furan	-9.32	0.72	11.37	12.22	10.75	11.27	10.36	10.54
Pyrrole	-8.66	1.38	10.71	12.88	10.09	11.93	9.70	11.20
Thiophene	-9.22	0.24	11.27	11.74	10.66	10.79	10.26	10.06
2,5-Diaminothiophene	-8.20	0.23	10.26	11.72	9.64	10.77	9.25	10.03
2,5-Dimethylthiophene	-8.96	0.19	11.02	11.69	10.40	10.74	10.00	10.01
2,5-bis(trifluoromethyl)thiophene	-10.33	-1.42	12.38	10.08	11.77	9.14	11.37	8.40
S-Methylthiophenium ion	-15.28	-6.14	17.34	5.36	16.72	4.41	16.32	3.67
Thiophene 1-oxide	-9.65	-0.67	11.70	10.83	11.09	9.88	10.69	9.15
Thiophene 1,1-dioxide	-11.03	-1.42	13.08	10.08	12.47	9.14	12.07	8.40

ΔE_I = energy difference between LUMO of acetylene and HOMO of diene; ΔE_{II} = energy difference between LUMO of diene and HOMO of acetylene; ΔE_{III} = energy difference between LUMO of ethylene and HOMO of diene; ΔE_{IV} = energy difference between LUMO of diene and HOMO of ethylene; ΔE_V = energy difference between LUMO of cyclopropene and HOMO of diene; ΔE_{VI} = energy difference between LUMO of diene and HOMO of cyclopropene

important orbital interactions are between HOMO of diene and LUMO dienophile. In inverse Diels-Alder reactions, the situation is reversed; it is a LUMO diene controlled reaction. According to the FMO energy gaps presented in Table 5, furan

and pyrrole are HOMO diene controlled Diels-Alder reactions while thiophene with acetylene and ethylene as dienophiles are HOMO diene controlled. But, with a reactive dienophile such as cyclopropene, it becomes a LUMO diene controlled cycloaddition reaction. Of course, this can be changed by putting different substituents on the thiophene ring. Thus, by adding two amino groups in the 2 and 5 positions of the thiophene ring, the addition becomes strongly HOMO diene controlled or, by putting two trifluoromethyl groups on the thiophene ring, the addition becomes strongly LUMO controlled. The case is the same if sulfur is methylated or oxidized (Table 1). On the basis of the FMO energy gap, it is predicted that *S*-methylthiophenium cation is the most reactive diene for Diels-Alder reactions followed by thiophene 1,1-dioxide and thiophene 1-oxide (Table 5)

Although the FMO energy gap between reactants has been proven to be a very useful approach for the evaluation of reactivity for any reactants involved in pericyclic reactions, it has one big disadvantage: it cannot predict the outcome of the reactions that depend on secondary interactions between reactants in transition state structures such as steric repulsion and secondary orbital interactions that can substantially destabilize or stabilize the corresponding transition state structure.

Finding and optimizing transition state structures with semiempirical methods is a simple and straightforward process through which information about transition state structures can be obtained in a relatively simple fashion[21]. We have proposed using a change of FMO energy going from reactants to the transition state as a better test of reactivity [22]. For example, if the reaction is HOMO diene controlled, then the energy change of HOMO diene and LUMO dienophile into the HOMO and LUMO transition state structure is compared for a series of reaction pairs. The reaction pair that exhibits the lowest energy change is the most reactive one. To condense the number of examples, we considered only the cyclopropene addition to heterocyclic dienes. There are two possible transition state structures, *exo* and *endo* for each cycloaddition reaction. For these transition state structures, frontier orbital energies and their changes are presented in Table 6. The FMO energy change for the *S*-methylthiophenium ion cannot be compared with other heterocyclic systems because of its cationic character, which is expected to substantially increase its reactivity as a diene in Diels-Alder reactions. Basically, we have that same order of reactivity obtained with FMO energy gap, with pyrrole in the pyrrole, furan, and thiophene series being most readily involved in Diels-Alder cycloaddition reactions [16]. When comparing thiophene, 2,5-dimethylthiophene, and 2,5-bis(trifluoromethyl)thiophene, the last compound in the series is predicted to be most reactive while 2,5-dimethylthiophene is the least reactive one. This is what one would expect if the Diels-Alder reaction is LUMO diene controlled, but 2,5-diaminothiophene, having two strongly electron-donating groups, making it substantially less reactive than thiophene, was predicted to be as reactive as 2,5-bis(trifluoromethyl)thiophene (Table 6). In the series of thiophene, thiophene 1-oxide, and thiophene 1,1-dioxide, the increase in reactivity follows the same order. This could be expected on the basis of the uniformity of the ring bond order of these thiophenes (Tables 1, 2, and 3).

Table 6. Frontier molecular orbital energies for transition state structures for cyclopropene as a dienophile computed by AM1 semiempirical method

Nature of cycloaddition reaction	HOMO	LUMO	ΔE_I	ΔE_{II}	Σ
<i>endo</i> furan	-8.89	0.56	0.43	-0.48	0.91
<i>exo</i> furan	-8.76	0.61	0.56	-0.43	0.99
<i>endo</i> pyrrole	-8.57	0.82	0.09	-0.22	0.31
<i>exo</i> pyrrole	-8.52	0.85	0.14	-0.19	0.33
<i>endo</i> thiophene	-8.38	0.37	1.44	-0.13	1.57
<i>exo</i> thiophene	-8.53	0.37	1.29	-0.13	1.42
<i>endo</i> 2,5-diaminothiophene	-8.22	0.34	0.02	-0.70	0.72
<i>exo</i> 2,5-dimethylthiophene	-8.24	0.35	0.04	-0.69	0.73
<i>endo</i> 2,5-dimethylthiophene	-8.29	0.33	1.53	0.14	1.67
<i>exo</i> 2,5-dimethylthiophene	-8.45	0.33	1.37	0.14	1.51
<i>endo</i> 2,5-bis(trifluoromethyl)thiophene	-9.19	-1.08	0.63	0.34	0.97
<i>exo</i> 2,5-bis(trifluoromethyl)thiophene	-9.39	-1.07	0.43	0.35	0.78
up <i>endo</i> <i>S</i> -methylthiophenium ion	-13.53	-5.41	-3.71	2.43	6.14
down <i>endo</i> <i>S</i> -methylthiophenium ion	-13.27	-5.60	-3.45	2.69	6.14
up <i>exo</i> <i>S</i> -methylthiophenium ion	-13.34	-5.54	-3.52	2.62	6.14
down <i>exo</i> <i>S</i> -methylthiophenium ion	-13.09	-5.70	-3.27	2.87	6.14
up <i>endo</i> thiophene 1-oxide	-8.99	-0.41	0.83	0.26	1.09
down <i>endo</i> thiophene 1-oxide	-8.72	-0.28	1.10	0.39	1.49
up <i>exo</i> thiophene 1-oxide	-8.99	-0.41	0.83	0.26	1.09
down <i>exo</i> thiophene 1-oxide	-8.55	-0.40	1.27	0.27	1.54
<i>endo</i> thiophene 1,1-dioxide	-9.58	-1.02	0.24	0.40	0.64
<i>exo</i> thiophene 1,1-dioxide	-9.35	-1.13	0.47	0.29	0.76

ΔE_I = HOMO energy change going from reactants to transition state; ΔE_{II} = LUMO energy change going from reactants to transition state; $\Sigma = \Delta E_I + \Delta E_{II}$

For many Diels-Alder reactions, the FMO energy change going from reactants to the transition state structure predicts that addition of cyclopropene to thiophene would produce an *endo* cycloadduct, which is not necessarily true. For instance, it was experimentally confirmed that cyclopropene adds to furan derivatives, facilitating the formation of the *exo* cycloadduct [23]. This discrepancy of the FMO energy difference prediction from experimental data might be due to Secondary Orbital Interactions (SOI) that substantially stabilize transition state structures but are not reflected in FMO energies. The same explanation might be used to support the prediction that pyrrole is the most reactive of the three heterocycles, thiophene, pyrrole, and furan, despite the clear experimental evidence that furan is, by far, the most reactive as a diene for Diels-Alder reactions [24].

To clarify these observations, we computed bond orders for transition states and selected two SOI that were responsible for stabilizing transition state structures. For the *endo* addition of cyclopropene, the SOI between the methylene

hydrogen and the π -bond of the heterocycle were stabilizing interactions. For the *exo* transition state structure, the stabilizing interaction was the SOI between the methylene hydrogen of cyclopropene and the lone pair of a hetero atom of the heterocycle moiety of the transition state structure [25]. The results, with transition state structures formed with cyclopropenes as dienophiles and furan, pyrrole, and thiophene as dienes, are presented in Table 7. It is obvious that

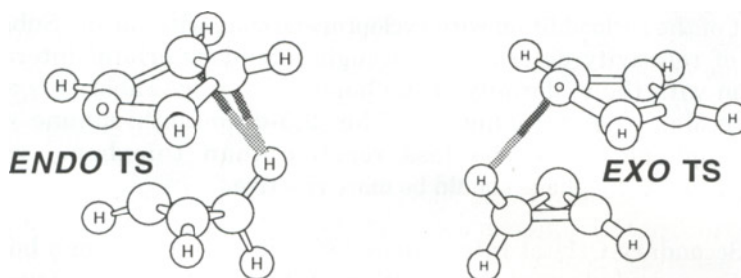


Table 7. Secondary Orbital Interactions (SOI) presented through bond orders between diene and dienophile moieties of transition state structures computed by AM1 semiempirical method

TS	Furan	Pyrrole	Thiophene	NH ₂ -Thio	CH ₃ -Thio	CF ₃ -Thio
<i>endo</i>	0.003553	0.003321	0.004168	0.003480	0.004059	0.004849
<i>exo</i>	0.004246	0.002790	0.014973	0.012703	0.014755	0.020637

^aThe transition state structure is presented only for furan but all other transition state structures can be derived from these two by adding corresponding substituents; NH₂-Thio = 2,5-diaminothiophene; CH₃-Thio = 2,5-dimethylthiophene

secondary orbital interactions are substantially higher in the case of the cyclopropene addition to furan than to pyrrole. Furthermore, SOI interactions in the *exo* transition state structure were much higher than in the *endo* transition state structure with furan as a diene (Table 7). It would not be surprising if furan is a better diene than pyrrole, and formation of *exo* cycloadduct is preferred. These two heterocycles, furan and pyrrole, can be compared because they have comparable sizes, while thiophene is substantially larger. For instance, the covalent radii of oxygen and nitrogen are 0.66 and 0.70 Å, respectively [26]. On the other hand, the sulfur radius is 1.04 Å. Therefore, only the SOI can be compared between substituted thiophene because the increase in the SOI in thiophene in comparison with furan could be entirely an effect of the larger size of the heteroatom, not as an effect of stabilization interactions between the reactants moieties of the transition state structure. If we compare bond order uniformity presented in Table 1, it can be seen that thiophene is actually less reactive than both furan and pyrrole.

A comparison of reactivity of substituted thiophenes through the SOI in their transition state structures with cyclopropene as the dienophile reveals some interesting features. The secondary orbital interactions increase in the following order: 2,5-diaminothiophene, 2,5-dimethylthiophene, thiophene, and 2,5-bis(trifluoromethyl)thiophene. It is very hard to determine if SOI in the *exo* transition state structures are stronger or weaker in comparison with SOI in the *endo* transition state structures. Considering that sulfur is much larger, it is reasonable to assume that the *endo* transition state for sulfur is more likely to be the product of the cycloaddition with cyclopropene than with furan. Subsequently, the order of reactivity determined through secondary orbital interactions in combination with the uniformity of the heterocycle ring (Table 1) is as follows: furan, pyrrole, and thiophene. The 2,5-diaminothiophene and 2,5-dimethylthiophene should be less reactive than thiophene, while 2,5-bis(trifluoromethyl)thiophene should be more reactive.

Table 8. Secondary Orbital Interactions (SOI) presented through bond orders between diene and dienophile moieties of transition state structures for cyclopropene addition to thiophene, thiophene 1-oxide, and thiophene 1,1-dioxide. The bond orders are computed by AM1 semiempirical method

TS	thiophene	thiophene 1-oxide		thiophene 1,1-dioxide
		O-up	O-down	
<i>endo</i>	0.004168	0.006849	0.006515	0.008546
<i>exo</i>	0.014973	0.005466	0.000806	0.000512

Let us now explore secondary orbital interactions in thiophene oxides. There are many possibilities in regard to the position of oxygen (Table 8). There is a substantial increase in SOI between two of the reactants' moieties of *endo* transition state structures. This observation is in full agreement with our previous observation that larger heteroatoms actually push dienophile moiety towards stronger SOI in an *endo* transition state structure. Conversely, SOI between reactants are substantially diminished in an *exo* transition state structure which is expected based on the nature of the sulfur atom in the thiophene oxides. One may also speculate that SOI can be present between the oxygen atom of thiophene 1-oxide or thiophene 1,1-oxide with the methylene hydrogen of cyclopropane moieties, but computed values were too small (0.005455) when compared with the one in unoxidized thiophene (0.014973) to have any significant influence on the reaction outcome. In thiophene 1,1-dioxide, SOI in the *endo* transition state structure is an indication that this should be a very reactive diene in Diels-Alder reactions.

Finally, we can turn our attention to the SOI interaction in transition state between cyclopentadiene as a dienophile and the *S*-methylthiophenium ion as a diene for Diels-Alder reactions. Theoretically, there are four different transition state structures in regard to position of the methyl group, as well as the dienophile in regard to the thiophenium ring. One can assume that because of the location of a positive charge on the sulfur atom of the thiophenium ion ring, the SOI in an *exo*

transition state structure should be negligible in comparison with SOI in an *endo* transition state structure. Therefore, formation of the *endo* cycloadduct between the *S*-methylthiophenium ion and cyclopentadiene is inevitable. Computed SOI fully agreed with these observations. SOI in the *up-endo* transition state structure was 0.009439, much higher than in thiophene and slightly higher than in thiophene 1,1-dioxide. At the same time, SOI in the *up-exo* transition state structure was only 0.004674, which was much smaller than in the case of thiophene and thiophene 1-oxide. In this study, we have selected the *S*-methylthiophenium ion as more reactive than thiophene 1,1-dioxide.

Our studies performed with the AM1 semiempirical method are of a more qualitative nature. Through comparison of bond orders of the heterocycles, FMO energy differences between reactants, FMO energy changes going from reactants to the transition state structures, and through SOI, we were able to determine the relative reactivity and reaction outcome in these cycloaddition reactions, but were not capable of actually determining if these cycloadditions are experimentally feasible. The best way to determine that is through computation of the transition state structure [27] for each and every reaction pair. We have demonstrated that AM1 semiempirical methods tend to produce activation barriers that are in a very narrow range [28]. Nevertheless, AM1 computational methods can be used to determine relative energies for a series of compounds. We have successfully used them for evaluation of the relative reactivity in many cycloaddition reactions [29]. Reaction barriers that are much closer can be obtained through a combination of hybrid DFT and AM1 semiempirical calculations such as B3LYP/6-31G(d)//AM1 [30] or by full B3LYP/6-31G(d) evaluation of activation barriers [31].

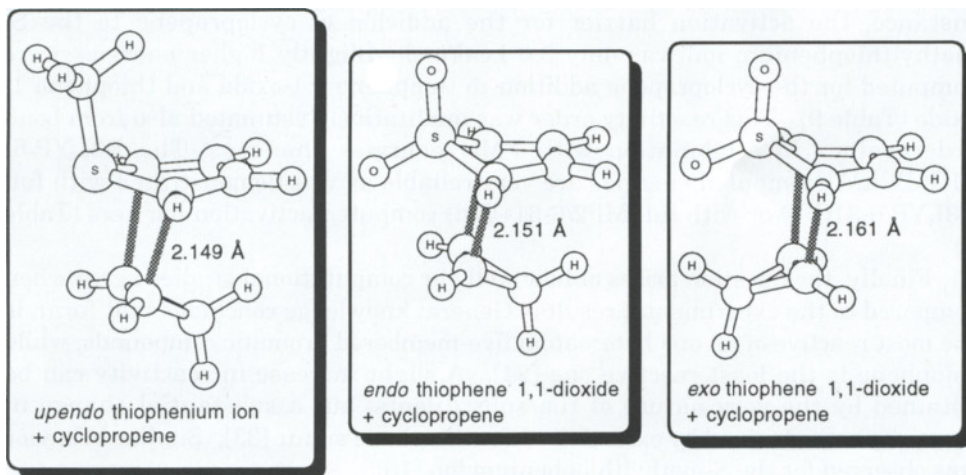


Figure 1. Some representative transition state structures computed by the AM1 semiempirical method.

All computed transition state structures are for symmetric and synchronous formation of both C-C bonds. Some representations of the series of compounds are presented in Figure 1. The bond distances of the newly-forming C-C bonds varied from 2 to almost 2.2 Å. They are characteristic transition state structures for Diels-Alder reactions [32]. The bond distances varied slightly for the two isomeric transition state structures, the *exo* and *endo* cyclopropene addition to thiophene 1,1-dioxide (Figure 1), demonstrating the stabilizing influence of steric repulsion interactions or SOI in *exo* and/or *endo* transition state structures.

Let us now evaluate the activation barriers for cycloaddition reactions (Table 9). According to AM1 computational studies, furan has a lower activation barrier than pyrrole, and pyrrole has a lower activation barrier than thiophene for cycloaddition reactions. This was the exact order of reactivity determined on the basis of the FMO changes going from reactant to transition state structures. For the cyclopropene addition to furan, the *exo* cycloadduct has a lower activation barrier and therefore, it should be the preferred product. For the cyclopropene addition to pyrrole, the computed activation barriers for two isomeric transition state structures were very close, indicating formation of a mixture of the products. On the other hand, as speculated on the basis of bond orders in the transition state structures between cyclopropene and thiophene, the *endo* transition state had a slightly lower activation barrier than its *exo* isomer.

The two substituted thiophenes (2,5-diaminothiophene and 2,5-dimethylthiophene), have higher activation barriers than thiophene, although 2,5-bis(trifluoromethyl)thiophene is more reactive. According to the B3LYP/6-31G(d)//AM1 computed activation barrier, all of their reactions might require forceful conditions. Our assumption that the aromaticity of the thiophene ring was responsible for low reactivity was confirmed by the computation of very low activation barriers for thiophene derivatives that have lower aromaticity. For instance, the activation barrier for the addition of cyclopropene to the *S*-methylthiophenium ion was only 5.0 kcal/mol. Slightly higher energies were computed for the cyclopropene addition to thiophene 1,1-oxide and thiophene 1-oxide (Table 9). This reactivity order was qualitatively estimated also from bond order values in combination with FMO energies changes. The B3LYP/6-31G(d)//AM1 computed energies are very reliable as was demonstrated with full B3LYP/6-31G(d) or with full MP2/6-31+G(d) computed activation barriers (Table 9)

Finally, the question arises of how well our computational studies agree when compared to the experimental results. General knowledge concludes that furan is the most reactive of all one heteroatom five-membered aromatic compounds, while thiophene is the least reactive one [24]. A slight increase in reactivity can be obtained by the very nature of the substituents, but a substantial change in reactivity was obtained by oxidation of the thiophene sulfur [33]. Similar behavior was observed for the *S*-methylthiophenium ion [16].

Table 9. Activation barrier for Diels-Alder reactions with heterocycles as dienes computed using the AM1 semiempirical (ΔE_I) and B3LYP/6-31G(d)//AM (ΔE_{II}) hybrid DFT approach

Cycloaddition reaction	ΔE_I	ΔE_{II}	ΔE_{III}
furan + acetylene	36.5	30.2	
furan + ethylene	28.1	27.0	
<i>endo</i> furan + cyclopropene	27.4	18.7	16.9 ^a
<i>exo</i> furan + cyclopropene	26.1	18.4	16.2 ^a
pyrrole + acetylene	41.5	33.5	
pyrrole + ethylene	33.5	32.1	
<i>endo</i> pyrrole + cyclopropene	31.7	22.4	
<i>exo</i> pyrrole + cyclopropene	30.3	22.1	
thiophene + acetylene	50.4	39.0	
thiophene + ethylene	43.1	35.3	36.7 ^b
<i>endo</i> thiophene + cyclopropene	40.5	25.2	
<i>exo</i> thiophene + cyclopropene	40.0	26.8	
<i>endo</i> 2,5-diaminothiophene + cyclopropene	40.5	22.3	
<i>exo</i> 2,5-diaminothiophene + cyclopropene	39.2	22.9	
<i>endo</i> 2,5-dimethylthiophene + cyclopropene	43.4	23.0	
<i>exo</i> 2,5-dimethylthiophene + cyclopropene	43.1	24.6	
<i>endo</i> 2,5-bis(trifluoromethyl)thiophene + cyclopropene	44.6	21.7	
<i>exo</i> 2,5-bis(trifluoromethyl)thiophene + cyclopropene	43.8	23.5	
<i>up</i> S-methylthiophenium ion + ethylene	24.6	16.8	18.2 ^c
<i>down</i> S-methylthiophenium ion + ethylene	30.0	19.7	22.8 ^c
<i>up-endo</i> S-methylthiophenium ion + cyclopropene	21.5	5.0	
<i>down-endo</i> S-methylthiophenium ion + cyclopropene	27.4	9.1	
<i>up-exo</i> S-methylthiophenium ion + cyclopropene	25.9	12.1	
<i>down-exo</i> S-thiophenium ion + cyclopropene	34.5	23.7	
<i>up-endo</i> thiophene 1-oxide + cyclopropene	29.3	7.2	
<i>up-exo</i> thiophene 1-oxide + cyclopropene	32.2	16.8	
<i>down-exo</i> thiophene 1-oxide + cyclopropene	33.8	12.6	
<i>endo</i> thiophene 1,1-dioxide + cyclopropene	27.8	8.0	
<i>exo</i> thiophene 1,1-dioxide + cyclopropene	33.8	17.4	

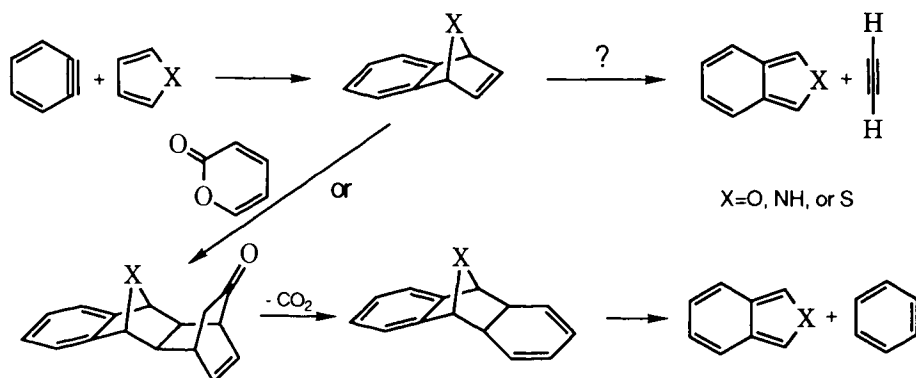
^aComputed with B3LYP/6-31G(d)//B3LYP/6-6-31G(d) theory model. ^bComputed with MP3/6-31+G(d)//MP2/6-31+G(d) [16]. ^cActivation barriers for ethylene addition to S-methylthiophenium ion computed with MP3/6-31+G(d)//MP2/6-31+G(d) [16].

3. 2. Addition of benzyne to furan, pyrrole, and thiophene

As mentioned above, a major problem in using heterocyclic aromatic compounds as dienes in the Diels-Alder reaction is their highly aromatic character which hinders their involvement in reactions that include direct participation of their π -bonds. Some heterocycles such as pyrrole would, in many instances, rather participate in Michael-addition type of reactions than in Diels-Alder

reactions. The reactions can be enforced if very reactive dienophiles are used, or if heterocyclic compounds can be properly selected and derivatized [34]. One of the most reactive dienophiles for Diels-Alder reactions is benzyne. Benzyne as an intermediate was proposed by Roberts in 1953 [35]. He studied the reaction between chlorobenzene and potassium amide which yields aniline. When chlorobenzene-1- ^{14}C is used as the starting material, approximately 50 percent of the ^{14}C in the product is found in the 1-position and approximately 50 percent in the 2-position. The overall substitution must be achieved by an elimination-addition mechanism with benzyne as an intermediate [36]. The existence of benzyne as an intermediate in the reaction was also directly demonstrated by a trapping experiment in the Diels-Alder reaction with anthracene, leading triptycene [36]. Since then, benzyne and its derivatives were used as very powerful dienophiles for Diels-Alder reactions [37].

An example of cycloaddition reactions that involve transformation of five-membered heterocycles with one heteroatom into benzo[*c*]-fused heterocycles through a sequence of Diels-Alder reactions is presented in Scheme 1.



Scheme 1. Two proposed pathways for the preparation of benzo[*c*]-fused heterocycles

According to our postulate, the more uniform the bond orders in the ring are, the less willing the aromatic compound is to participate in Diels-Alder reactions as a diene. Let us now apply this approach in order to compare the reactivity of benzene and benzyne with three five-membered aromatic heterocyclic compounds. By definition and also by our AM1 semiempirical calculations, there is no bond order deviation from uniformity in the benzene ring. On the other hand, benzene represents an ideal aromatic system with full π -bond delocalization that is reflected in its ring bond order uniformity (Table 10). The highest deviation of bond orders from uniformity was obtained for benzyne when a π -orbital of a triple bond was also included. That is, of course, not an entirely proper approach because this π -orbital belongs to the carbon skeleton ring and therefore is perpendicular to the six *p*-atomic orbitals of benzene. When this bond was excluded, then a very uniform aromatic system with a computed bond order deviation of only 0.19 was

achieved. Nevertheless, by including triple bond π -orbitals in computation of the bond order, the bond order deviation was very high, indicating centers of disruption as C(1) and (C2), the most reactive of the four aromatic compounds. The results are presented in Table 1. For the aromatic heterocycle series, furan was found to have the least uniform bond order distribution; we, therefore, predict that it will be the most reactive in reactions that involve direct participation of the furan π -bond.

Table 10. Bond orders and bond order deviation from average bond orders of some aromatic compounds

Compounds	BO ₁	BO ₂	BO ₃	BO ₄	BO ₅	BO ₆	SOB	BOD
benzene	1.417	1.417	1.417	1.417	1.417	1.417	8.502	0.0
furan	1.104	1.670	1.190	1.670	1.104		6.738	1.290
pyrrole	1.181	1.555	1.285	1.555	1.181		6.757	0.814
thiophene	1.180	1.612	1.251	1.612	1.180		6.835	0.980
benzyne	2.359	1.346	1.422	1.397	1.422	1.346	9.292 ^a	1.622 ^a

BO = bond order of ring's bond; SBO = sum of bond orders; BOD = sum of bond order deviation from an average ring bond order. The second π -bond is not a part of the ring π -bonds and should be excluded; if BO₁ = 1.359 then SBO = 8.292 and BOD = 0.19

The differences between pyrrole and thiophene were very small. According to calculations presented in Table 1, one might conclude that pyrrole is more aromatic due to the predicted higher bond order uniformity of the ring. But it has to be noted that AM1 predicted a C-heteroatom bond that was too short for thiophene and slightly longer C-heteroatom bonds for furan and pyrrole. For instance, the AM1 computed C-O, C-N, and C-S bond distances that were 1.395, 1.391, and 1.672 Å, while experimental values were 1.362, 1.370, and 1.714 Å. Therefore, pyrrole and furan have the same error in uniformity that was eliminated by comparing them. Conversely, the computational error for thiophene was in the opposite direction, producing higher bond deviation than it should be. Consequently, we cannot firmly state that pyrrole has a higher stability due to its ring bond order uniformity in comparison with thiophene.

By examining bond order uniformity, it is clear that benzyne is quite a reactive species if a triple bond is involved in the reaction, but not very reactive if a double bond is involved in the reaction. The furan ring had the highest bond order deviation and, as a result, it should be the most reactive five-membered heterocycle with one heteroatom studied here.

Let us now explore the reactivity of these compounds as dienes (furan, pyrrole, thiophene) with a dienophile (benzyne) in Diels-Alder reactions. One approach that, for a long time, has been widely employed by chemists, is the use of Frontier Molecular Orbital (FMO) [19] energy gap between two of the reactants. According to this theory, the most reactive reactant pair will be the one that has a lower FMO energy gap. The reaction is predicted to be HOMO diene-controlled. If

we consider the FMO energy difference between reactants as a measure of their reactivity, very controversial results can be obtained (Table 11). In this study, benzene was found to be a better dienophile for Diels-Alder addition than ethylene. The most reactive heterocycle as diene for Diels-Alder reaction was pyrrole and then thiophene. The least reactive was furan. An opposite prediction was made on the basis of bond order uniformity (Table 10). These findings also contradicted the basic principles of organic chemistry which contend that benzene does readily participate in cycloaddition reactions; and another argument is that furan is commonly used as a diene for Diels-Alder reactions, as it is one of the most reactive aromatic heterocycles. That definitively demonstrates that the FMO energy gap difference between reactants as a measure of reactivity should be regarded with caution.

Table 11. Frontier Molecular Orbital (FMO) energies (eV) and the energy gaps (eV) between a heterocycle and benzyne

Reactant	HOMO	LUMO	ΔE_I	ΔE_{II}	ΔE_{III}	ΔE_{IV}	ΔE_V	ΔE_{VI}
ethylene	-10.551	1.438						
benzene	-9.653	0.554						
benzyne	-9.909	-0.698						
furan	-9.317	0.723	10.755	11.274	9.871	10.376	8.619	10.632
pyrrole	-8.657	1.378	10.095	11.929	9.211	11.031	7.959	11.287
thiophene	-9.218	0.238	10.656	10.789	9.772	9.891	8.520	10.147

$\Delta E_I = \text{LUMO}_{\text{ethylene}} - \text{HOMO}_{\text{heterocycle}}$; $\Delta E_{II} = \text{LUMO}_{\text{heterocycle}} - \text{HOMO}_{\text{ethylene}}$;
 $\Delta E_{III} = \text{LUMO}_{\text{benzene}} - \text{HOMO}_{\text{heterocycle}}$; $\Delta E_{IV} = \text{LUMO}_{\text{heterocycle}} - \text{HOMO}_{\text{benzene}}$;
 $\Delta E_V = \text{LUMO}_{\text{benzyne}} - \text{HOMO}_{\text{heterocycle}}$; $\Delta E_{VI} = \text{LUMO}_{\text{heterocycle}} - \text{HOMO}_{\text{benzyne}}$

Let us now employ our approach of necessary FMO energy changes for transformation of reactants into corresponding transition state structures. Transition state structures, using the example of furan as a diene, are presented in Figure 2. It is interesting to note that

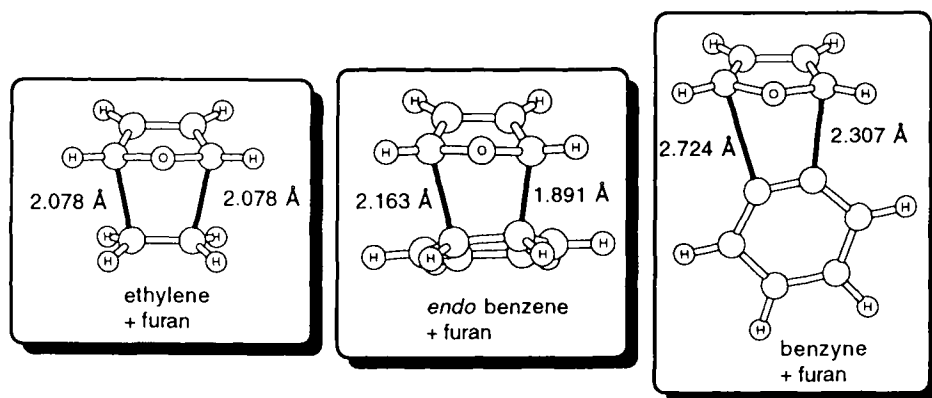


Figure 2. Some transition state structures for the dienophile addition to furan.

transition state structure for the addition of benzyne to furan is not symmetrical, while both transition state structures for ethylene and for benzyne addition have a plane of symmetry (concerted synchronous mechanisms of Diels-Alder addition). Optimization of the transition states with benzyne as a dienophile was not possible without using the keyword "biradical", indicating that the transition state might have biradicaloid character.

The necessary FMO energy change for transformation of reactants into transition state structures that will be reviewed are shown in Table 12. As defined earlier, the reactant pair that requires smaller changes in frontier orbital energy changes should form the transition state with lower energy, (smaller activation barrier), and therefore, should be the most reactive reactant pair. The sum of the FMO energy changes undoubtedly shows that the reaction is HOMO diene and LUMO dienophile controlled making the sum of FMO energy changes, SE_1 , substantially smaller than for other possible combinations of FMO energies, SE_2 (Table 11). The order of reactivity for dienophiles is benzyne, ethylene, and then benzene; this order would be expected on the basis of our common knowledge of organic chemistry. Furthermore, for reactions between furan and benzyne, lower FMO orbital energy changes are required to reach the transition state than for thiophene and benzyne. This is what one would expect from our previous theoretical, as well as experimental results. In this way, it was demonstrated that FMO energy changes required for transformation of reactants into their corresponding transition state structures certainly represent a more reliable approach for assessment of reactivity than simple FMO energy gap between reactants.

Table 12. Frontier Molecular Orbital energy change (eV) necessary to transform reactants into transition state structures

TS	HOMO	LUMO	ΔE_I	ΔE_{II}	ΔE_{III}	ΔE_{IV}	Σ_1	Σ_2
A	-9.037	0.791	0.280	1.514	0.068	-0.647	0.927	1.582
B	-7.987	-0.103	1.330	1.666	-0.826	-0.657	1.987	2.492
C	-8.751	-0.543	0.566	1.158	-1.266	0.155	0.721	1.424
D	-8.468	-0.496	0.750	1.441	-0.734	0.202	0.952	2.175

A = furan + ethylene; B = *endo* furan + benzene; C = furan + benzyne; D = thiophene + benzyne; $\Delta E_I = \text{HOMO}_{\text{TS}} - \text{HOMO}_{\text{diene}}$; $\Delta E_{II} = \text{HOMO}_{\text{TS}} - \text{HOMO}_{\text{dienophile}}$; $\Delta E_{III} = \text{LUMO}_{\text{TS}} - \text{LUMO}_{\text{diene}}$; $\Delta E_{IV} = \text{LUMO}_{\text{TS}} - \text{LUMO}_{\text{dienophile}}$; $\Sigma_1 = |\Delta E_I| + |\Delta E_{IV}|$; $\Sigma_2 = |\Delta E_{II}| + |\Delta E_{III}|$

The question that we should answer now is not only one concerning the relative reactivity of these dienes and dienophiles, but additionally, whether the addition of these dienophiles (ethylene, benzene, and benzyne) to furan and thiophene is experimentally feasible. That can only be properly addressed by computing activation barriers. Semiempirical methods tend to produce very similar and very narrow differences in activation barriers for different diene-dienophile reactant

pairs with different reactivities. This problem can be remedied by computing B3LYP/6-31G(d) activation barriers or computing activation barriers with the less expensive B3LYP/6-31G(d)//AM1 computational approach. We have demonstrated that this computational approach produced activation barriers that were as accurate as full B3LYP/6-31G(d) calculations. Computed activation barriers for these reactions are presented in Table 13. It was clearly demonstrated that the addition of

Table 13. Computed activation barriers for Diels-Alder reactions

reaction species	HOF	E	ΔE_I	ΔE_{II}
ethylene	16.5	-78.58702		
benzene	22.0	-232.24803		
benzyne	140.5	-230.90903		
furan	3.0	-230.01791		
thiophene	27.4	-552.99828		
ethylene + furan	47.5	-308.56192	28.0	27.0
<i>endo</i> benzene + furan	74.0	-462.18295	49.0	52.1
benzyne + furan	147.8	-460.92582	4.3	0.7
benzyne + thiophene	179.6	-783.89440	11.7	8.1

HOF = heat of formation (kcal/mol) computed by AM1; E = total energy (a.u.) computed with B3LYP/6-31G(d)//AM1; ΔE_I = activation barrier (kcal/mol) computed by AM1; ΔE_{II} = activation barrier (kcal/mol) computed with B3LYP/6-31G(d)//AM1.

benzyne to both furan and thiophene had very low activation barriers and should be carried out at room temperature. The addition of ethylene in comparison to the addition of benzene to furan had a lower activation barrier, as was predicted on the basis of the necessary change in FMO energy in transformation of reactants into corresponding transition state structures.

In regard to transformation of the benzyne cycloadduct product between benzyne and furan into benzo[*c*]furan, the direct transformation by elimination of acetylene will not occur because computed activation barriers for this reaction were too high. The AM1 computed activation barrier for acetylene elimination was 57.0 kcal/mol, while the activation barriers for the addition of α -pyrone (30.4 kcal/mol) and the elimination of carbon dioxides and benzene (29.9 kcal/mol) were much more energetically favorable (Scheme 1). These computational results are in full agreement with experimental evidence [38].

3. 3. Cycloaddition reactions with pyrrole as a diene for Diels-Alder reaction

The discovery of the structure of 7-azabicyclo[2.2.2]heptene isolated from the Ecuadorian poison frog, *Epipedobates tricolor* [39], has caused a rebirth of investigation of Diels-Alder reactions with pyrrole as a diene [40]. Subsequently, a large number of 7-azabicyclo[2.2.1]-heptane and 7-azabicyclo[2.2.1]hept-2-ene derivatives have been synthesized and protected by patents [41]. We have

previously explored the possibility of using pyrrole as a diene for the Diels-Alder reaction [42]. It was demonstrated that the reaction is HOMO pyrrole controlled with the addition of classical dienophiles such as ethylene, as well as heterodienophiles such as nitrosyl hydride or oxygen. Although Frontier Molecular Orbital (FMO) energy gaps for the reaction of pyrrole with those dienophiles are smaller than with cyclopentadiene, the computed activation barriers are substantially higher, and the reactions are not experimentally feasible [42]. Because the FMO energy gap considers orbital energies for separated reactions, it is not appropriate for systems in which substantial electronic interactions exist in the transition state structures. Therefore, we have explained the low reactivity of pyrrole, in reactions with hetero dienophiles, as the repulsion due to the lone pair effect in their transition state structures [42]. Of course, this is true for heteroatoms, but low reactivity of pyrrole toward dienophiles such as acetylene, ethylene and its derivatives cannot be fully explained on the basis of orbital repulsion interaction in the transition states.

Table 14. The FMO energies (eV) and FMO energy gap between pyrrole derivatives and acetylene, and cyclopentadiene and acetylene computed by AM1 semiempirical methods

Comp.	HOMO	LUMO	A	B	C	D	E	F
I	-9.079	0.481	11.132	11.981	8.004	11.909	10.528	10.198
II	-8.657	1.378	10.710	12.877	7.582	12.805	10.106	11.095
III	-8.222	1.265	10.276	12.764	7.148	12.692	9.671	10.982
IV	-10.224	-0.424	12.278	11.074	9.150	11.002	11.673	9.293
V	-9.525	-0.792	11.578	10.707	8.451	10.634	10.974	8.925
VI	-10.894	-1.396	12.947	10.103	9.819	10.031	12.343	8.321
VII	-8.606	1.331	10.659	12.830	7.532	12.823	10.055	11.048
VIII	-9.345	0.222	11.398	11.721	8.270	11.649	10.794	9.939
IX	-9.204	-0.039	11.257	11.460	8.129	11.387	10.653	9.678
X	-9.560	-1.334	11.613	10.164	8.485	10.092	11.009	8.383
XI	-8.687	0.000	10.740	11.500	7.612	11.428	10.136	9.717
XII	-13.903	-4.299	15.956	7.200	12.829	7.128	15.352	5.418
XIII	-15.146	-5.520	17.199	5.979	14.071	5.907	16.595	4.197

I = cyclopentadiene; II = pyrrole; III = 2,5-dimethylpyrrole; IV = 2,5-di(trifluoromethyl)pyrrole; V = 2,5-diformylpyrrole; VI = 2,5-disulfonylpyrrole; VII = 1-methylpyrrole; VIII = 1-trifluoromethylpyrrole; IX = 1-formylpyrrole; X = 1-sulfonylpyrrole; XI = 1-formyl-2,5-dimethylpyrrole; XII = lithium-pyrrole; XIII = pyrrolium ion; A = LUMO_{acetylene} - HOMO_{diene}; B = LUMO_{diene} - HOMO_{acetylene}; C = LUMO_{diformylacetylene} - HOMO_{diene}; D = LUMO_{diene} - HOMO_{diformylacetylene}; E = LUMO_{dimethoxyacetylene} - HOMO_{diene}; F = LUMO_{diene} - HOMO_{methoxyacetylene}.

Let us first compare the reactivity of pyrrole derivatives with cyclopentadiene, on the basis of FMO energy gap, with acetylene as a dienophile (Table 14). According to the FMO energy gap between reactants, the acetylene addition to the

majority of pyrrole derivatives is neither strongly LUMO nor HOMO dienophile controlled. Except for the pyrrolium ion that should be considered separately from this group because of its positive charge, all other pyrrole derivatives in reaction pairs with acetylene have a FMO energy gap between 10-13 eV. This is too high for a cycloaddition to be experimentally feasible. On the other hand, due to the exceptionally low LUMO energy of the pyrrolium ion, the FMO gap between this cation and acetylene was predicted to be only 5.979 eV (Table 14). Therefore pyrrole, and substituted pyrroles, do not exhibit either strong HOMO or LUMO diene controlled Diels-Alder reactions due to their high aromatic character. However, the FMO energy gap can be decreased by using electron rich or electron poor dienophiles. To demonstrate this effect, we have computed FMO energy gaps with diformylacetylene and dimethoxyacetylene as dienophiles (Table 14). The FMO energy gaps with diformylacetylene as a dienophile were substantially lower than in the case of acetylene, suggesting that this compound, as well as other dienophiles with strong electron withdrawing groups such as esters, anhydrides, and nitriles, might be good dienophiles in reactions with derivatives of pyrroles. A similar conclusion can be made for electron rich dienophiles.

Lewis acids have been widely used to catalyze Diels-Alder reactions when thermal conditions were not efficient [43]. A limitation of the Lewis acid catalyzed Diels-Alder cycloaddition reaction has often been found to be due to the sensitivity of the substrates to the strongly acidic media. For instance, when considering the addition of phenylacetylene derivatives to 1-silyloxy-pyrrole, it was found that the Lewis acids (AlCl_3 , BF_3 , TiCl_4) led to decomposition of starting materials, while the thermal processes afforded only negligible amounts of the desired cycloadduct [44]. The successful preparation of the cycloadduct product was achieved with lithium perchlorate in ether. This approach did not produce a very acidic reaction medium, but considerably lowered the LUMO pyrrole energy, almost as much as protonation by itself (Table 14). The final effect was that the reaction became a strongly LUMO diene controlled Diels-Alder reaction.

We have already mentioned that the low reactivity of pyrrole as a diene for the Diels-Alder reaction is due to its high aromaticity. One way to determine this effect is through substituents and their effect on the stability of the pyrrole ring through uniformity of the ring bond orders by deviation from bond order uniformity (Table 15). It is obvious that pyrrole has delocalized double bond throughout the ring with a slightly increased π -bond character between C2-C3 and C4-C5. The lowest π -bond character is between N1-C2 and C5-N1. In order for pyrrole to be an ideal dienophile, then π -bond character would have to be located on the C2-C3 and C4-C5 bonds of the pyrrole ring. Because of this we can postulate that the most reactive pyrrole derivative would be the one with strong location of π -bond character for these two bonds, or the one that has the highest deviation from ring bond order uniformity. One would expect that by putting certain substituents on the pyrrole ring the aromatic character of the pyrrole ring could be changed. According to our deviations from the average pyrrole ring bond order, almost every substituent in the 2 and 5 position will increase aromaticity of the pyrrole ring (Table 15). For instance, 2,5-dimethylpyrrole has more ring bond order uniformity

(deviation was 0.584) than pyrrole itself (deviation 0.812). Interestingly, 2,5-bis(trifluoromethyl)pyrrole also had a higher bond order uniformity that was almost identical to 2,5-dimethylpyrrole (Table 15). In fact, pyrrole with the strongest electron withdrawing substituent such as sulfonyl (2,5-disulfonylpyrrole), had the most bond order uniform pyrrole ring (Table 15). Asymmetric substitution did not noticeably diminish aromaticity. Thus, 2-sulfonylpyrrole had a slightly less uniform pyrrole ring (deviation 0.482) than 2,5-disulfonylpyrrole (Table 15). Therefore, adding substituents on the pyrrole ring in the 2 and 5 positions makes it even less desirable as a diene for Diels-Alder reactions.

Table 15. Bond orders and average bond order deviation computed for derived pyrrole by AM1 semiempirical methods

compound	N1-C2	C2-C3	C3-C4	C4-C5	C5-N1	Σ
pyrrole	1.181	1.555	1.285	1.555	1.181	6.757
deviation (ABO 1.351)	0.170	0.170	0.203	0.066	0.203	0.812
2,5-dimethylpyrrole	1.166	1.522	1.296	1.522	1.166	6.672
deviation (ABO 1.334)	0.168	0.188	0.038	0.188	0.168	0.584
2,5-bis(trifluoromethyl)pyrrole	1.193	1.483	1.339	1.483	1.193	6.691
deviation (ABO 1.338)	0.145	0.145	0.001	0.145	0.145	0.581
2,5-diformylpyrrole	1.177	1.446	1.359	1.446	1.177	6.606
deviation (ABO 1.338)	0.144	0.125	0.038	0.125	0.144	0.576
2,5-disulfonylpyrrole	1.240	1.405	1.426	1.405	1.240	6.716
deviation (ABO 1.343)	0.103	0.062	0.083	0.062	0.103	0.413
2-sulfonylpyrrole	1.131	1.420	1.401	1.439	1.307	6.698
deviation (ABO 1.340)	0.209	0.080	0.061	0.099	0.033	0.482
1-methylpyrrole	1.064	1.190	1.646	1.190	1.064	6.154
deviation (ABO 1.231)	0.167	0.041	0.415	0.041	0.167	0.831
1-trifluoromethylpyrrole	1.092	1.623	1.228	1.623	1.092	6.658
deviation (ABO 1.332)	0.240	0.291	0.104	0.291	0.240	1.166
1-formylpyrrole	1.088	1.648	1.204	1.648	1.088	6.676
deviation (ABO 1.335)	0.247	0.313	0.131	0.313	0.247	1.251
1-sulfonylpyrrole	1.064	1.677	1.190	1.677	1.064	6.672
deviation (ABO 1.334)	0.270	0.343	0.144	0.343	0.270	1.370
lithium-pyrrole	1.076	1.608	1.179	1.608	1.976	6.444
deviation (ABO 1.289)	0.213	0.319	0.110	0.319	0.213	1.383
pyrrolium ion	0.923	1.853	1.047	1.853	0.923	6.599
deviation (ABO 1.320)	0.397	0.533	0.273	0.533	0.397	2.133
1-formyl-2,5-dimethylpyrrole	1.069	1.612	1.214	1.612	1.069	6.576
deviation (ABO 1.315)	0.246	0.297	0.101	0.297	0.246	1.187

ABO = average pyrrole ring bond order; Σ = some of the ring bond orders or sum of deviation from average ring bond order.

Let us now examine the effects of substituents at the N(1) position of the pyrrole ring on its ring's bond order uniformity. Opposite to carbon attached

substituents, 1-substituted pyrroles have a less uniform bond order pyrrole ring than pyrrole itself. Methyl groups have only a slight effect on the distribution of π -bond character on the pyrrole ring. One would postulate that substituents that can diminish electron density on the nitrogen ring through inductive or resonance effects should localize π -bond character between C2-C3 and C4-C5 bonds. These types of substituents are formyl and sulfonyl. In these two cases, strong deviation from uniformity was observed; therefore, it was expected that they would also show higher reactivity as dienes for Diels-Alder reactions. Maximal localization of π -bonds and hence, stronger deviation from uniformity of bond orders of the pyrrole ring can be obtained if nitrogen is quaternized as is the case with the pyrrolium cation (Table 15). For this cation, we predicted high activity because it represents an ideal diene for Diels-Alder reaction.

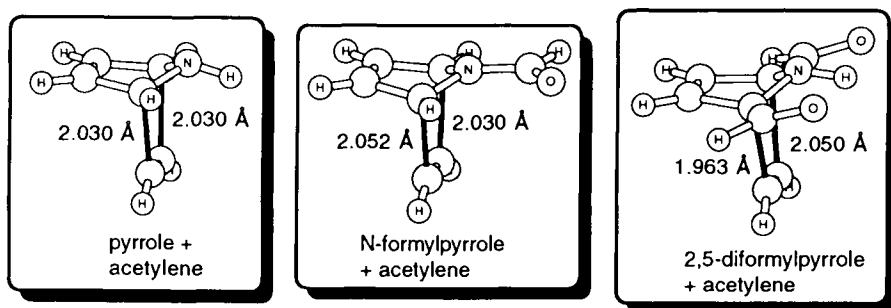


Figure 3. Some typical examples of transition state structures for acetylene addition to pyrrole and its derivatives.

All this information about the reactivity of pyrrole and its derivatives as dienes for Diels-Alder reactions was obtained from the values computed on two separated reactants. It is much more appropriate to compare changes of the FMO energies for reactants to reach transition state structures. Some representative transition state structures for acetylene addition to pyrrole derivatives are presented in Figure 3. The transition state structures are for synchronous formation of both CC bonds with slightly asymmetric transition state structures, although both diene and dienophiles have a plane of symmetry that coincides with a plane of symmetry for the transition state structure. All of these transition state structures have similar bond distances for CC bonds, which is true for almost all Diels-Alder reactions. The FMO orbital changes for transformation of reactants into transition state structures are presented in Table 16.

The most important values are the sums of the two possible frontier orbital changes; this considers dienes to change HOMO and acetylene to change LUMO energies (S_1), and the other combination a diene to change LUMO and acetylene to change HOMO energy (S_2). This energy more closely represents the energy changes for the course of the reaction than a separate examination of FMO orbital diene-dienophile reaction pair. For cyclopentadiene, pyrrole, 2,5-dimethylpyrrole,

1-methylpyrrole, 1-trifluoromethylpyrrole, and 1-formylpyrrole reacting with acetylene the reaction is HOMO diene controlled or a normal Diels-Alder cycloaddition reaction. According to calculations, 1-methylpyrrole is the best diene for this addition. If stronger, electron demanding dienophiles such as diformylacetylene are used, the necessary FMO orbital change is only 0.954 eV (Table 16). This cycloaddition should be experimentally feasible.

Table 16. Frontier molecular orbital energy (eV) changes going from reactants to transition state structures with acetylene as a dienophile computed with the AM1 semiempirical method

Diene in the TS	HOMO	LUMO	A	B	C	D	Σ_1	Σ_2
I	-8.822	0.724	0.257	0.243	2.677	-1.329	1.586	2.920
II	-8.998	1.091	-0.341	-0.287	2.501	-0.962	1.303	2.788
II ^a	-7.955	-1.326	-0.702	-2.704	3.472	-0.252	0.954	6.176
III	-8.648	1.142	-0.426	-0.123	2.851	-0.911	1.337	2.974
IV	-10.125	-0.406	0.099	0.018	1.374	-2.459	2.558	1.392
V	-9.570	-0.502	-0.045	0.290	1.929	-2.555	2.600	2.219
VI	-9.874	-1.452	1.020	-0.056	1.625	-3.505	3.525	1.681
VII	-8.909	1.164	-0.303	-0.167	2.590	-0.889	1.192	2.757
VIII	-9.459	0.340	-0.114	0.118	2.040	-1.713	1.827	2.158
IX	-9.250	0.329	-0.046	0.368	2.249	-1.724	1.770	2.617
X ^b	-8.574	0.118	0.630	0.157	1.143	1.331	1.961	1.300
XI	-9.605	-1.054	-0.045	0.280	1.894	-3.107	3.152	2.174
XII	-14.186	-5.105	0.960	0.415	-2.687	-7.158	8.118	3.102

I = cyclopentadiene; II = pyrrole; III = 2,5-dimethylpyrrole; IV = 2,5-di(trifluoromethyl)pyrrole; V = 2,5-diformylpyrrole; VI = 2,5-disulfonylpyrrole; VII = 1-methylpyrrole; VIII = 1-trifluoromethylpyrrole; IX = 1-formylpyrrole; X = 1-sulfonylpyrrole; XI = 1-formyl-2,5-dimethylpyrrole; XII = pyrrolium ion; ^afor diformylacetylene addition to pyrrole; ^bfor 1,2-dimethoxyacetylene addition to 1-formylpyrrole; A = HOMO_{TS} - HOMO_{diene}; B = LUMO_{TS} - LUMO_{diene}; C = HOMO_{TS} - HOMO_{acetylene}; D = LUMO_{TS} - LUMO_{acetylene}; $\Sigma_1 = |A| + |D|$; $\Sigma_2 = |B| + |C|$.

Interestingly, systems that we have predicted to be slightly more aromatic are also less reactive than pyrrole itself on the basis of the pyrrole ring bond order uniformity. For 1-substituted pyrroles, it was predicted that, when a larger deviation of the ring bond order uniformity existed, lower reactivity than that of pyrrole was predicted (Table 16). This finding is not at all surprising since the FMO gaps between pyrrole and acetylene as reactants are 10.710 eV for the HOMO and 12.877 eV for LUMO diene controlled reaction. The effect of the substituent on the pyrrole nitrogens should be noticeable. The pyrrole LUMO energy should be lowered substantially to allow an efficient molecular orbital overlap between reactants. Of course, 1-substituted pyrroles should be good dienes in a reaction with a more reactive acetylene such as dimethoxyacetylene (Table 14); this would make energy changes going from LUMO 1-formylpyrrole and

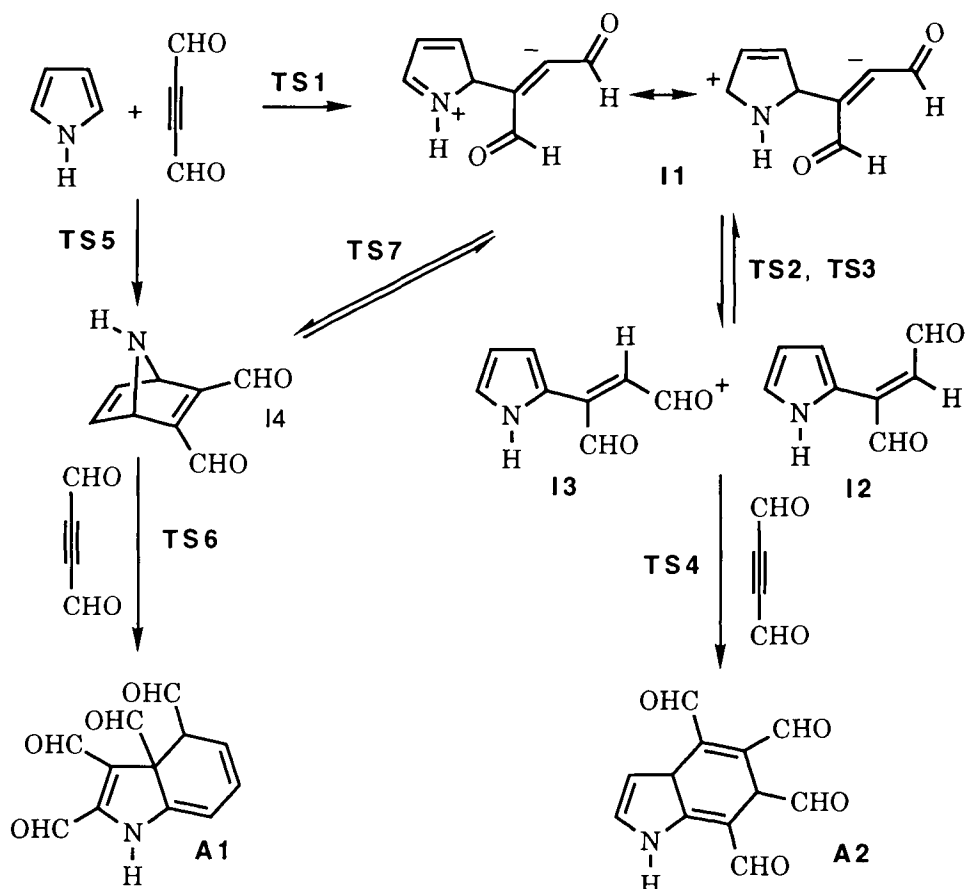
HOMO 1,2-dimethoxyacetylene comparable to the FMO changes for acetylene addition to pyrrole. As stated previously, the addition of acetylene to the pyrrolium ion should be examined separately because of its charge; the full localization of double bonds on the pyrrole ring (an ideal diene for a Diels-Alder reaction), and the charge in the pyrrolium cation result in an exceptionally low LUMO energy.

Table 17. Activation barriers (kcal/mol) for acetylene addition to pyrrole derivatives computed by AM1 and B3LYP/6-31G(d)//AM1 computational approaches

transition state with	A	B	C	D	ΔE_1	ΔE_2
acetylene	54.8		-77.3254			
pyrrole	39.9	136.2	-210.1633	-287.4353	41.5	33.5
2,5-dimethylpyrrole	23.4	122.6	-288.8003	-366.0726	44.4	33.3
2,5-bis(trifluoromethyl)pyrrole	272.4	173.1	-884.2307	-961.5043	44.5	32.5
2,5-diformylpyrrole	-21.5	76.1	-436.8164	-514.0778	42.8	40.2
2,5-disulfonylpyrrole	-65.7	41.8	-1307.2079	-1384.4801	52.7	33.4
1-methylpyrrole	44.4	141.1	-249.4738	-326.7461	41.9	33.3
1-trifluoromethylpyrrole	-96.9	-0.8	-547.1893	-624.4600	41.3	34.3
1-formylpyrrole	13.7	110.3	-323.4958	-400.7681	41.8	33.3
1-sulfonylpyrrole	-4.3	91.5	-758.6882	-835.9611	41.0	33.0
lithium-pyrrole	175.3	261.9	-294.7924	-217.5119	31.8	28.2
pyrrolium ion	212.6	298.0	-210.4804	-287.7877	30.6	11.3

A = AM1 computed heat of formation for pyrrole derivatives (kcal/mol); B = AM1 computed heat of formation of transition state structure with acetylene; C = Total energy computed by B3LYP/6-31G(d)//AM1 (Hartree); D = Total energy for transition state structure with acetylene evaluated with B3LYP/6-31G(d) on AM1 geometries (Hartree); ΔE_1 = activation barrier computed by AM1; ΔE_2 = activation barrier computed with B3LYP/6-31G(d) on AM1.

On the basis of the FMO and the bond order, it is fair to say that the addition of acetylene to pyrrole is not experimentally feasible. This reaction can be enforced by properly substituted pyrrole and acetylene as well as by using strong acids, but which might diminish the aromatic character. To confirm our studies, we have evaluated activation barriers for those reactions by AM1 and B3LYP/6-31G(d) theory models (Table 17). While semiempirical AM1 methods predicted a slightly lower activation barriers for 1-substituted pyrroles, the B3LYP/6-31G(d) computed activation energies were all around 33 kcal/mol, regardless of the nature of the substituent attached to the pyrrole ring. A noticeable decrease in the activation energy was observed for cases in which the lithium ion and proton catalyzed acetylene addition to pyrrole (Table 17).



Scheme 2. Possible reaction pathways in addition of diformylacetylene to pyrrole.

The addition of the more reactive acetylene based dienophile, diformylacetylene to pyrrole, may now be focused upon. Possible reaction pathways for this reaction are presented in Scheme 2. Because of the presence of two of the reactants which can stabilize both positive (pyrrole) and negative (diformylacetylene) charges well, it is reasonable to propose formation of an ionic intermediate **I1**. Both positive and negative charges in this structure are stabilized through delocalization (resonance). Intermediate **I1** can be rearranged through a [1,4] hydrogen shift to either intermediate **I3** or **I4**. The reaction should be suprafacial and thermally feasible because the transition state **TS2** should have aromatic character (two electrons and positive charge). The other possibility is a [1,5] hydrogen shift in the pyrrole ring that should have a comparable activation barrier. Intermediates **I3** and **I4** should be better dienes for the Diels-Alder reaction and should further react with diformylacetylene through **TS4**. Intermediates can be reoriented in a such way that cycloadduct **I4** can be formed. On the other hand, this cycloadduct can be directly formed from pyrrole and

diformylacetylene through **TS5**. The intermediate **I4** can undergo a further cycloaddition reaction forming 1:2 cycloadduct product but these three different pathways should produce two different 1:2 cycloadduct products (Scheme 2).

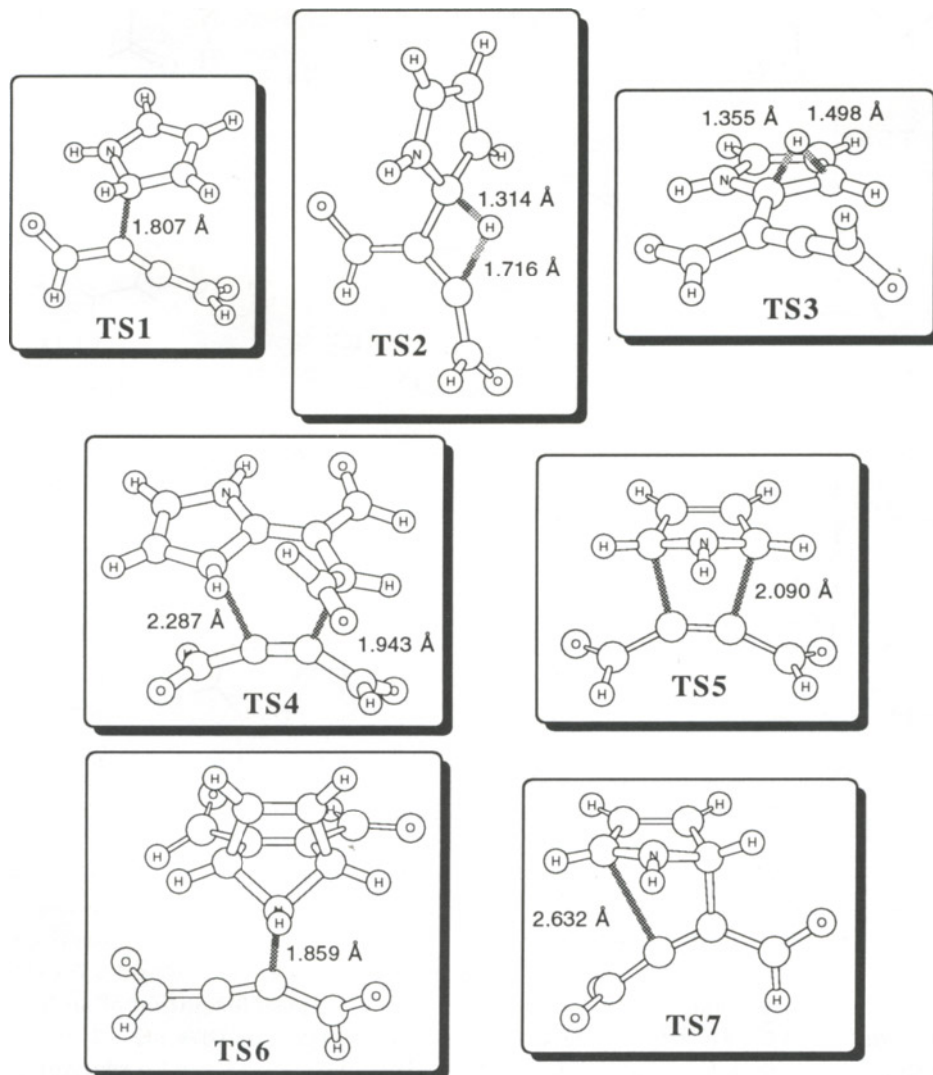


Figure 4. Possible transition state structures for diformylacetylene addition to pyrrole.

The AM1 optimized structures of all transition state structures from Scheme 2 are presented in Figure 4. The transition state structures presented in Scheme 1 are ones that could be expected on the basis of a general knowledge of chemical transformations. Estimated activation barriers were substantially lower (Table 18) than those generated for acetylene addition to pyrrole derivatives,

Table 18. Activation barriers (kcal/mol) for diformylacetylene addition to pyrrole computed by AM1 and B3LYP/6-31G(d)//AM1 computational approaches.

Chemical structure	A	B	ΔE_1	ΔE_2
diformylacetylene	-5.8	-303.960546		
pyrrole	39.9	-210.163390		
I1	60.7	-514.095807		
I2	-6.3	-514.209526		
I3	-6.6	-514.210495		
I4	35.9	-514.142035		
TS1	65.9	-514.097948	31.8	16.3
TS2	82.1	-514.069140	21.4	16.7
TS3	85.1	-514.060726	24.4	22.0
TS4	22.7	-818.142181	35.1	18.1
TS5	72.0	-514.085708	37.9	24.0
TS6	54.1	-818.093909	24.0	5.5
TS7	63.6	-514.100018	2.9	2.6

A = AM1 computed heat of formation (kcal/mol); B = total energy for transition state structure with acetylene evaluated with B3LYP/6-31G(d) on AM1 geometries (Hartree); ΔE_1 = activation barrier computed by AM1; ΔE_2 = activation barrier computed with B3LYP/6-31G(d) on AM1.

(Table 17). For instance, the AM1 computed activation barrier for acetylene addition to pyrrole should be 41.5 kcal/mol (Table 17). B3LYP energy evaluation on AM1 geometries predicted an activation barrier at 33.5 kcal/mol. This is too high to be reached under thermal conditions without causing formation of a substantial amount of byproducts. On the other hand, direct cycloaddition of diformylacetylene to pyrrole through similar transition state structures (**TS5**) had an activation barrier 24.0 kcal/mol (Table 18). According to our calculation, this addition should not occur through **TS5** but thorough **TS1** which represents a Michael type of addition [16]. The B3LYP/6-31G(d) estimated activation energy for this reaction was only 16.3 kcal/mol suggesting that the zwitterionic intermediate **I1** is more preferable to form than direct cycloaddition to intermediate **I4** (Scheme 2). Of course, this ionic structure is very high on the reaction potential energy surface and can be stabilized through three different channels; through [1,5]-hydrogen shift (**TS3**, Figure 2) producing another zwitterion structure, through [1,4]-hydrogen shift (**TS2**) producing a neutral 2-vinylpyrrole derivative (**I2** and **I3**) or reconcile charges in a such way (**TS7**) that the Diels-Alder cycloadduct (**I4**) would be formed. The transition state for scrambling hydrogen in pyrrole (**TS3**) was predicted to have the highest activation energy (22 kcal/mol) of all three possible pathways. The lowest possible activation barrier (2.6 kcal/mol) was achieved through reconciliation of charges in **I1** and by producing Diels-Alder cycloadduct **I4**. Therefore, formation of cycloadduct **I4** is energetically preferred over the formation of vinylpyrrole derivatives **I2** and **I3**. If intermediate **I2** were formed, a second molecule of diformylacetylene could be

added with a moderate activation barrier of (18.1 kcal/mol, Table 18). On the other hand, a second molecule of diformylacetylene is predicted to add to Diels-Alder adduct **I4** with formation of a new zwitterionic intermediate that will again reconcile charges through formation of a 1:2 Diels-Alder adduct (**A1**). The activation energy for the formation of the first zwitterionic structure is predicted to be only 5.5 kcal/mol. In general, the reaction is expected to first form zwitterionic intermediate **I1** that will recombine into Diels-Alder adduct **I4**. This will later react with one molecule of the dienophile, producing a new zwitterionic structure that will recombine into a 1:2 Diels-Alder adduct **A1**.

This type of cycloaddition reaction has been well explored experimentally. Diels, Alder, Winckler and Peterson [45] explored the product of the reaction between 1-methylpyrrole and dimethyl acetylenedicarboxylate. They suggested product type **A2** as a 1:2 cycloadduct. If this is true, the reaction should go through formation of a vinylpyrrole intermediate similar to **I2** and **I3**. It was later shown [46] that the actual product of this addition is product type **A1**, as predicted through our computational studies. Actually, the same reaction product was obtained with pyrrole [47] as a diene as outlined in our study. Furthermore, when an electron-withdrawing group was placed on the nitrogen atom of pyrrole, the pyrrole aromatic ring was found [48] to be more reactive as a diene toward acetylenic dienophiles as predicted by our bond order uniformity study. In many cases, products of a Michael addition were isolated instead of the desired cycloadduct [49]. To prevent Michael addition at the C(2) position of the pyrrole ring, 1-acyl-2,5-disubstituted pyrroles were used [50]. The results of our computational study of diformylacetylene addition to 1-formyl-2,5-pyrrole fully agree with experimental studies. The Diels-Alder transition state has an energy of activation that is 2.3 kcal/mol lower than the energy of the Michael type of transition state structure (Figure 5).

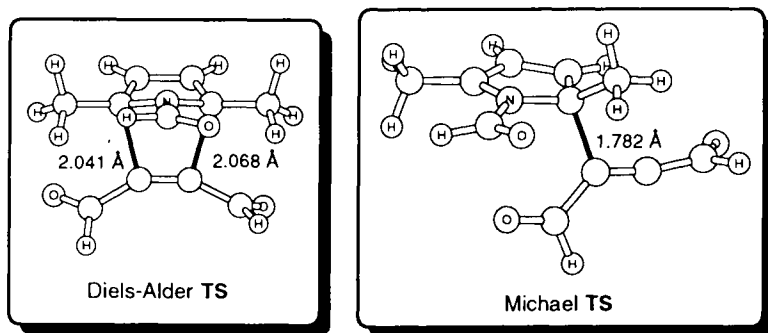


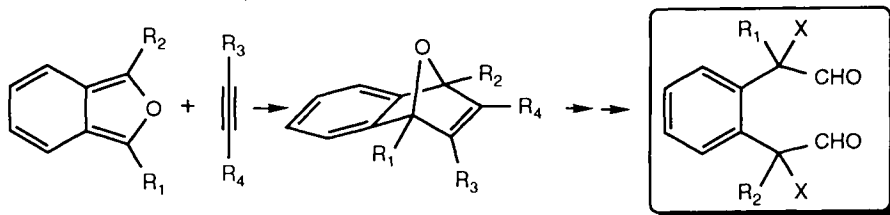
Figure 5. Two possible transition state for reaction between diformylacetylene and 1-formyl-2,5-dimethylpyrrole.

We can summarize that, generally, derivatives of pyrrole do not react with low or moderately reactive dienophiles such as acetylene. There are at least two reasons: FMO energies are too low (HOMO) or too high (LUMO) and pyrrole has

high aromatic character (very uniform ring bond orders). Substituents at the C(1) position can substantially decrease the aromatic character of pyrrole and consequently increase its reactivity. Nevertheless, reactive dienophiles such as diformylacetylene are necessary for the reaction. If pyrrole does not have substituents at the C(2) and C(5) position, formation of Michael type adducts might be preferred. To get a qualitative picture of the reaction potential, a single point B3LYP/6-31G(d) computational study is required. Our computational study is in complete agreement with experimental data.

3.4. Diels-Alder reactions with benzo[*b*] and benzo[*c*]-fused heterocycles

Five-membered heterocyclic compounds with a fused benzene ring are, from a theoretical point of view, ideal starting materials for the preparation of complex organic compounds that contain a 1,2-disubstituted benzene ring. One possible transformation begins with benzene[*c*]-fused heterocycles through a Diels-Alder reaction with acetylene derivatives. The formed cycloadduct can be ozonized; the furan ring can be easily opened generating the desirably functionalized 1,2-disubstituted benzene (Scheme 3). Some theoretical studies pertaining to the



Scheme 3. Possible transformation of a benzo-fused heterocycle into 1,2-functionalized benzene derivatives.

reactivity of benzo-fused heterocycles in cycloaddition reactions was published previously. While benzo[*b*] heterocyclic compounds are stable compounds, benzo[*c*] heterocycles are very reactive species, which are usually difficult to isolate in pure form [15]. For instance, existence of benzo[*c*]furan was unequivocally proven by Fieser and Haddadin [51]. There are many methods for preparing benzo[*c*]furan but it is quite difficult to obtain in pure form [52]. Some of the methods which produce relatively pure compounds were published in 1972. These include the pyrolyses of 1,4-epoxy-1,2,3,4-tetrahydronaphthalene at 650°C and 0.1 torr [53]. We have studied the reactivity of benzo-fused heterocycles through B3LYP/6-31G(d) evaluation of relative energies, as well as magnetic properties, although only our results in the evaluation of the benzene-fused furan heterocycle reactivity through uniformity of the heterocycle, benzene ring bond order uniformity, and Frontier Molecular Orbital (FMO) energies are presented here.

The aromatic character of benzo[*b*]-fused heterocycles was evaluated through uniformity of their ring bond order (Table 19). It is very interesting to mention that little aromatic stability for benzene is lost by [*b*] fusing with furan, pyrrole,

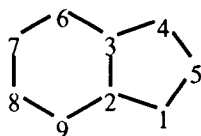


Table 19. Bond order uniformity in the heterocyclic and benzene ring of the benzo[*b*]heterocycle as computed by the AM1 semiempirical method

Bonds	Benzo[<i>b</i>]furan		Benzo[<i>b</i>]pyrrole		Benzo[<i>b</i>]thiophene	
	BO	BOD	BO	BOD	BO	BOD
	Heterocyclic ring					
X(1)-C(2)	1.060	-0.194	1.123	-0.143	1.097	-0.180
C(2)-C(3)	1.297	0.043	1.274	0.006	1.316	0.039
C(3)-C(4)	1.102	-0.152	1.161	-0.107	1.140	-0.137
C(4)-C(5)	1.742	0.488	1.643	0.376	1.704	0.427
C(5)-X(1)	1.067	-0.186	1.136	-0.132	1.129	-0.148
Σ	6.267	1.063	6.337	0.764	6.387	0.932
ABO	1.253	0.213	1.267	0.153	1.277	0.186
	Benzene ring					
C(2)-C(3)	1.297	-0.080	1.274	-0.091	1.316	-0.059
C(3)-C(6)	1.342	-0.035	1.301	-0.063	1.296	-0.079
C(6)-C(7)	1.460	0.083	1.501	0.136	1.507	0.132
C(7)-C(8)	1.371	-0.006	1.327	-0.038	1.323	-0.052
C(8)-C(9)	1.452	0.075	1.502	0.137	1.505	0.130
C(9)-C(2)	1.340	-0.037	1.283	-0.082	1.305	-0.070
Σ	8.261	0.316	8.188	0.547	8.253	0.523
ABO	1.377	0.053	1.365	0.091	1.376	0.087
Total BOD		1.379		1.311		1.455

BO = bond order; BOD = bond order deviation from average bond order; Σ = sum of bond orders or sum bond order deviation from average bond order; ABO = average ring bond order.

or thiophene. Actually, the difference in bond order deviation in the benzene ring from the average bond order is very small. The closest resemblance with separated benzene was obtained for benzo[*b*]furan, which is the least aromatic of the three heterocycles (Table 1). The highest uniformity for the heterocycles was observed for pyrrole. Overall, benzo[*b*]pyrrole should be the most aromatic, while benzo[*b*]furan should be the least aromatic of these three heterocycles.

For the benzo[*c*]-fused heterocycles presented in Table 20, the same interesting features are possessed. All the heterocycles have a much more uniform ring bond order than their benzo[*b*] isomers. Benzene bond order uniformity here is considerably more disrupted by the presence of the five-membered heterocycle (Table 20). According to the ring's bond orders, the most uniform heterocycle is benzo[*c*]pyrrole with a total bond order deviation of 1.718.

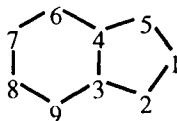


Table 20. Bond order uniformity in the heterocyclic and benzene ring of benzo[*c*]heterocycles as computed by the AM1 semiempirical method

Bonds	Benzo[<i>c</i>]furan		Benzo[<i>c</i>]pyrrole		Benzo[<i>c</i>]thiophene	
	BO	BOD	BO	BOD	BO	BOD
	Heterocyclic ring					
X(1)-C(2)	1.133	-0.158	1.225	-0.065	1.240	-0.069
C(2)-C(3)	1.525	0.234	1.401	0.111	1.434	0.125
C(3)-C(4)	1.140	-0.152	1.199	-0.092	1.198	-0.111
C(4)-C(5)	1.525	0.234	1.401	0.111	1.434	0.125
C(5)-X(1)	1.133	-0.158	1.225	-0.065	1.240	-0.069
Σ	6.455	0.935	6.451	0.445	6.544	0.500
ABO	1.291	0.187	1.290	0.089	1.309	0.100
	Benzene ring					
C(3)-C(4)	1.140	-0.184	1.199	-0.138	1.198	-0.138
C(4)-C(6)	1.117	-0.206	1.161	-0.176	1.149	-0.186
C(6)-C(7)	1.710	0.387	1.655	0.318	1.669	0.334
C(7)-C(8)	1.146	-0.177	1.190	-0.147	1.178	-0.157
C(8)-C(9)	1.710	0.387	1.655	0.318	1.669	0.334
C(9)-C(3)	1.117	-0.206	1.161	-0.176	1.149	-0.186
Σ	7.939	1.547	8.021	1.273	8.013	1.335
ABO	1.323	0.258	1.337	0.212	1.335	0.223
Total BOD		2.482		1.718		1.835

BO = bond order; BOD = bond order deviation from average bond order; Σ = sum of bond orders or sum of bond order deviation from average bond order; ABO = average ring bond order.

This difference of bond order deviations for benzo[*c*]pyrrole and benzo[*c*]thiophene suggest that they have comparable aromaticity. From the presented results, it is obvious that benzo[*b*] heterocycles have a more even distribution of bond orders (delocalization); hence, they are more aromatic in comparison with their benzo[*c*] isomers. Furthermore, the difference in the energy between these two groups of isomers should be very high for benzofurans, while benzopyrrole and benzothiophene should have a similar energy difference between their [*b*] and [*c*] isomers. Considering the fact that benzene is more aromatic than any of the heterocyclic aromatic compounds, the system that has a higher benzene bond order deviation is less stable and therefore more reactive. In our case, our results showed bond order as benzo[*c*]furan first, followed by benzo[*c*]thiophene and finally benzo[*c*]pyrrole.

To confirm this finding, we computed energy differences between these two series of isomers. For the sake of discussion of relative stability of these

compounds, through isodesmic reactions one can imagine fusing benzene with a five-membered heterocycle with the elimination of ethylene. The energy of this imaginary reaction might also suggest the relative stability of benzo-fused heterocycles. All calculations suggested that benzo[*b*]-fused heterocycles are more stable than benzo[*c*]heterocycles, which is in perfect agreement with the computed bond order deviation from ring uniformity (Table 19 and 20). The enthalpy for the imaginary reaction between heterocycle and benzene, with producing benzo-fused heterocycle and ethylene, is also a good measure of their stability. This reaction, as for all cases, is endothermic; the more endothermic the system is, the less stable it is. As mentioned previously, the fused benzene and heterocyclic system is less thermally favorable than two separated aromatic systems. The least unfavorable fusion is for benzo[*b*]furan, while the most unfavorable fusion is for benzo[*c*]furan. Almost identical energies were obtained for the imaginary ring fusion between benzene and pyrrole on one side and benzene and thiophene on the opposite side (Table 21). This indicated that they may possess similar aromatic properties. Therefore, the most favorable diene for the Diels-Alder reaction should be benzo[*c*]furan.

Table 21. Energy difference (kcal/mol) between two isomers and energy for the imaginary reaction of benzene fusing with a five-membered heterocycle

Method	Benzofuran			Benzopyrrole			Benzothiophene		
	ΔE_I	ΔE_{II}	ΔE_{III}	ΔE_I	ΔE_{II}	ΔE_{III}	ΔE_I	ΔE_{II}	ΔE_{III}
A	7.1	13.9	20.9	6.6	11.3	17.8	7.1	10.9	18.0
B	14.3	5.3	19.6	9.2	6.6	15.8	12.1	6.6	18.8

A = AM1; B = B3LYP/6-31G(d)//AM1; ΔE_I = energy difference between benzo[*b*]heterocycle and benzo[*c*]heterocycle; ΔE_{II} = imaginary enthalpy of reaction benzene + heterocycle into benzo[*b*]heterocycle plus ethylene; ΔE_{III} = imaginary enthalpy of reaction benzene + heterocycle into benzo[*c*]heterocycle plus ethylene.

Let us now turn our focus to Frontier Molecular Orbital (FMO) energies as a measure for the reactivity of benzo-fused heterocycles in reaction with acetylene, ethylene, and cyclopropene. According to this approach, the most reactive reactant pair would be one that has the smallest FMO gap between the two reactants. The FMO and their energy gap for Diels-Alder reaction with benzo-fused heterocycles is presented in Table 22. According to these calculations, benzo[*c*]heterocycles are much better dienes for Diels Alder reactions than their benzo[*b*]isomers. In all cases of acetylene, ethylene, and cyclopropene, addition to the benzo[*c*]heterocycle, the reaction is HOMO diene controlled. Furthermore, all calculations agreed that benzo[*c*]pyrrole is the most reactive diene for the Diels-Alder reaction due to its very high HOMO energy. This finding is in total opposition to our bond uniformity approach. Imaginary benzene and heterocycle fusion reaction energy selects benzo[*c*]pyrrole as at least the same, if not higher stability as benzo[*c*]thiophene. It is, however, significantly more stable than benzo[*c*]furan.

Thus, benzo[c]furan, not pyrrole, must be the most reactive diene for the Diels-Alder reaction.

Table 22. Frontier molecular orbital energy gaps (eV) for acetylene, ethylene, and cyclopropene addition to benzoheterocycles computed by the AM1 semiempirical method

diene	HOMO	LUMO	A	B	C	D	E	F
I	-9.010	-0.063	11.063	11.436	10.448	10.488	10.052	9.755
II	-8.263	-0.396	10.316	11.104	9.701	10.156	9.305	9.423
III	-8.403	0.300	10.456	11.799	9.841	10.851	9.446	10.119
IV	-7.796	0.142	9.849	11.641	9.234	10.693	8.838	9.960
V	-8.430	-0.166	10.483	11.334	9.868	10.386	9.472	9.653
VI	-8.340	-0.592	10.393	10.908	9.778	9.960	9.382	9.227

I = benzo[b]furan; II = benzo[c]furan; III = benzo[b]pyrrole; IV = benzo[c]pyrrole; V = with benzo[b]thiophene; VI = with benzo[c]thiophene; A = $LUMO_{\text{acetylene}} - HOMO_{\text{benzoheterocycle}}$; B = $LUMO_{\text{benzoheterocycle}} - HOMO_{\text{acetylene}}$; C = $LUMO_{\text{ethylene}} - HOMO_{\text{benzoheterocycle}}$; D = $LUMO_{\text{benzoheterocycle}} - HOMO_{\text{ethylene}}$; E = $LUMO_{\text{cyclopropene}} - HOMO_{\text{benzoheterocycle}}$; F = $LUMO_{\text{benzoheterocycle}} - HOMO_{\text{cyclopropene}}$

As we have demonstrated above, a more reliable way to adjust the reactivity of various dienes for Diels-Alder reactions is by estimating FMO energy changes necessary to transfer reactants into transition state structures. In many cases, more than one isomer of the cycloaddition reaction can be formed for which FMO energy gaps between reactants cannot be used to assess selectivity of the reaction. Our approach should allow us to determine energy preferences of one isomer over the other. The transition state structures as well as their energies for this study must be available. The AM1 geometries of the transition state structures will be presented later. There are two changes of FMO energies in the transformation of reactants into transition state structures: one with HOMO diene controlled (S_1) and the other a LUMO diene controlled (S_2) Diels Alder reaction (Table 23). Because cyclopropene is the most reactive dienophile studied here, we will only present the FMO study for the addition of cyclopropene to benzo-fused heterocycles. There are twelve possible transition state structures; for each a FMO change was computed. It is obvious that the reaction is HOMO diene controlled with the exception of *endo* cyclopropene addition to benzo[b]furan and *exo* cyclopropene addition to benzo[c]thiophene (Table 23). Similarly, the conclusion obtained from FMO energy gaps between reactants (Table 22) was that in all cases, benzo[c]heterocycles were much better dienes for Diels-Alder reactions than benzo[b]heterocycles. This is demonstrated by a lower FMO energy change if cyclopropene is coupled with the benzo[c]heterocycle as opposed to benzo[b]heterocycle as a diene (Table 23). A review of computed FMO energy changes(S_1) shows the order of reactivity of benzo-fused heterocycles as benzo[c]furan, benzo[c]pyrrole, benzo[c]thiophene, benzo[b]pyrrole, benzo[b]furan followed by benzo[b]thiophene. For the addition of cyclopropene to benzo[c]furan,

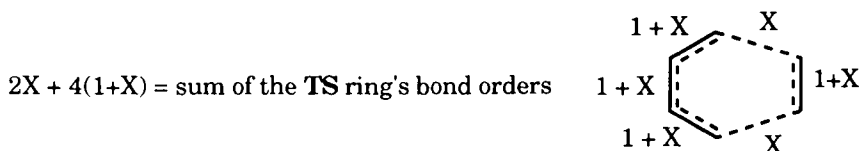
the *exo* isomeric transition state structure is predicted to have a lower activation energy while for the reaction with benzo[*c*]pyrrole, an opposite reaction outcome is expected.

Table 23. Frontier molecular orbital energy (eV) changes going from reactants to the transition state structures computed with the AM1 semiempirical method

Addition	HOMO	LUMO	A	B	C	D	Σ_1	Σ_2
<i>exo</i> I	-7.841	-0.725	1.169	-0.662	1.978	-1.767	2.936	2.640
<i>endo</i> I	-7.013	-1.582	1.997	-1.519	2.806	-2.624	4.621	4.325
<i>exo</i> II	-8.305	-0.228	-0.042	0.168	1.514	-1.270	1.312	1.682
<i>endo</i> II	-8.311	-0.210	-0.048	0.186	1.508	-1.252	1.300	1.694
<i>exo</i> III	-7.564	-0.692	0.839	-0.992	2.255	-1.734	2.573	3.247
<i>endo</i> III	-6.920	-1.333	1.483	-1.633	2.899	-2.375	3.858	4.532
<i>exo</i> IV	-8.150	0.066	-0.354	-0.076	1.669	-0.976	1.330	1.745
<i>endo</i> IV	-8.142	0.061	-0.346	-0.081	1.677	-0.981	1.327	1.758
<i>exo</i> V	-6.929	-1.611	1.501	-1.445	2.890	-2.653	4.154	4.335
<i>endo</i> V	-7.027	-1.585	1.403	-1.419	2.792	-2.627	4.030	4.211
<i>exo</i> VI	-8.490	-0.390	-0.150	0.202	1.329	-1.432	1.582	1.531
<i>endo</i> VI	-8.403	-0.424	-0.063	0.168	1.416	-1.466	1.529	1.584

I = benzo[*b*]furan; II = benzo[*c*]furan; III = benzo[*b*]pyrrole; IV = benzo[*c*]pyrrole; V = with benzo[*b*]thiophene; VI = benzo[*c*]thiophene; A = HOMO_{TS} - HOMO_{diene}; B = LUMO_{TS} - LUMO_{diene}; C = HOMO_{TS} - HOMO_{cyclopropene}; D = LUMO_{TS} - LUMO_{cyclopropene}; $\Sigma_1 = |A| + |D|$; $\Sigma_2 = |B| + |C|$.

Now we would like to use a transition state ring bond order uniformity (π -molecular orbital delocalization) as a measure of its stability, and therefore the selectivity between two or more isometric transition state structures. A view that transition state structures can be classified as aromatic and antiaromatic is widely accepted in organic chemistry [54]. A stabilized aromatic transition state will lead to a lower activation barrier. Also, it can be said that a more uniform bond order transition state will have lower activation barriers and will be allowed. An ideal uniform bond order transition state structure for a six-membered transition state structure is presented in Scheme 4. According to this definition, a six-electron transition state can be defined through a bond order distribution with an average bond order X . Less deviation from these ideally distributed bond orders is present in a transition state which is more stable. Therefore, it is energetically preferred over the other transition state structures.



Scheme 4. Distribution of bond orders in the six-member ideal aromatic transition state structure.

This bond order deviation from an ideal transition state structure to an example of cyclopropene added to a benzo-fused heterocycle may now be applied. Before we examine bond order deviation from an ideal transition state, we can take a look at the sums of rings bond order in the transition state structures. To simplify this picture, we will focus only on *exo* transition state structures between cyclopropene and benzo[*c*]heterocycles. Previously, we mentioned that the Diels-Alder reaction with benzo[*c*]heterocycles as dienes is a HOMO controlled diene reaction, therefore, an electron rich (higher sum of bond order) transition state structure should be energetically preferred. If this is the case, the order of reactivity should be benzo[*c*]furan, benzo[*c*]thiophene, and then benzo[*c*]pyrrole, which is exactly the same as determined on the basis FMO energy change (Table 23).

The transition state with the highest Bond Order Deviation (BOD) also has a higher energy. From the example of cyclopropene addition to benzo-fused furans, we can perfectly demonstrate this approach. The BOD for both *exo* and *endo* transition state structures of cyclopropene addition to benzo[*b*]furan were higher than for addition to benzo[*c*]furan (Table 24). On the other hand, the transition

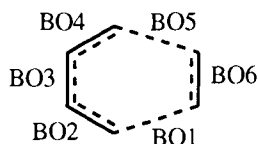


Table 24. Bond orders for the transition state ring and SOI in transition state structures computed by AM1 semiempirical methods with cyclopropene as dienophile

Addition	BO1	BO2	BO3	BO4	BO5	BO6	BO7	BO8	BOD
<i>exo</i> I	0.256	1.194	1.394	1.273	0.467	1.439	0.004		0.577
<i>endo</i> I	0.151	1.219	1.316	1.129	0.744	1.219	0.000	0.045	0.934
<i>exo</i> II	0.323	1.233	1.262	1.233	0.323	1.542	0.004		0.460
<i>endo</i> II	0.342	1.232	1.262	1.232	0.342	1.519	0.002	0.002	0.477
<i>exo</i> III	0.256	1.175	1.438	1.222	0.512	1.396	0.003		0.693
<i>endo</i> III	0.190	1.199	1.384	1.120	0.739	1.217	0.000	0.035	1.013
<i>exo</i> IV	0.376	1.159	1.282	1.159	0.376	1.470	0.003		0.622
<i>endo</i> IV	0.388	1.169	1.280	1.169	0.388	1.457	0.002	0.002	0.615
<i>exo</i> V	0.087	1.202	1.422	1.124	0.751	1.223	0.023		1.141
<i>endo</i> V	0.180	1.227	1.357	1.140	0.764	1.209	0.000	0.039	0.991
<i>exo</i> VI	0.385	1.181	1.288	1.181	0.385	1.464	0.018		0.584
<i>endo</i> VI	0.385	1.210	1.281	1.210	0.385	1.465	0.003	0.003	0.534

I = benzo[*b*]furan; II = benzo[*c*]furan; III = benzo[*b*]pyrrole; IV = benzo[*c*]pyrrole; V = with benzo[*b*]thiophene; VI = benzo[*c*]thiophene; X is bond order ideal delocalization computed from formula presented in Scheme 2; BOD = sum of bond order deviation from uniform bond order distribution in the transition state structures.

state for *exo* cyclopropane addition to benzo[*c*]furan has a slightly more uniform bond order transition state structure than for the *endo* transition state structures. Clearly, it is the *exo* cycloadduct product between cyclopropane and benzo[*c*]furan that has the most stable transition state structure. The *exo* selectivity in cyclopropene addition to benzo[*c*]furan can be explained by the stabilization effect of Secondary Orbital Interactions (SOI) in the transition state structure between the lone pair of furan oxygen and methylene hydrogen of cyclopropene. There are also SOI in the *endo* transition state structure between the methylene hydrogen of cyclopropene and π -orbitals on the double bond in formation. The n-H SOI are stronger in comparison with π -H SOI and formation of an *exo* isomer and is preferred, as in the case of cyclopropene addition to isolated furan. If we consider only bond order uniformity in the six-membered transition state structures, the *endo* addition of cyclopropene to both benzo[*c*]pyrrole and benzo[*c*]thiophene should be preferred over *exo* addition. This might not be necessarily true if we take into account that, for instance, the n-H SOI in the *exo* transition state structure with benzothiophene is extremely strong. We can then assume that the formation of an *exo* cycloadduct, as in case of cyclopropene addition to both benzo[*c*]pyrrole and benzo[*c*]thiophene, might be also slightly preferred over *endo* cycloaddition. Later, we will support these observations with computations of the activation energies for all possible reaction pathways.

Our qualitative computational studies verified that activation barriers of both AM1 semiempirical and B3LYP density are functional on the theoretical level. The transition state structures for different combinations of benzo-fused heterocycles with various dienophiles are very similar and therefore only transition structures between benzo-fused furans and cyclopropene will be presented (Figure 6). The transition state structure for cyclopropene addition to benzo[*b*]furan is for concerted mechanisms and is asynchronous. Higher asynchronicity is observed in *endo* isomers. Lower asynchronicity seen in *exo* isomers may be attributed to a strong n-H SOI between the two reactants moieties of the transition state structure. On the other hand, the two transition state structures for cyclopropene addition to benzo[*c*]furan have a plane of symmetry indicating synchronous formation of both CC bonds. If we apply the Hammond postulate [55], we also have to determine which of two isomeric transition state structures will have a lower energy. According to the Hammond postulate the transition state structure that is closer in geometry to the reactants should be also closer in the energy. Therefore, this transition state should have the lower energy. If we consider only C-C bond distances in *exo* and *endo* transition state structures for cyclopropene addition to benzo[*c*]furan, then the *exo* transition with longer C-C bond, distance is closer to reactants, hence the more stable. This conclusion is in full compliance with minimal bond order deviation from an ideal transition state structure (Table 24). Studies select this structure as the one with the lowest activation barrier.

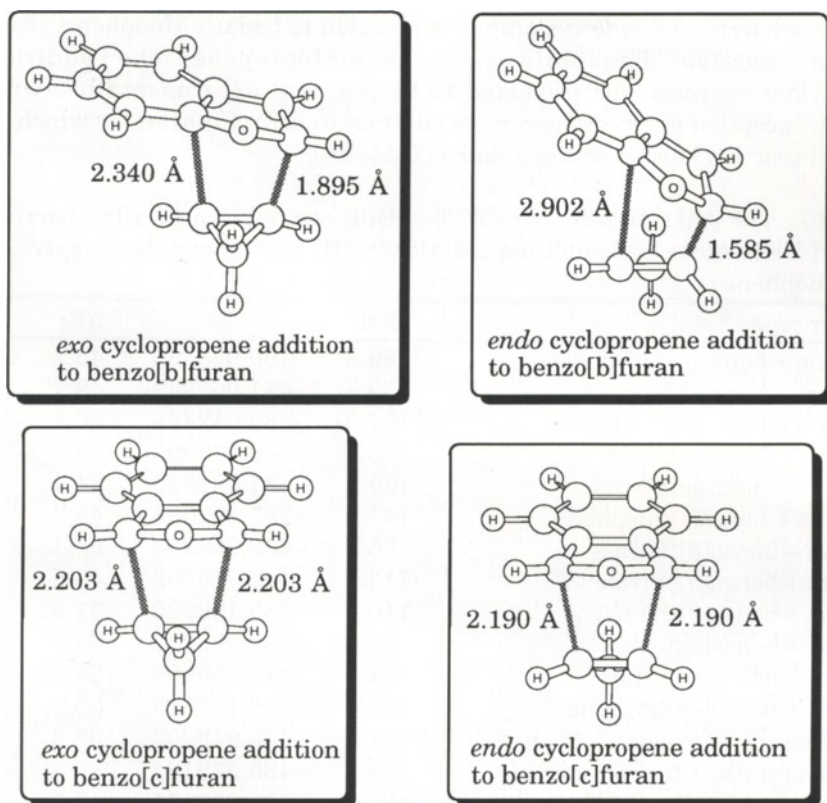


Figure 6. Four possible transition state structures for cyclopropene addition to benzo[b] and benzo[c]furan computed by the AM1 semiempirical method.

The AM1 and B3LYP/6-31G(d) activation barriers for the Diels-Alder reaction with benzo-fused heterocycles are presented in Table 25. These energies are in perfect agreement with our qualitative observations presented above. Regardless of the dienophile selected, the activation energies for its addition to the benzo[b]heterocycle were substantially higher than for its addition to the benzo[c]heterocycle. As expected on the basis of FMO energy analyses, Diels-Alder reactions with acetylene as the dienophile are the least feasible of all cycloaddition reactions studied here. Estimated activation energies for addition to benzo[b]heterocycles were around 40 kcal/mol, which are too high to be achieved under normal reaction conditions. Acetylene addition to benzo[c]furan was computed to have an activation energy of only 18.9 kcal/mol and should be feasible experimentally (Table 25). Generally, activation barriers with ethylene were lower than with acetylene as a dienophile, and with cyclopropene even lower (Table 25). For instance, addition of ethylene to all benzo[c]-fused heterocycles should be experimentally feasible with a higher activation barrier of 24.4 kcal/mol for ethylene addition to benzo[c]thiophene. Of course, for the addition of cyclopropene, even lower activation barriers were computed. Thus 16.9 was the highest

activation barrier for *endo* cyclopropene addition to benzo[*c*]thiophene. Even for such a reactive dienophile such as cyclopropene, the addition to benzo[*b*]heterocycles was predicted to be practical experimentally, with the possible exception of *exo* cyclopropene addition to benzo[*b*]furan for which a 28.4 kcal/mol reaction barrier was estimated (Table 25).

Table 25. The AM1 and The B3LYP/6-31G(d) computed activation barriers for selected Diels-Alder cycloaddition reactions with benzofuran, benzopyrrole, and benzothiophene

reaction type	HOF	E	ΔE_I	ΔE_{II}
acetylene + benzo[<i>b</i>]furan	120.8	-460.928333	45.2	42.4
acetylene + benzo[<i>b</i>]pyrrole	159.7	-441.065949	49.7	46.0
acetylene + benzo[<i>b</i>]thiophene	154.0	-783.912465	56.8	38.7
acetylene + benzo[<i>c</i>]furan	110.8	-460.942991	28.1	18.9
acetylene + benzo[<i>c</i>]pyrrole	150.9	-441.087810	34.4	23.1
acetylene + benzo[<i>c</i>]thiophene	147.9	-783.892088	43.7	39.4
ethylene + benzo[<i>b</i>]furan	75.7	-462.208252	38.4	30.9
ethylene + benzo[<i>b</i>]pyrrole	115.3	-442.328909	43.6	45.2
ethylene + benzo[<i>b</i>]thiophene	110.8	-785.160020	51.9	47.5
ethylene + benzo[<i>c</i>]furan	64.5	-462.208252	20.1	16.7
ethylene + benzo[<i>c</i>]pyrrole	104.8	-442.350109	26.6	22.7
ethylene + benzo[<i>c</i>]thiophene	102.4	-785.177698	36.5	24.4
<i>exo</i> cyclopropene + benzo[<i>b</i>]furan	131.1	-500.242907	35.5	28.4
<i>exo</i> cyclopropene + benzo[<i>b</i>]pyrrole	169.3	-480.379182	39.3	32.8
<i>exo</i> cyclopropene + benzo[<i>b</i>]thiophene	162.7	-823.203175	45.5	39.6
<i>endo</i> cyclopropene + benzo[<i>b</i>]furan	129.6	-500.238488	34.0	31.1
<i>endo</i> cyclopropene + benzo[<i>b</i>]pyrrole	168.1	-480.375076	38.1	35.4
<i>endo</i> cyclopropene + benzo[<i>b</i>]thiophene	161.2	-823.207174	44.0	37.1
<i>exo</i> cyclopropene + benzo[<i>c</i>]furan	121.9	-500.250434	19.2	9.4
<i>exo</i> cyclopropene + benzo[<i>c</i>]pyrrole	160.7	-480.395742	24.2	13.2
<i>exo</i> cyclopropene + benzo[<i>c</i>]thiophene	157.7	-823.221607	33.5	16.0
<i>endo</i> cyclopropene + benzo[<i>c</i>]furan	123.7	-500.247997	21.0	10.9
<i>endo</i> cyclopropene + benzo[<i>c</i>]pyrrole	162.9	-480.391950	26.4	15.6
<i>endo</i> cyclopropene + benzo[<i>c</i>]thiophene	159.6	-823.220062	35.4	16.9

HOF = heat of formation computed by AM1; E = total energy (a.u.) computed by B3LYP/6-31G(d)/AM1; ΔE_I = activation barrier (kcal/mol) computed by AM1; ΔE_{II} = activation barrier (kcal/mol) computed by B3LYP/6-31G(d)/AM1.

Let us now evaluate *exo-endo* selectivity in cyclopropene addition to benzo[*c*]heterocycles. On the basis of bond order analysis, the *exo* addition of cyclopropene to benzo[*c*]furan was selected over the *endo* addition; while in the case of benzo[*c*]pyrrole and benzo[*c*]thiophene, it was suggested that SOI might be responsible for formation of an *exo* cycloadduct product. The computed activation energies for cyclopropene addition to the benzo[*c*]-fused heterocycle clearly favors formation of the *exo* cycloadduct. The activation barrier with benzo[*c*]furan was

predicted to be a mere 9.4 kcal/mol, while the least reactive *exo* cyclopropene addition to benzo[*c*]thiophene was 16.0 kcal/mol. Therefore, our bond order uniformity as well as SOI studies are in agreement with the computed activation barriers for these cycloaddition reactions.

The experimental results fully support our computational studies. Benzo[*b*]furan, benzo[*b*]pyrrole, and benzo[*b*]thiophene do not take part in Diels-Alder reactions. The benzo[*c*]-fused heterocycles function as highly reactive dienes in [4+2] cycloaddition reactions. Thus, all benzo[*c*]furan, isoindole (benzo[*c*]pyrrole) and benzo[*c*]thiophene yield Diels-Alder adducts with reactive dienophiles such as maleic anhydride [56].

4. DIELS-ALDER REACTIONS WITH FIVE-MEMBERED HETEROCYCLES WITH TWO HETEROATOMS

The reactivity of five membered heterocycles with two heteroatoms as dienes with at least one nitrogen for Diels-Alder reactions is also very low. In fact, there is not much experimental data in this area of research, except for addition of dienophiles to oxazole, better known as the Kondratyeva reaction [57]. The main reason for their low reactivity is high heterocycle aromaticity; delocalization of molecular π -orbitals that should be part of the cycloaddition reaction. That can be explained from FMO energy differences between aromatic heterocycles as well as by bond order uniformity of heterocycles with two heteroatoms

The FMO energy gaps between these heterocycles and dienophiles (ethylene, acetylene, and cyclopropene) were compared with the same energy gaps for cyclopentadiene and furan as dienes for the Diels-Alder cycloaddition reaction (Table 26). According to these energy differences, the addition of acetylene to heterocycles with two heteroatoms was HOMO 1,2- and 1,3-diazole controlled. For oxygen and sulfur containing heterocycles it was LUMO diene controlled. The energy gap was higher for all studied heterocycles than it was for cyclopentadiene as diene. Only in two instances, 1,3-diazole and 1,3-thiazole, were heterocycles with two heteroatoms selected on the basis of FMO energy gap as more reactive than furan. In a reaction with a more reactive dienophile such as cyclopropene, the heterocycles with two heteroatoms (1,2-oxazole; 1,2-thiazole; 1,3-thiazole) were predicted to be quite reactive, in some cases, substantially more reactive than cyclopentadiene (Table 26). According to the FMO energy gap between reactants, the 1,2-oxazole, 1,2-thiazole, and 1,3-thiazole should be exceptionally good dienes for the Diels-Alder reaction with moderate and highly reactive dienophiles.

Another static approach that can give us the relative reactivity of heterocycles as dienes for Diels-Alder reactions is evaluation of their aromatic stability through the ring bond order uniformity. If, for a moment, we examine reactivity of the heterocycle on the basis of FMO energy gap with cyclopropene as a dienophile, it is obvious that the most reactive heterocycle is 1,3-thiazole. It had a FMO energy gap of only 9.609 eV (Table 26). That finding is almost

Table 26. Frontier molecular orbital (FMO) energy (eV) gap between the heterocycles with two heteroatoms and acetylene, ethylene, and cyclopropene computed with AM1 semiempirical method

Diene	HOMO	LUMO	A	B	C	D	E	F
CPD	-9.079	0.482	11.132	11.982	10.517	11.033	10.121	10.301
furan	-9.317	0.723	11.370	12.223	10.755	11.274	10.359	10.542
1,2-diazole	-9.706	0.955	11.759	12.455	11.144	11.506	10.748	10.774
1,2-oxazole	-10.466	0.175	12.519	11.675	11.904	10.726	11.508	9.994
1,2-thiazole	-9.540	-0.102	11.593	11.398	10.978	10.449	10.582	9.717
1,3-diazole	-9.159	0.977	11.212	12.477	10.597	11.528	10.201	10.796
1,3-oxazole	-9.891	0.307	11.944	11.807	11.329	10.858	10.933	10.126
1,3-thiazole	-9.698	-0.210	11.751	11.290	11.136	10.341	10.740	9.609

CPD = cyclopentadiene; A = $LUMO_{\text{acetylene}} - HOMO_{\text{heterocycle}}$; B = $LUMO_{\text{heterocycle}} - HOMO_{\text{acetylene}}$; C = $LUMO_{\text{ethylene}} - HOMO_{\text{heterocycle}}$; D = $LUMO_{\text{heterocycle}} - HOMO_{\text{ethylene}}$; E = $LUMO_{\text{cyclopropene}} - HOMO_{\text{heterocycle}}$; F = $LUMO_{\text{heterocycle}} - HOMO_{\text{cyclopropene}}$

diametrically different from our ring bond order uniformity. In this finding, the 1,3-thiazole ring bond order deviation from uniformity was only 0.985 (Table 27). That should make it quite less reactive than both cyclopentadiene and furan. According to our calculations, 1,3-diazole (imidazole) had almost perfect distribution of π -molecular orbitals in the ring, making it the most aromatic of all heterocycles with two heteroatoms studied here. It is, therefore, the least susceptible to participate in the Diels-Alder reactions as a diene. Again, the FMO energy gap between 1,2-diazole and 1,3-diazole with cyclopropene as a dienophile prefers an incorrect isomer (1,3-diazole) to be a better diene although bond order uniformity clearly selects 1,3-diazole as a highly aromatic, low reactive heterocycle. According to bond order uniformity, the most reactive diene for the Diels-Alder reaction was, as expected, cyclopentadiene. But from heterocycles with two heteroatoms, 1,2- and 1,3-oxadiazoles were selected. Although bond order uniformity had a clear advantage in predicting reactivity of the heterocycles over FMO energy gap, it cannot account for the nature of bond formation in the transition state structure because reactants were considered separately. To further investigate what kind of influence the nature of the bond in formation has on the reactivity of heterocycles, we will separately present our computational results for addition of the dienophiles to heterocycles that contain heteroatoms in the 1,2- and 1,3-positions.

One can also determine the reactivity of two isomeric heterocycles in regard to their stability, with the notion that the less stable heterocycle will also be more reactive. That is, of course, not true because the reactivity of these two groups of heterocycles in Diels-Alder reactions do not depend only on their relative energies but also on the nature of the bonds that are formed in the course of reaction. Heterocycles with heteroatoms in the 1 and 2 positions of the ring form one bond

Table 27. The ring bond orders and bond order deviation for heterocycles with two heteroatoms computed with the AM1 semiempirical method

	BO12	BO23	BO34	BO45	BO51	SBO	ABO	BOD
CPD	1.002	1.849	1.061	1.849	1.002	6.763	1.353	1.986
furan	1.104	1.670	1.190	1.670	1.104	6.738	1.348	1.290
1,2-diazole	1.172	1.580	1.262	1.560	1.195	6.769	1.354	0.865
1,2-oxazole	1.108	1.702	1.156	1.689	1.101	6.756	1.351	1.377
1,2-thiazole	1.176	1.640	1.205	1.629	1.178	6.828	1.366	1.076
1,3-diazole	1.171	1.567	1.394	1.554	1.174	6.860	1.372	0.798
1,3-oxazole	1.083	1.676	1.193	1.657	1.103	6.712	1.342	1.296
1,3-thiazole	1.159	1.632	1.257	1.590	1.186	6.824	1.365	0.985

CPD = cyclopentadiene; $BO_{X,Y}$ = bond order between atoms X and Y in the heterocycle ring; SBO = sum of ring bond orders; ABO = average bond order; BOD = sum of bond order deviation from the average ring bond order.

that is a C-N bond. Usually, these transition state structures have substantially higher energies than the isomeric transition state structure with both C-C bonds in formation. Nevertheless, let us determine the relative energies of two isomeric groups of heterocycles (Table 28). The semiempirical AM1 methods selects 1,3-oxazole as the most stable isomer of the six different heterocycles. The B3LYP computed energies agreed that heterocycles with heteroatoms in the 1 and 3 positions are more stable than one with heteroatoms in the 1 and 2 positions. That does not mean that the latter are better suited as dienes for Diels-Alder reactions. The difference in energy of the C-N bond formation in relation to the C-C bond formation might be higher than the difference between two isomeric heterocycles.

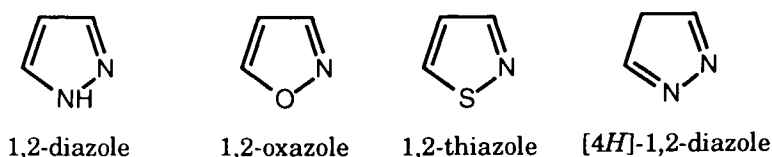
Table 28. The computed energies of different heterocycles with two heteroatoms

Heterocycle	HOF	E	ΔE_I	ΔE_{II}
1,2-diazole	65.6	-226.194086		
1,2-oxazole	42.9	-246.026458		
1,2-thiazole	38.0	-569.038114		
1,3-diazole	50.8	-226.209688	-14.5	-9.8
1,3-oxazole	12.5	-246.064475	-30.4	-23.9
1,3-thiazole	38.6	-569.043560	0.6	-3.4

HOF = heat of formation (kcal/mol) computed with the AM1; Total energies (a.u.) computed with B3LYP/6-31G(d); ΔE_I = energy difference (kcal/mol) between two isomeric heterocycles computed by AM1; ΔE_{II} = energy difference (kcal/mol) between two isomeric heterocycles computed with B3LYP/6-31G(d)//AM1.

4.1. Addition of acetylene, ethylene, and cyclopropene to heterocycles with heteroatoms in the 1 and 2 positions

The structures of heterocycles used for evaluation due to their suitability as dienes for Diels-Alder reactions are presented in Scheme 5. The theoretical investigation of reactivity of [4*H*]-1,2-diazole as a high energy tautomer of 1,2-



Scheme 5. The heterocycles with two heteroatoms in the 1 and 2 positions.

diazole will be presented later on in this chapter. If we consider diene moiety of the heterocycle with localized double bonds as presented in Scheme 5, then we will have formation of two new bonds, one of which must include formation of bond with a heteroatom. In this case, the transition state is expected to be highly unsynchronous and, consequently, the activation barrier for these reactions were predicted to be higher than for all carbon transition state structures. To investigate this postulate, we will explore reactivity of 1,2-diazole, 1,2-oxazole, and 1,2-thiazole through the change of FMO energy going from reactants to the transition state structures, changes of uniformity heterocycle ring bond order going from reactants to the transition state structure, and activation barriers for the Diels-Alder reaction with acetylene, ethylene, and cyclopropene in comparison with the same reaction with cyclopentadiene and furan as dienes.

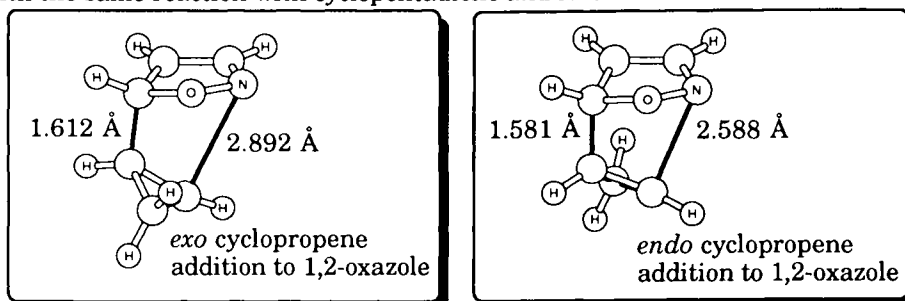


Figure 7. Transition state structures for cyclopropene addition to 1,2-oxazole.

The transition state structures between these heterocycles and dienophiles are quite similar for different dienes. We are, therefore, presenting transition state structures only for the reaction with 1,2-oxazole (Figure 7) because it was predicted on the basis of the ring bond order uniformity to be the most reactive one. The required FMO energy change for transformation of reactants into transition state structures is a much more reliable way to assess reactivity of the heterocycles as dienes than simple FMO energy gap between diene and dienophile. In combination with the Hammond postulate, FMO energy change can select the

reaction pair that requires the smallest energy change and therefore, it should be the most reactive. In our case (Table 29), the least energy change for the reaction with the heterocycle as a diene was for furan 1,2-oxazole or 1,2-thiazole as predicted on the basis of FMO energy gap between reactants (Table 26). This finding was more reasonable, although the FMO energy change going from 1,2-thiazole and cyclopropene to the transition state structure had a FMO energy change that was too small, smaller than for cyclopropene addition to cyclopentadiene which occurs under mild reaction conditions. This result might be an artifact of the AM1 computational study and will be tested by evaluation of activation barrier for the Diels-Alder reaction.

Table 29. Frontier Molecular Orbital (FMO) energy (eV) changes going from reactants to transition state structures with heterocycles with two heteroatoms in the 1 and 2 positions as dienes computed with semiempirical method

TS	HOMO	LUMO	A	B	C	D	Σ_1	Σ_2
Ia	-8.821	0.723	2.679	0.258	1.330	0.241	2.920	1.588
Ib	-8.694	0.665	1.857	0.385	1.388	0.103	1.960	1.773
Ic	-8.515	0.548	1.304	0.564	1.505	0.066	1.370	2.069
Id	-8.632	0.622	1.187	0.447	1.431	0.140	1.327	1.878
IIa	-9.241	0.830	2.259	0.076	1.223	0.107	2.366	1.299
IIb	-9.037	0.791	1.514	0.280	1.262	0.068	1.582	1.542
IIc	-8.764	0.608	1.055	0.553	1.445	-0.115	1.170	1.998
IId	-8.893	0.565	0.926	0.424	1.488	-0.158	1.084	1.912
IIIa	-8.209	-0.484	3.291	1.497	2.537	-1.439	4.730	4.034
IIIb	-7.949	-0.533	2.602	1.757	2.586	-1.488	4.090	4.343
IIIc	-7.502	-1.103	2.317	2.204	3.156	-2.058	4.375	5.360
IIId	7.674	-1.011	2.145	2.032	3.064	-1.966	4.111	5.096
IVa	-8.410	-0.887	3.291	2.056	2.940	-1.062	4.353	4.996
IVb	-8.181	-0.928	2.602	2.285	2.981	-1.103	3.705	5.266
IVc	-7.551	-1.660	2.317	2.915	3.713	-1.835	4.152	6.628
IVd	-7.801	-1.456	2.145	2.665	3.509	-1.631	3.776	6.174
Va	-8.195	-0.837	3.305	1.345	2.890	-0.835	4.140	4.235
Vb	-8.616	0.042	1.935	0.925	2.011	0.144	2.079	2.936
Vc	-8.746	0.061	1.073	0.794	1.992	0.163	1.236	2.786
Vd	-7.445	-1.347	2.374	2.095	3.400	-1.245	3.619	5.495

I = TS with cyclopentadiene; II = TS with furan; III = TS with 1,2-diazole; IV = TS with 1,2-oxazole; V = 1,2-thioazole; a = for addition of acetylene; b = for addition of ethylene; c = for *exo* addition of cyclopropene; d = for *endo* addition of cyclopropene; A = $\text{HOMO}_{\text{TS}} - \text{HOMO}_{\text{dienophile}}$; B = $\text{HOMO}_{\text{TS}} - \text{HOMO}_{\text{heterocycle}}$; C = $\text{LUMO}_{\text{dienophile}} - \text{LUMO}_{\text{TS}}$; D = $\text{LUMO}_{\text{TS}} - \text{LUMO}_{\text{heterocycle}}$; $\Sigma_1 = |A| + |D|$; $\Sigma_2 = |B| + |C|$.

A new approach that we would like to introduce here to evaluate the reactivity of aromatic heterocycles that have asymmetric transition state structures is a comparison of the changes of the ring bond order between the heterocycle and

heterocycle in the transition state structure. By applying the Hammond postulate, the transition state structure that will have a heterocycle with the least required ring bond order reorganization should be electronically closest to the reactants and should have the lowest activation barrier. This heterocycle should be the most reactive as a diene for the Diels-Alder reaction. The AM1 computed ring bond order changes are presented in Table 30. As one would expect, bond orders BO₂₃ and BO₄₅ were decreasing in the course of the reaction (negative bond orders). In conventional bond order representation, it is a transformation of two double bonds of dienophile into two single bonds. At the same time, two new single bonds and one new double bond (BO₃₄) are formed. To reach the transition state at some point, this rearrangement of electron density on the heterocycle moiety is necessary.

Table 30. Bond orders and bond order deviation of the heterocycle ring going from reactant to the transition state structure

TS	BOC ₁₂	BOC ₂₃	BOC ₃₄	BOC ₄₅	BOC ₅₁	SBOC
I _a	-0.015	-0.465	0.385	-0.465	-0.465	1.345
I _b	-0.013	-0.456	0.373	-0.456	-0.013	1.311
I _c	-0.008	-0.416	0.329	-0.416	-0.008	1.177
I _d	-0.013	-0.413	0.321	-0.413	-0.013	1.173
II _a	-0.038	-0.434	0.399	-0.434	-0.038	1.343
II _b	-0.039	-0.424	0.387	-0.424	-0.039	1.313
II _c	-0.023	-0.384	0.340	-0.384	-0.024	1.155
II _d	-0.034	-0.386	0.339	-0.386	-0.035	1.180
III _a	-0.037	-0.337	0.385	-0.479	-0.221	1.459
III _b	-0.033	-0.304	0.342	-0.451	-0.207	1.337
III _c	0.008	-0.259	0.296	-0.475	-0.224	1.262
III _d	-0.019	-0.272	0.311	-0.495	-0.248	1.345
IV _a	-0.015	-0.392	0.417	-0.585	-0.121	1.530
IV _b	-0.013	-0.333	0.351	-0.535	-0.105	1.337
IV _b	0.003	-0.207	0.224	-0.532	-0.118	1.078
IV _d	-0.007	-0.281	0.299	-0.596	-0.146	1.329
V _b	-0.088	-0.36	0.369	-0.400	-0.123	1.340
V _c	-0.063	-0.372	0.370	-0.398	-0.084	1.287
V _d	-0.172	-0.492	0.300	-0.267	-0.009	1.240

I = TS with cyclopentadiene; II = TS with furan; III = TS with 1,2-diazole; IV = TS with 1,2-oxazole; V = 1,2-thioazole; a = for addition of acetylene; b = for addition of ethylene; c = for *exo* addition of cyclopropene; d = for *endo* addition of cyclopropene; BOC_{X.Y} = bond order change for bonds between atoms X and Y in the heterocycle ring required to achieve transition state structures; SBOC = sum of the heterocycle bond order changes.

The minimal bond order changes for transformation of a dienophile into transition state structures perfectly agree with computed activation barriers for the addition of ethylene, acetylene, and cyclopropane to cyclopentadiene and furan

as a diene. Ethylene is always least reactive, while cyclopropene is the most reactive diene. Two possible isomeric transition state structures for the cyclopropene addition to cyclopentadiene require almost identical bond order changes on the diene ring, indicating that formation of the mixture of *exo* and *endo* cycloadducts should be formed with a slight excess of the *exo* cycloadduct. On the other hand, the formation of the *exo* cycloadduct with furan as a diene and cyclopropene as a dienophile is preferable and should be the major product of the reaction. Both 1,2-diazole and 1,2-thiazole are predicted to be less reactive than furan as a dienophile for the Diels-Alder reaction. Although 1,2-oxazole was predicted to be less reactive as a diene than furan in the reactions with acetylene, ethylene, and for the *endo* addition with cyclopropene, it was, surprisingly, estimated to be more reactive than furan for the *exo* cyclopropane addition reaction.

To determine the reliability of these computational approaches, we have computed activation barriers for these reactions at both AM1 *ab initio* and B3LYP/6-31G(d) DFT theory levels (Table 30). Knowing that activation barriers for the addition of cyclopropene to furan computed at the B3LYP/6-31G(d)/AM1 theory level were 18.7 and 18.4 kcal/mol for an *endo* and *exo* cycloaddition reaction (Table 8) and it is experimentally feasible, it becomes obvious that none of the cycloaddition reactions presented in Table 25 should be able to be accomplished experimentally. All activation barriers were around 40 kcal/mol or higher, with the exception of the cyclopropane addition to 1,2-oxazole (Table 31). The comparison

Table 31. The AM1 and the B3LYP/6-31G(d) computed activation barriers for the acetylene, ethylene, and cyclopropene additions to 1,2- diazole, oxadiazole and thiazole

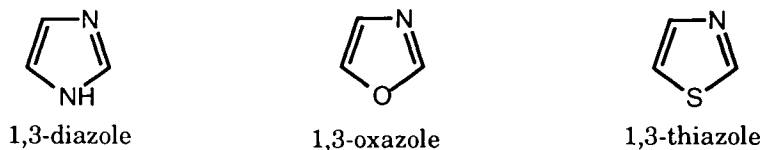
reactant or reaction	HOF	E	ΔE_I	ΔE_{II}
acetylene + 1,2-diazole	178.3	-303.442657	57.9	48.2
ethylene + 1,2-diazole	133.0	-304.704633	50.9	48.0
<i>exo</i> cyclopropene + 1,2-diazole	183.2	-342.752184	42.8	37.3
<i>endo</i> cyclopropene + 1,2-diazole	183.0	-342.754897	42.6	35.6
acetylene + 1,2-oxazole	148.3	-323.281919	50.6	43.9
ethylene + 1,2-oxazole	102.8	-324.546103	43.4	42.3
<i>exo</i> cyclopropene + 1,2-oxazole	154.0	-362.588141	36.3	35.1
<i>endo</i> cyclopropene + 1,2-oxazole	153.4	-362.593658	35.7	31.6
acetylene + 1,2-thiazole	158.5	-646.261924	65.7	63.8
ethylene + 1,2-thiazole	116.7	-647.549554	62.2	47.4
<i>exo</i> cyclopropene + 1,2-thiazole	170.6	-685.594340	57.8	38.5
<i>endo</i> cyclopropene + 1,2-thiazole	164.6	-685.575508	51.8	50.3

HOF = heat of formation computed by AM1; E = total energy (a.u.) computed with B3LYP/6-31G(d)/AM1; ΔE_I = activation barrier (kcal/mol) computed by AM1; ΔE_{II} = activation barrier (kcal/mol) computed with B3LYP/6-31G(d)/AM1.

of the reactivity through minimal ring bond order rearrangement is appropriate only if the reactivities of 1,2-diazole, 1,2-oxazole, and 1,2-thiazole are compared

among themselves. This method turned out to be unreliable for comparisons of reactivity among these three heterocycles with furan or cyclopentadiene because, in the course of the reaction with those dienes, two C-C bonds were formed, contrary to 1,2-heterocycles where only one C-heteroatom bond was formed. Therefore, we can conclude on the basis of bond order and FMO energy change that 1,2-oxazole is most reactive; however, on the basis of the computed activation barriers, even this reaction should not be experimentally feasible.

4.2. Addition of acetylene, ethylene, and cyclopropene to heterocycles with heteroatoms in the 1 and 3 positions



Scheme 6. The heterocycles with two heteroatoms in the 1 and 3 positions

By using 1,3-substituted heterocycles (Scheme 6), we have the ability to perform Diels-Alder cycloaddition reactions in which both formed bonds will be C-C bonds; therefore, their reactivity can be compared with a higher degree of certainty to the reactivity of heterocycles with one heteroatom or to classical dienophiles

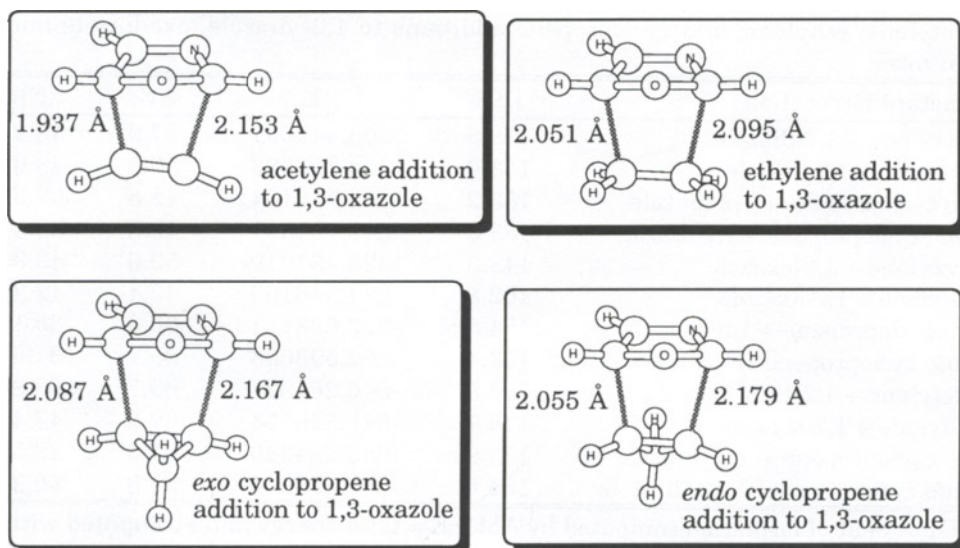


Figure 8. Transition state structures for acetylene, ethylene, and cyclopropene addition to 1,3-oxazole.

such as cyclopropene. All computed transition state structures are much more synchronous (Figure 8), than in the case of heterocycles with a heteroatom in 1,3-position (Figure 7). In fact, they are very similar to the transition state structures for the furan reaction with the same dienophiles. Therefore, we should be able to closely analyze the reactivity of these heterocycles as dienes for Diels-Alder reactions in comparison with the known reactivity of dienes such as cyclopentadiene and furan. To achieve this goal, we used two of our approaches; the necessary FMO energy change for transformation of reactants into corresponding transition state structures, and the ring bond order reorganization required for the heterocycle to achieve the transition state reorganization.

For the sake of simplicity in our presentation, we are presenting only our results for cyclopropene addition to the heterocycles. All conclusions drawn from this study can be applied to the addition of acetylene and ethylene to heterocycles, knowing that cyclopropene was the most reactive while acetylene was the least reactive dienophile. The frontier molecular orbital energies and frontier orbital energy changes for *exo* and *endo* cyclopropene addition to heterocycles which have heteroatoms in 1,3-position are presented in Table 32. The cycloaddition reaction with 1,3-diazole is for normal, LUMO dienophile and HOMO diene controlled Diels Alder reactions. On the other hand, both cycloaddition reactions with 1,3-oxazole and 1,3-thiazole were LUMO diene, and HOMO dienophile controlled. In all three heterocycles, in order to reach the *exo* transition state structure, a lower energy change was required than was needed for isomeric *endo* transition state structures. So, as the Hammond postulate states, the *exo* transition state structure should have had lower energies. According to the FMO energy change, the most reactive was 1,3-oxazole, while the least reactive dienophile was 1,3-thiazole (Table 32). This reactivity order might not be correct because the same approach selected 1,3-oxazole to be more reactive than furan, which should not be true based on the experimental observations.

Table 32. Frontier molecular orbital (FMO) energy (eV) for the transition state structure and the necessary changes in reactants to reach this transition state of cyclopropene addition to heterocycles which have heteroatoms in their 1,3-position

TS	HOMO	LUMO	A	B	C	D	Σ_1	Σ_2
I _{exo}	-8.777	0.422	1.042	0.382	0.620	-0.555	1.597	1.002
I _{endo}	-8.723	0.364	1.096	0.436	0.678	-0.613	1.709	1.114
II _{exo}	-9.096	0.205	0.723	0.795	0.837	-0.102	0.825	1.632
II _{endo}	-9.074	0.128	0.745	0.817	0.914	-0.179	0.924	1.731
III _{exo}	-8.890	-0.031	0.929	0.808	1.073	0.179	1.108	1.881
III _{endo}	-8.675	-0.121	1.144	1.023	1.163	0.089	1.233	2.186

I = TS with 1,3-diazole as diene; II = TS with 1,3-oxazole as diene; III = TS with 1,3-thiazole as diene; A = HOMO_{TS} - HOMO_{dienophile}; B = HOMO_{TS} - HOMO_{heterocycle}; C = LUMO_{dienophile} - LUMO_{TS}; D = LUMO_{TS} - LUMO_{heterocycle}; $\Sigma_1 = |A| + |D|$; $\Sigma_2 = |B| + |C|$.

The other method to determine reactivity for reactions with synchronous concerted cyclic transition state structures is evaluation of the transition state ring aromaticity through bond order deviation. The results of the *exo* cyclopropene addition to the heterocycles and to cyclopentadiene are presented in Table 33. The higher the sum of bond order deviation from average bond order (x) is, the lower aromatic character the transition state structure has. The most reactive dienophile was cyclopentadiene, followed by furan, and then heterocycles. The most reactive heterocycle with heteroatoms in 1,3-position was 1,3-oxazole as was predicted on the basis of the FMO energy changes (Table 32). The least reactive was 1,3-diazole, as one would expect on the basis of experimental observations. It is very difficult to rely on the transition state structure bond order deviation to determine the experimental feasibility of a reaction but, because SBOD for furan and 1,3-oxadiazole were very similar, one can conclude that the cycloaddition with 1,3-oxadiazole is also experimentally feasible.



Table 33. Bond order uniformity for a six-membered transition state structure of an *exo* cyclopropane addition to a heterocycle with two heteroatoms

diene	BOC ₂₃	BOC ₃₄	BOC ₄₅	BOC ₅₆	BOC ₆₇	BOC ₂₇	SBOD
cyclopentadiene	1.433	1.390	1.433	0.362	1.482	0.362	0.234
furan	1.286	1.530	1.286	0.378	1.471	0.378	0.449
1,3-diazole	1.235	1.620	1.181	0.451	1.400	0.413	0.702
1,3-oxadiazole	1.298	1.545	1.243	0.406	1.455	0.378	0.487
1,3-thiazole	1.274	1.598	1.208	0.440	1.411	0.420	0.603

Average bond order deviation is computed from the formula $6X + 4 = \text{sum of the ring bond orders}$ as explained in Scheme 4; BOC_{X.Y} = bond order change of bonds between atoms X and Y in the heterocycle ring required to achieve transition state structures; SBOD = sum of the bond order deviations.

To verify our qualitative reactivity assessment on the basis of FMO energy changes and transition state bond order deviation, we computed activation barriers for those Diels-Alder reactions (Table 34). It was quite obvious that the order of reactivity of dienophiles, regardless of the heterocycle, was cyclopropene, ethylene, and then acetylene. This order of reactivity was true for many Diels-Alder reactions. The predicted order of reactivity through the FMO energy changes (Table 32) is in full agreement with computed activation barriers: the most reactive diene was 1,3-oxazole whereas 1,3-thiazole was the least reactive heterocycle (Table 34). Considering activation barriers, cyclopropene addition should be experimentally feasible to all three heterocycles, while in the case of alkenes, the reaction that should have practical applications should be the reaction with 1,3-oxazole as a dienophile. This was also confirmed experimentally [57].

Table 34. The AM1 and the B3LYP/6-31G(d) computed activation barriers for acetylene, ethylene, and cyclopropene addition to 1,3-diazole, oxadiazole and thiazole

reactant or reaction	HOF	E	ΔE_I	ΔE_{II}
acetylene + 1,3-diazole	148.0	-303.481102	42.4	33.9
ethylene + 1,3-diazole	101.8	-304.747668	34.5	30.8
<i>exo</i> cyclopropene + 1,3-diazole	157.0	-342.793393	31.4	21.2
<i>endo</i> cyclopropene + 1,3-diazole	158.5	-342.790726	32.9	22.9
acetylene + 1,3-oxazole	103.9	-323.342976	36.6	29.4
ethylene + 1,3-oxazole	57.0	-324.612028	28.0	24.8
<i>exo</i> cyclopropene + 1,3-oxazole	113.4	-362.655323	26.1	16.8
<i>endo</i> cyclopropene + 1,3-oxazole	115.1	-362.652411	27.8	18.6
acetylene + 1,3-thiazole	142.3	-646.308854	48.9	37.7
ethylene + 1,3-thiazole	96.6	-647.579750	41.5	31.9
<i>exo</i> cyclopropene + 1,3-thiazole	152.2	-685.623616	38.8	23.5
<i>endo</i> cyclopropene + 1,3-thiazole	152.7	-685.622401	39.3	24.3

HOF = heat of formation computed by AM1; E = total energy (a.u.) computed by B3LYP/6-31G(d)/AM1; ΔE_I = activation barrier (kcal/mol) computed by AM1; ΔE_{II} = activation barrier (kcal/mol) computed with B3LYP/6-31G(d)/AM1

5. DIELS-ALDER REACTIONS WITH FIVE-MEMBERED HETEROCYCLES WITH THREE HETEROATOMS

There is not much literary available in the literature on five-membered heterocycles with three heteroatoms as dienes for Diels-Alder reactions. There are some reports that focus on specially activated 1,3,4-oxadiazoles that participate as dienes in the Diels-Alder reactions. But to the best of our knowledge, there are no experimental results that would suggest using five-membered heterocycles with heteroatoms in the 1,2,3-positions as dienes for Diels-Alder reactions. Keeping in mind the fact that one C-heteroatom bond should be formed in the course of the reaction, this reaction, though highly valuable for the experimental organic chemist, should not be experimentally feasible because of the high activation barrier.

As the first, simple step, which is to determine the reactivity of heterocycles with three heteroatoms, we are presenting our results on FMO energy gap between reactants and bond order uniformity of the heterocyclic ring. The FMO energy gap between reactants is the approach that might give reasonable relative reactivity values for very similar compounds if there are no substantial electronic rearrangement or steric interactions in the course of the reaction. The addition of acetylene and ethylene to these heterocycles is LUMO heterocycle (diene) and HOMO dienophile controlled Diels-Alder reaction. With more a reactive dienophile such as cyclopropene, FMO control of the reaction changes from one heterocycle to the other. For example, in the case of triazoles, the reaction is a HOMO

heterocycle controlled cycloaddition reaction while for oxadiazoles and thiadiazoles, the reaction is LUMO heterocycle (diene) controlled. It was also predicted that oxadiazole should be the most reactive and triazole should be least reactive. The most reactive heterocycle in this group was predicted to be 1,2,3-oxadiazole (Table 35).

Table 35. Frontier molecular orbital (FMO) energy (eV) gap between the heterocycles with three heteroatoms and acetylene, ethylene, and cyclopropene computed by AM1 semiempirical method

TS with	HOMO	LUMO	A	B	C	D	E	F
I	-10.183	0.423	12.236	11.923	11.621	10.974	9.141	10.242
II	-11.014	-0.387	13.067	11.113	12.452	10.164	9.972	9.432
III	-10.221	-0.684	12.274	10.816	11.659	9.867	9.179	9.135
IV	-10.331	0.545	12.384	12.045	11.769	11.096	11.373	10.364
V	-11.894	-0.378	13.947	11.122	13.332	10.173	12.936	9.441
VI	-9.886	-0.450	11.939	11.050	11.324	10.101	10.928	9.369
VII	-10.033	0.516	12.086	12.016	11.471	11.067	8.991	10.335
VIII	-10.828	-0.177	12.881	11.323	12.266	10.374	9.786	9.642
IX	-10.849	-0.733	12.902	10.767	12.287	9.818	9.807	9.086

I = 1,2,3-triazole; II = 1,2,3-oxadiazole; III = 1,2,3-thiadiazole; IV = 1,2,5-triazole; V = 1,2,5-oxadiazole; VI = 1,2,5-thiadiazole; VII = 1,3,4-triazole; VIII = 1,3,4-oxadiazole; IX = 1,3,4-thiadiazole; A = LUMO_{acetylene} - HOMO_{heterocycle}; B = LUMO_{heterocycle} - HOMO_{acetylene}; C = LUMO_{ethylene} - HOMO_{heterocycle}; D = LUMO_{heterocycle} - HOMO_{ethylene}; E = LUMO_{cyclopropene} - HOMO_{heterocycle}; F = LUMO_{heterocycle} - HOMO_{cyclopropene}

The ring bond order deviation from uniformity partially agreed with the order of reactivity computed on the basis of FMO energy gaps. The least aromatic was 1,2,5-oxadiazole, while 1,2,3-thiadiazole should be most aromatic (Table 36). The order of reactivity was oxadiazole, triazole, thiadiazole in all 1,2,3-, 1,2,5- and 1,3,4-series of the three heteroatom heterocycles. Except for 1,3,4-oxadiazole, the two other 1,3,4- five-membered heterocycles were predicted to be more reactive than their 1,2,3- isomers (Table 36). The prediction that 1,2,5-oxadiazole was the most reactive heterocycle as a diene for Diels-Alder reaction was unacceptable due to the fact that two C-N bonds should be formed in the course of the reaction, which usually requires an exceptionally high activation barrier.

Let us now explore the relative energies of these three groups of heterocycles. We have to be aware of the fact that only for heterocycles with heteroatoms simultaneously in the 1,3, and 4 positions two CC bonds are formed, while in the case of heterocycles with heteroatoms in 1,2, and 5 positions, two CN bond should be formed. The heterocycles with heteroatoms in the 1,2, and 3 positions are between these two cases. We have demonstrated above that heterocycles that have heteroatom-heteroatom bonds are usually less stable when compared with those that have heteroatoms separated by carbon atoms. As a result, one can

select 1,3,4-heterocycles as the most stable of the five-membered heterocycles with three heteroatoms. The relative energies for these three groups of

Table 36. The ring bond orders and bond order deviation for heterocycles with three heteroatoms computed by AM1 semiempirical method

Heterocycle	BO ₁₂	BO ₂₃	BO ₃₄	BO ₄₅	BO ₅₁	SBO	ABO	BOD
1,2,3-triazole	1.151	1.625	1.244	1.565	1.190	6.775	1.355	0.960
1,2,3-oxadiazole	1.077	1.751	1.145	1.679	1.105	6.757	1.351	1.454
1,2,3-thiadiazole	1.692	1.243	1.411	1.389	1.680	7.415	1.483	0.812
1,2,5-triazole	1.181	1.588	1.235	1.588	1.181	6.773	1.355	0.934
1,2,5-oxadiazole	1.100	1.734	1.111	1.734	1.100	6.779	1.356	1.513
1,2,5-thiadiazole	1.155	1.689	1.126	1.689	1.155	6.814	1.363	1.298
1,3,4-triazole	1.171	1.558	1.295	1.558	1.170	6.762	1.350	0.963
1,3,4-oxadiazole	1.089	1.653	1.219	1.653	1.089	6.703	1.341	1.250
1,3,4-thiadiazole	1.176	1.590	1.302	1.590	1.176	6.834	1.367	0.893

BO_{X,Y} = bond order between atoms X and Y in the heterocycle ring; SBO = sum of ring bond orders; ABO = average bond order; BOD = sum of bond order deviation from the average ring bond order.

heterocycles confirmed this speculation (Table 37). The most stable heterocycle should be 1,3,4-oxadiazole. As mentioned above, that does not necessarily mean that this heterocycle is also the least reactive in its role as a diene for Diels-Alder reactions. This finding will be defended later by the computation of activation barriers with various dienophiles.

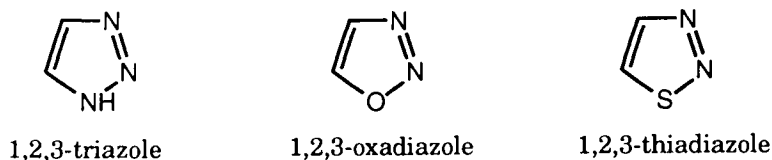
Table 37. The computed energies of different heterocycles with three heteroatoms

Heterocycle	HOF	E	ΔE_I	ΔE_{II}	ΔE_{III}	ΔE_{IV}
1,2,3-triazole	86.4	-242.217031				
1,2,3-oxadiazole	62.1	-262.054621				
1,2,3-thiadiazole	59.9	-585.053271				
1,2,5-triazole	92.3	-242.224633	5.9	-4.8		
1,2,5-oxadiazole	85.5	-262.039387	23.4	9.6		
1,2,5-thiadiazole	49.8	-585.077517	-10.1	-15.2		
1,3,4-triazole	72.9	-242.229057	-13.5	-7.5	-19.4	-2.8
1,3,4-oxadiazole	32.9	-262.082599	-29.2	-17.6	-52.6	-27.1
1,3,4-thiadiazole	62.0	-585.060191	2.1	-4.3	12.2	10.9

HOF = heat of formation (kcal/mol) computed by the AM1; Total energies (a.u.) computed by B3LYP/6-31G(d); ΔE_I = relative energy (kcal/mol) with respect to 1,2,3-heterocycles computed by AM1; ΔE_{II} = relative energy (kcal/mol) with respect to 1,2,3-heterocycles computed by B3LYP/6-31G(d)//AM1; ΔE_{III} = relative energy (kcal/mol) with respect to 1,2,5-heterocycles computed by AM1; ΔE_{IV} = relative energy (kcal/mol) with respect to 1,2,5-heterocycles computed by B3LYP/6-31G(d)//AM1

5.1. Addition of acetylene, ethylene, and cyclopropene to heterocycles with heteroatoms in 1, 2, and 3 positions

The structures of heterocycles used for evaluation due to their suitability as dienes for Diels-Alder reactions are presented in Scheme 6. It is easy to see that if



Scheme 6. The heterocycles with three heteroatoms in 1, 2, and 3 positions.

we use these heterocycles as dienes for Diels-Alder reactions, the formation of a C-N bond is inevitable. From our studies presented above, we know that the transition state structures with heterocycles that include formation of the C-N bond have considerably higher energies than isomeric transition state structures that have two C-C bonds formed. The transition state structures for the addition of dienophiles to heterocycles presented in Scheme 6 are quite similar to the transition state structures for heterocycles with heteroatoms in the 1,2-position (Figure 7). The transition state structures are highly asymmetric with almost fully formed C-C bond and starting to form C-N bond.

As we have discussed above, the FMO energy gap between reactants is not a very reliable approach for evaluation of relative reactivity of heterocycles because the electronic as well as the steric interactions that can stabilize or destabilize transition state structures are not incorporated in these calculations. A much better approach should be based on the necessary FMO energy changes going from reactants into transition state structures. Because all results obtained with acetylene and ethylene as dienophiles were similar to the one for cyclopropane as a dienophile, only results including the last dienophile and heterocycles with heteroatoms in 1,2,3 positions FMO energy changes are presented (Table 38). It is clear that necessary FMO energy changes for reactants to become part of the transition state structures suggest that reactivity of the heterocycles with three heteroatoms in 1,2,3- positions is low. All of these reactions are HOMO dienophile and LUMO diene (heterocycle) controlled cycloaddition reactions. The most reactive heterocycle as a diene for Diels-Alder reaction was 1,2,3-thiadiazole

The bond order uniformity for the transition state with those heterocycles is not an adequate approach for evaluation of their reactivity because their transition state structures are highly asynchronous. Therefore, we have evaluated their reactivity by estimating the activation barriers for the cycloaddition reactions with acetylene, ethylene, and cyclopropene as dienophiles (Table 39). All computed activation barriers were very high and are not expected to be experimentally feasible. That is fully supported by the fact that, at this moment, there is no experimental evidence that five-membered heterocycles with

Table 38. Frontier molecular orbital (FMO) energy (eV) for the transition state structure and necessary changes to reach the corresponding transition state with cyclopropene as dienophile

TS	HOMO	LUMO	A	B	C	D	Σ_1	Σ_2
I _{exo}	-8.764	0.608	1.055	0.553	1.445	-0.115	1.170	1.998
I _{endo}	-8.893	0.565	0.926	0.424	1.488	-0.158	1.084	1.912
II _{exo}	-7.997	-1.423	1.822	2.186	2.465	1.846	3.668	4.651
II _{endo}	-8.121	-1.358	1.698	2.062	2.400	1.781	3.479	4.462
III _{exo}	-7.947	-2.073	1.872	3.067	3.115	1.686	3.558	6.182
III _{endo}	-8.287	-1.773	1.532	2.727	2.815	1.386	2.918	5.542
IV _{exo}	-8.160	-1.798	1.659	2.061	2.840	1.114	2.773	4.901
IV _{endo}	-8.086	-1.890	1.733	2.135	2.932	1.206	2.939	5.067

I = TS with furan as diene; II = TS with 1,2,3-diazole as diene; III = TS with 1,2,3-oxazole as diene; IV = TS with 1,2,3-thiazole as diene; A = HOMO_{TS} - HOMO_{cyclopropene}; B = HOMO_{TS} - HOMO_{heterocycle}; C = LUMO_{cyclopropene} - LUMO_{TS}; D = LUMO_{TS} - LUMO_{heterocycle}; $\Sigma_1 = |A| + |D|$; $\Sigma_2 = |B| + |C|$

heteroatoms in 1,2,3 positions are capable to participate in Diels-Alder reaction as dienes. For instance, the lowest activation barrier, 31.2 kcal/mol, computed for the *endo* cyclopropene addition to 1,2,3-oxadiazole was almost twice higher than for the same addition to furan. The 1,2,3-oxadiazole was selected to be the most reactive heterocycle of those three heterocycles (Table 38), in contrast to the FMO energy changes.

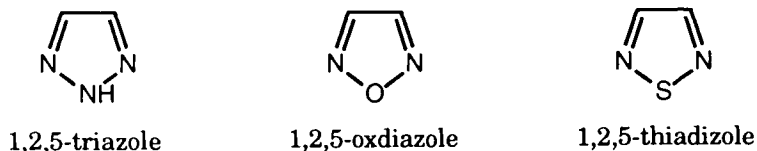
Table 39. The AM1 and B3LYP/6-31G(d) computed activation barriers for selected Diels-Alder cycloaddition reactions with benzofuran, benzopyrrole, and benzothiophene as dienes

Reactant or reaction	HOF	E	ΔE_I	ΔE_{II}
acetylene + 1,2,3-triazole	200.8	-319.467266	59.6	47.2
ethylene + 1,2,3-triazole	155.4	-320.726636	52.5	48.6
<i>exo</i> cyclopropene + 1,2,3-triazole	204.6	-358.776302	43.4	36.6
<i>endo</i> cyclopropene + 1,2,3-triazole	204.3	-358.776796	43.1	36.3
acetylene + 1,2,3-oxadiazole	166.9	-339.312027	50.0	42.7
ethylene + 1,2,3-oxadiazole	120.9	-340.573175	42.3	43.0
<i>exo</i> cyclopropene + 1,2,3-oxadiazole	170.6	-378.616001	33.7	35.2
<i>endo</i> cyclopropene + 1,2,3-oxadiazole	170.4	-378.622481	33.5	31.2
acetylene + 1,2,3-thiadiazole	184.5	-662.302830	69.8	47.6
ethylene + 1,2,3-thiadiazole	133.2	-663.561217	56.8	49.6
<i>exo</i> cyclopropene + 1,2,3-thiadiazole	180.1	-701.614789	45.4	35.2
<i>endo</i> cyclopropene + 1,2,3-thiadiazole	179.5	-701.614066	44.8	35.6

HOF = heat of formation computed by AM1; E = total energy (a.u.) computed by B3LYP/6-31G(d)/AM1; ΔE_I = activation barrier (kcal/mol) computed by AM1; ΔE_{II} = activation barrier (kcal/mol) computed with B3LYP/6-31G(d)/AM1

5.2. Addition of cyclopropene to heterocycles with heteroatoms in the 1, 2, and 5 positions

These heterocycles (Scheme 7) have a plane of symmetry perpendicular to the plane of the molecule; therefore, they are expected to form symmetric transition state structures for cycloaddition reactions with symmetric dienophiles such as



Scheme 7. The heterocycles with three heteroatoms in 1, 2, and 5 position.

acetylene, ethylene, and cyclopropene. Formation of two C-N bond in the heterocycle should be very unfavorable, therefore the activation barrier for this reaction should be very high. The FMO energy gap, as well as the bond uniformity of the heterocyclic rings were not adequate to explain the low reactivity of 1,2,5-heterocycles.

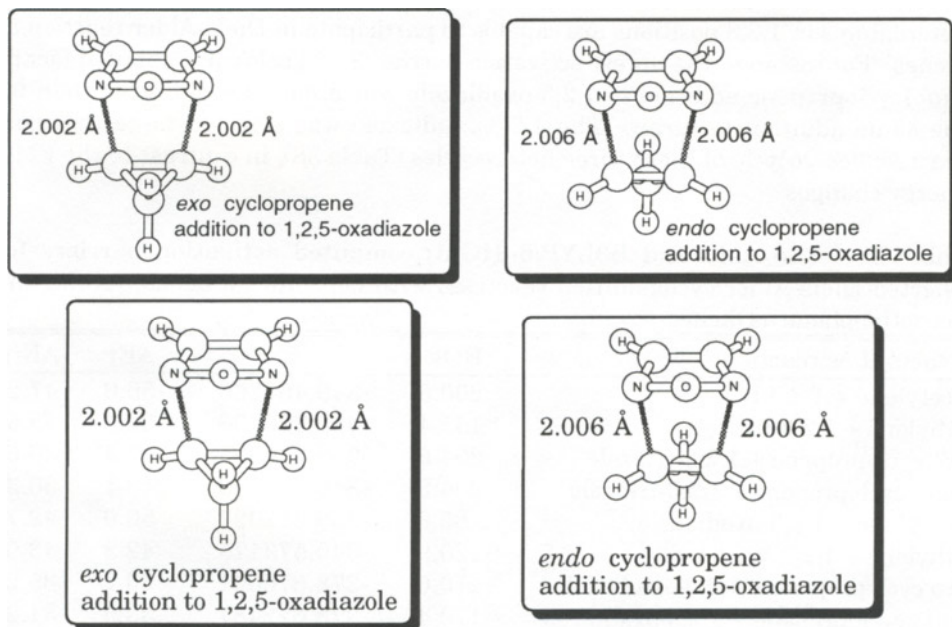


Figure 9. Transition state structures for cyclopropene addition to 1,2,5-oxadiazole.

Transition state structures are for synchronous concerted mechanism of Diels Alder reactions and are very similar to each other. The two isomeric structures for the cyclopropene addition to 1,2,5-oxadiazole are presented in Figure 9. Sometimes the AM1 semiempirical method can compute asymmetric transition state structures that are, in energy, very close to the symmetric transition state structures presented in Figure 9.

Here we present only activation barriers for heterocycles with heteroatoms in 1,2,5- positions in reaction with the most reactive dienophile, cyclopropene (Table 40). It demonstrates our previous postulate that the activation barriers were too high, even with very reactive dienophiles such as cyclopropene. The most reactive was 1,2,5-oxadiazole but not 1,2,5-thiadiazole as predicted on the basis of FMO energy gap between reactants. The results indicated that the reaction should not be experimentally achievable. To the best of our knowledge, there is no experimental evidence that heterocycles with heteroatoms in 1,2,5- positions might be acceptable dienes for Diels-Alder reactions.

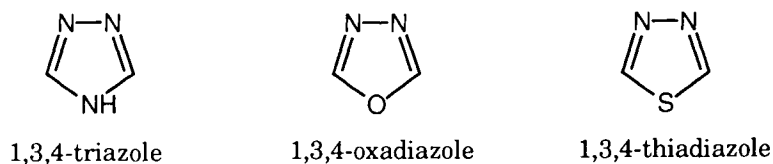
Table 40. The AM1 and the B3LYP/6-31G(d) computed activation barriers for selected Diels-Alder cycloaddition reactions with benzofuran, benzopyrrole, and benzothiophene as dienes

Reactant or reaction	HOF	E	ΔE_I	ΔE_{II}
<i>exo</i> cyclopropene + 1,2,5-triazole	229.3	-358.759817	62.2	51.7
<i>endo</i> cyclopropene + 1,2,5-triazole	229.2	-358.762127	62.1	50.2
<i>exo</i> cyclopropene + 1,2,5-oxadiazole	220.6	-378.598026	60.4	37.0
<i>endo</i> cyclopropene + 1,2,5-oxadiazole	221.7	-378.597956	61.5	37.0
<i>exo</i> cyclopropene + 1,2,5-thiadiazole	200.7	-701.614556	76.1	50.5
<i>endo</i> cyclopropene + 1,2,5-thiadiazole	199.7	-701.616743	75.1	49.1

HOF = heat of formation computed by AM1; E = total energy (a.u.) computed by B3LYP/6-31G(d)/AM1; ΔE_I = activation barrier (kcal/mol) computed by AM1; ΔE_{II} = activation barrier (kcal/mol) computed by B3LYP/6-31G(d)/AM1

5.3. Addition of acetylene, ethylene, and cyclopropene to heterocycles with heteroatoms in 1, 2, and 4 positions

The heterocycles with heteroatoms in 1, 2, and 4 positions (Scheme 8) have the highest probability to be engaged in Diels-Alder reactions as dienes because two new CC bonds should be formed in the course of the reaction. Due to their symmetry, one can also expect that transition states for reactions with symmetric dienophiles should be symmetric and support synchronous formation of both C-C bonds. The AM1 computed transition state structures



Scheme 8. The heterocycles with three heteroatoms in 1, 3, and 4 position.

fully supported this assumption. All transition state structures with 1,3,4-triazole, and 1,3,4-thiadiazole as dienes were similar to the transition state structures for the acetylene, ethylene, and cyclopropene additions to 1,3,4-oxadiazole as a diene, presented in Figure 10.

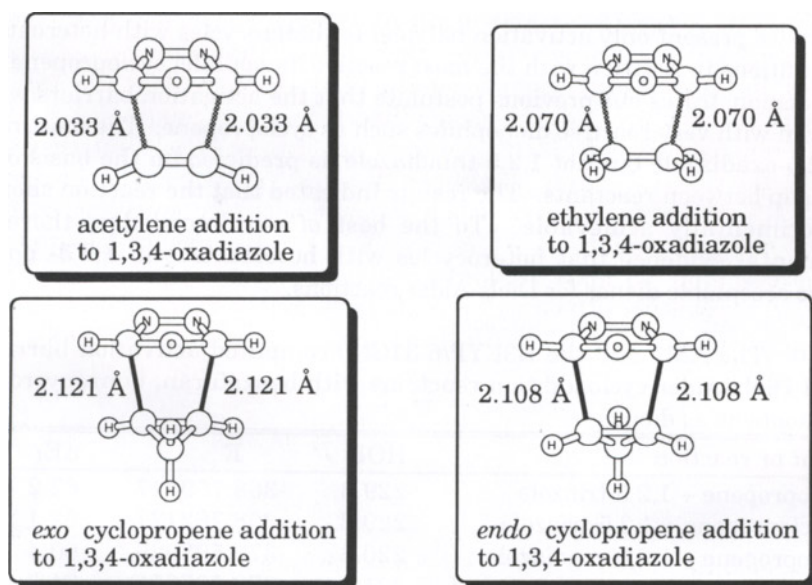


Figure 10. The AM1 computed transition state structures for ethylene, ethylene, and cyclopropene addition to 1,3,4-oxadiazole.

Table 41. Frontier Molecular Orbital (FMO) energy (eV) changes going from reactants to transition state structures with heterocycles with three heteroatoms in 1, 3, and 4 positions as dienes for Diels-Alder reaction. The values were computed by the AM1 semiempirical method

TS	HOMO	LUMO	A	B	C	D	Σ_1	Σ_2
Ia	-9.838	0.385	1.662	0.195	1.668	-0.131	1.793	1.863
Ib	-9.697	0.372	0.854	0.336	1.066	-0.144	0.998	1.402
Ic	-9.376	0.144	0.443	0.657	0.898	-0.372	0.815	1.555
Id	-9.311	0.127	0.508	0.722	0.915	-0.389	0.897	1.637
IIa	-10.164	0.095	1.336	0.664	1.958	0.272	1.608	2.622
IIb	-10.095	0.083	0.456	0.733	1.355	0.260	0.716	2.088
IIc	-9.763	-0.123	0.056	1.065	1.165	0.054	0.110	2.230
IId	-9.727	-0.156	0.092	1.101	1.198	0.021	0.113	2.299
IIIa	-9.313	-0.363	2.187	1.536	2.416	0.370	2.557	3.952
IIIb	-9.197	-0.345	1.354	1.652	1.783	0.388	1.742	3.435
IIIc	-9.308	-0.364	0.511	1.541	1.406	0.369	0.880	2.947
IIId	-9.068	-0.396	0.751	1.781	1.438	0.337	1.088	3.219

I = TS with 1,3,4-triazole; II = TS with 1,3,4-oxadiazole; III = 1,3,4-thiadiazole; a = for addition of acetylene; b = for addition of ethylene; c = for *exo* addition of cyclopropene; d = for *endo* addition of cyclopropene; A = $\text{HOMO}_{\text{TS}} - \text{HOMO}_{\text{dienophile}}$; B = $\text{HOMO}_{\text{TS}} - \text{HOMO}_{\text{heterocycle}}$; C = $\text{LUMO}_{\text{dienophile}} - \text{LUMO}_{\text{TS}}$; D = $\text{LUMO}_{\text{TS}} - \text{LUMO}_{\text{heterocycle}}$; $\Sigma_1 = |A| + |D|$; $\Sigma_2 = |B| + |C|$.

Because these transition state structures are symmetric, all previously used approaches for determining relative reactivity of the heterocycles should be applicable in these cases. The cycloaddition is a HOMO dienophile and LUMO heterocycle (diene) controlled cycloaddition reaction with exceptionally low demand for orbital energy changes in reactants to achieve the electronic contribution present in the transition state structure (Table 41). There is no doubt that 1,3,4-oxadiazole is the most reactive of all five-membered heterocycles with three heteroatoms. However, the question remains as to whether this heterocycle is more reactive than, for instance, furan or even cyclopentadiene. To answer this question, the deviation of bond order uniformity in the six-membered ring being formed was computed (Table 42). The bond order uniformity selected 1,3,4-

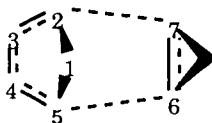


Table 42. Bond order uniformity for six-membered transition state structure of an *exo* cyclopropane addition to heterocycle with two heteroatoms

Diene	BOC ₂₃	BOC ₃₄	BOC ₄₅	BOC ₅₆	BOC ₆₇	BOC ₂₇	SBOD
cyclopentadiene	1.433	1.390	1.433	0.362	1.482	0.362	0.234
furan	1.286	1.530	1.286	0.378	1.471	0.378	0.449
1,3,4-triazole	1.169	1.692	1.169	0.454	1.382	0.454	0.746
1,3,4-oxadiazole	1.229	1.614	1.229	0.411	1.436	0.411	0.638
1,3,4-thiadiazole	1.196	1.677	1.196	0.443	1.399	0.443	0.786

Average bond order deviation is computed from the formula: $6X + 4 = \text{sum of the ring's bond orders}$ as explained in Scheme 4; BOC_{X,Y} = bond order change for bonds between atoms X and Y in the heterocycle ring required to achieve transition state structures; SBOD = sum of the bond order deviations.

oxadiazole as the most reactive heterocycle with three heteroatoms; however, its transition state ring bond order deviation was substantially higher than cyclopentadiene and furan deviations. To determine reaction feasibility of 1,3,4-oxadiazole as a diene for the Diels-Alder reaction, the activation barriers in reaction with acetylene, ethylene, and cyclopropene were computed (Table 43).

The reaction barrier firmly supported our previous finding that 1,3,4-oxadiazole should react with highly reactive dienophiles. In fact, it was predicted that the reaction between 1,3,4-oxadiazole and cyclopropene should be possible under moderate reaction conditions (room temperature). For reactions with dienophiles of low reactivity such as ethylene, forceful reaction conditions or even activation of the diene or dienophile are required. Both 1,3,4-triazole and 1,3,4-thiadiazole were predicted to have activation barriers that were ~ 6 kcal/mol and correspond to comparable reactivities. In all cycloaddition reactions with cyclopropene as a dienophile, an *exo* cycloadduct is predicted to be a major or exclusive product, which is in agreement with some of our previous studies of cycloaddition reactions with furan as a diene and cyclopropene as a dienophile.

Table 43. The AM1 and The B3LYP/6-31G(d) computed activation barriers for selected Diels-Alder cycloaddition reactions with heterocycles that have heteroatoms in 1, 3, and 4 position

Reaction	HOF	E	ΔE_I	ΔE_{II}
acetylene + 1,3,4-triazole	172.1	-319.500830	44.4	33.7
ethylene + 1,3,4-triazole	125.2	-320.765958	35.8	31.4
<i>exo</i> cyclopropene + 1,3,4-triazole	181.0	-358.810965	33.3	22.4
<i>endo</i> cyclopropene + 1,3,4-triazole	182.4	-358.807335	34.7	24.6
acetylene + 1,3,4-oxadiazole	125.3	-339.362477	37.6	28.6
ethylene + 1,3,4-oxadiazole	77.6	-340.630984	28.2	24.2
<i>exo</i> cyclopropene + 1,3,4-oxadiazole	134.3	-378.673571	26.6	16.7
<i>endo</i> cyclopropene + 1,3,4-oxadiazole	136.1	-378.669761	28.4	19.1
acetylene + 1,3,4-thiadiazole	164.8	-662.327826	48.0	36.3
ethylene + 1,3,4-thiadiazole	118.4	-663.597477	39.9	31.2
<i>exo</i> cyclopropene + 1,3,4-thiadiazole	174.5	-701.641214	37.7	22.9
<i>endo</i> cyclopropene + 1,3,4-thiadiazole	174.9	-701.638490	38.1	24.6

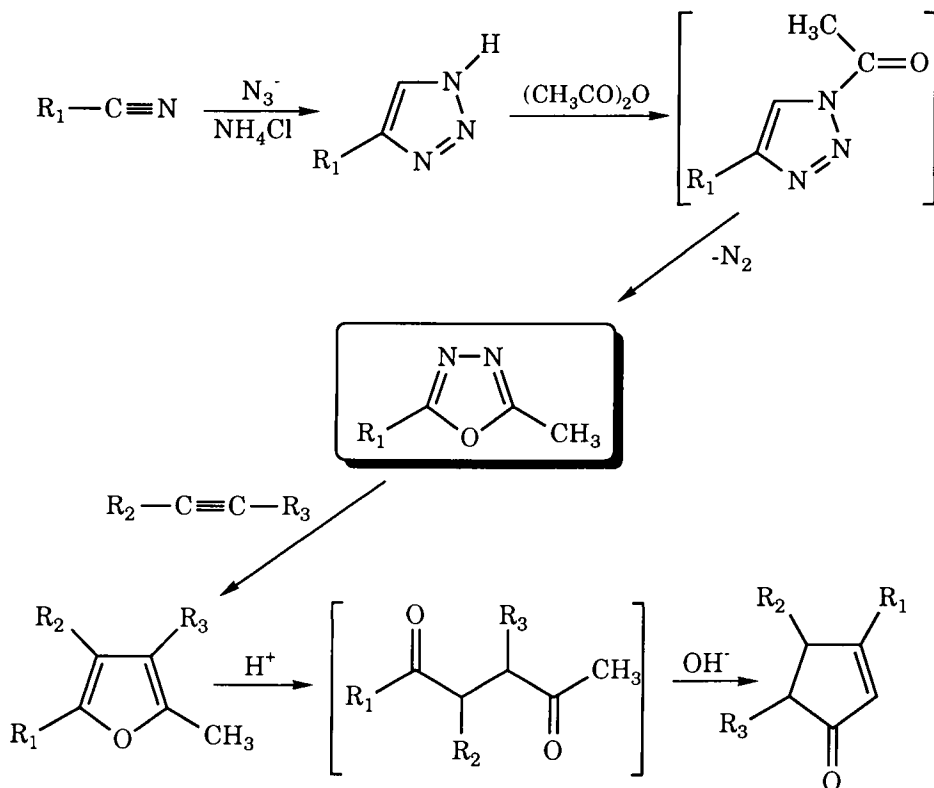
HOF = heat of formation computed by AM1; E = total energy (a.u.) computed by B3LYP/6-31G(d)/AM1; ΔE_I = activation barrier (kcal/mol) computed by AM1; ΔE_{II} = activation barrier (kcal/mol) computed by B3LYP/6-31G(d)/AM1.

5.4. Further investigation of the role of 1,3,4-oxadiazole as a diene in Diels-Alder reactions

1,3,4-Oxadiazole is a particularly interesting compound because it can be very easily synthesized from organic nitriles and acid chlorides or anhydrides through formation of tetrazoles [58]. By substitution of two nitrogens with alkyne derivatives, one can obtain many substituted furans that would otherwise, through ring opening and closing, be transferred into prostaglandins [59]. This is presented in Scheme 9. Therefore, exploring their capability of being dienes for Diels-Alder reactions is important because they are an essential part of accomplishing this chemical transformation.

As mentioned above, the cycloaddition reaction with 1,3,4-oxadiazole is predicted to be LUMO diene (heterocycle) controlled. That definitely suggests that with electron-withdrawing substituents in the two and five positions of the heterocycle ring, the heterocycle should become more reactive as a diene for Diels-Alder reactions. To study the usefulness of 1,3,4-oxadiazole and its derivatives as dienes for the Diels-Alder reaction, we are presenting the results of our theoretical study of the cyclopropene addition to 2,5-di(trifluoromethyl)-1,3,4-oxadiazole. The AM1 computed FMO energy gap for this reaction pair was only 8.00266 eV in comparison to 9.64149 eV FMO energy gap between LUMO of 1,3,4-oxadiazole and HOMO of cyclopropene. Therefore, the computed activation barrier for the cyclopropene addition to 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole should be very

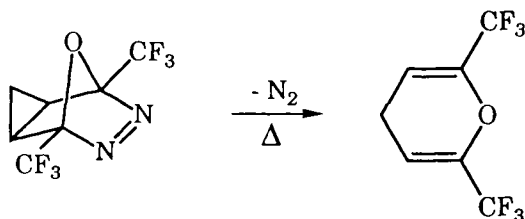
small. The AM1 computed activation barriers were 25.0 kcal/mol for *endo* and 23.5 kcal/mol for *exo* cycloaddition. These values were approximately 3 kcal/mol lower than what was predicted for unsubstituted 1,3,4-oxadiazole (Table 43). One might speculate that the B3LYP/6-31G(d)//AM1 computed activation barriers should be around 13 kcal/mol for *endo* and 10 kcal/mol for the *exo* cyclopropene addition to 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole. In fact, the computed energies (12.7 and 11.0 kcal/mol, respectively) were very close to this prediction suggesting that properly substituted 1,3,4-oxadiazoles should be very good dienes for Diels-Alder reaction.



Scheme 9. One of the possible ways of preparation of prostaglandin precursors from simple starting materials through 1,3,4-oxadiazole derivatives as key intermediates.

The question remains: will it be possible to isolate the cycloadduct between 1,3,4-oxadiazole and the dienophiles from the reaction mixture? Let us examine the case that is likely to be very easy experimentally; the addition of cyclopropene to 3,5-bis(trifluoromethyl)-1,3,4-oxadiazole. This cycloadduct actually represents a cyclic azo compound (Scheme 10). Azo compounds are known to decompose very easily producing nitrogen molecules and the corresponding radicals. The energy for the decomposition can be either thermal or photochemical [60]. In

thermal decomposition, it has been established that the temperature at which decomposition occurs depends on the nature of the substituents. Thus, azomethane decomposes to the methyl radical at a temperature above 400°C, while azo compounds derived from an allyl group decompose at around 100°C. Considering the fact the C-C bond of the cyclopropane ring of the cycloadduct (Scheme 10) has a π bond character and that the formed radical is tertiary with an ether bond, it is therefore more than likely that the transformation outlined in Scheme 3 should occur with a moderate activation barrier with a reaction temperature far below 100°C. If possible, this would be an excellent approach to synthesize cyclic vinyl ethers (γ -pyrans). To verify the credibility of the reaction scheme we have performed an AM study of the transformation is presented in Scheme 10.



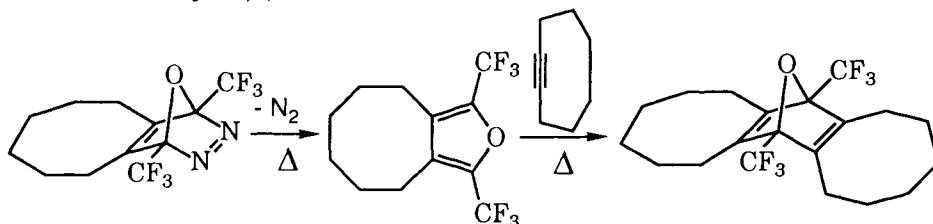
Scheme 10. Possible thermal decomposition of the cycloadduct between cyclopropene and 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole.

The AM1 estimated activation barrier for nitrogen elimination was 17.5 kcal/mol. This means that isolation of the cycloadduct between cyclopropene and 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole might be very hard. If we explore other cycloadducts between cyclopropene and 2 and 5 disubstituted 1,3,4-oxadiazoles and their activation barriers for nitrogen elimination, we can see that transition state structures are very similar with differences in C-N distance for the carbomethoxy substituent of 0.536 Å, and the methyl substituent of 0.605 Å instead of trifluoromethyl substitution. These results proved that all the transition state structures were very similar. That was also true for computed activation barriers. With carbomethoxy as a substituent, the activation barrier was 14.7 kcal/mol and with methyl as substituent, 16.7 kcal/mol. All of these values suggested that isolation of the cycloadduct might not be possible and that a substituted pyran should be expected as the final cycloaddition product.

One might anticipate that a similar effect as in the case of the strain cyclopropene exhibits as a dienophile would be present for the even less reactive acetylene. For instance, the frontier energy gap between LUMO of 1,3,4-oxadiazole and HOMO cyclooctyne was 9.81137 eV in comparison with 11.32244 eV for the acetylene addition to 1,3,4-oxadiazole. Other factor related to reactivity of acetylene in the cycloaddition reaction was the strain energy release going from a triple bond in the starting material to a double bond in the product. It can be estimated through the partial heat of hydrogenation of acetylene to ethylene and cyclooctyne to *cis*-cyclooctyne. The heat of hydrogenation of acetylene to

ethylene is -43.5 kcal/mol while the heat of hydrogenation of cyclooctyne to cyclooctene is -53.8 kcal/mol; 10.2 kcal/mol can be attributed to the strain energy release. This difference should bring the activation barrier for the acetylene addition (predicted to be 28.6 kcal/mol, Table 43) to around 20 kcal/mol. That value should be obtained under normal reaction conditions. The FMO orbital energy gap between LUMO of substituted 1,3,4-oxadiazoles and HOMO of cyclooctyne were 9.93129 eV for 3,5-dimethyl-1,3,4-oxadiazole, 9.81137 eV for 1,3,4-oxadiazole, and 8.17254 eV for 3,5-bis(trifluoromethyl)-1,3,4-oxadiazole as the dienes. The last FMO energy gap was slightly higher than the one obtained for the cyclopropene addition to 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole (8.00266 eV). One could then expect that the cyclooctyne addition to 1,3,4-oxadiazole might be experimentally achievable.

As determined on the basis of the FMO energies, the addition of cyclooctyne to substituted 1,3,4-oxadiazoles is easier than the addition of acetylene. Nevertheless, computed activation barriers (AM1 semiempirical method) were too high (33.6 kcal/mol for 2,5-dimethyl-1,3,4-oxadiazole and 27.7 kcal/mol for bis(trifluoromethyl)-1,3,4-oxadiazole). Considering the fact that the AM1 computational method tends to compute activation barriers that are almost the same for distinctively different reactive homologues of the same series and having established a correlation between AM1 computed and B3LYP computed energies, it is possible to estimate the activation barrier for the cyclooctyne addition to 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole. Our best estimate was around 20 kcal/mol.



Scheme 11. Conversion of the cycloadduct of cyclooctyne and 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole into a derivative of furan or even into a bicyclic adduct after the second addition of cyclooctyne

The accumulation of the cycloaddition product is related to its thermal stability in regard to nitrogen elimination. Here, elimination of nitrogen is even more pronounced because of two reasons: the presence of the double C-C bond instead of a cyclopropane moiety (Scheme 11) and because it can produce corresponding furan derivatives. Furan is actually one of the rare aromatic heterocyclic compounds that easily participates in Diels-Alder reactions as a moderately active diene. Therefore, it is also reasonable to postulate that the furan derivative obtained after elimination of nitrogen is more reactive than 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole. Additionally, the cycloadduct with a second molecule of cyclooctyne would be a final product of the cycloaddition reaction. To explore this possibility further, a semiempirical study of cycloadduct stability and activation barrier needed for cyclooctyne to react with furan was performed.

The AM1 computed activation barrier for the elimination of nitrogen from the cycloadduct between cyclooctyne and 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole was near 7.2 kcal/mol. The fact that the AM1 computational method overestimated activation barriers, the activation barrier for nitrogen elimination must be even smaller than 7.2 kcal/mol. In any case, computed activation energy for decomposition of this cycloadduct was substantially lower than it was for its formation. Therefore, we can safely state that trapping a bicyclic product would be very difficult and that the product of the reaction is definitively a derivative of furan as outlined in Scheme 11.

As discussed above, furans are five-membered heterocycles that readily undergo HOMO dienophile controlled (normal) Diels-Alder reactions. In our case, a tetra substituted furan has two electron donating and two strongly electron withdrawing substituents, therefore it might be less reactive than unsubstituted furan, or even activated 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole. For instance, the HOMO energy was -9.31685 eV and LUMO energy was 0.72282 eV for furan, while the same energies for the substituted furan as shown in Scheme 3 were -10.41022 eV and -0.83062 eV. With frontier orbital energies for cyclooctyne of -9.98847 eV and 1.44243 eV, both cycloadditions were actually LUMO diene controlled. The FMO energy gaps were 10.71129 eV and 9.15785 eV, tetra substituted furan being chosen as the more reactive one, although the FMO energy gap for the cyclooctyne addition to 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole (8.17254 eV) was substantially lower than for addition to the substituted furan. Even though comparison of the FMO gap for different aromatic systems should be taken with caution because the difference of aromatization is not fully incorporated in frontier orbital energies, the estimated difference favors the second addition by a very high margin. This should also be reflected in the difference of their reactivity. To confirm this observation, the activation barrier for the second cyclooctyne was computed. The AM1 estimated the barrier to be around 5 kcal/mol. This confirmed that the final product of the cyclooctyne addition to 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole was the bisadduct outlined in Scheme 11.

To make cycloaddition with 1,3,4-oxadiazole as a diene feasible, it is not necessary to use highly strained molecular systems such as cyclopropene and *cis*-cyclooctyne. Because the reaction is LUMO diene controlled, if the dienophile has an exceptionally high energy HOMO orbital, then the FMO energy gap between these two reactants should be very small and facilitate the cycloaddition reaction. That is the case for highly electron rich dienophiles such as aminoacetylene. Its HOMO energy was -9.48123 eV with the FMO energy gap relative to LUMO of 2,5-bis(trifluoromethyl)-1,3,4-oxadiazole of 7.6653 eV. This FMO energy gap selected aminoacetylene-2,5-bis(trifluoromethyl)-1,3,4-oxadiazole, a very reactive pair, for the Diels-Alder reaction.

All computational studies were in agreement with experimental findings. To the best of our knowledge, there are no experimental data that suggest the possibility of a cycloaddition reaction between 1,3,4-oxadiazole or electron rich 1,3,4-oxadiazoles with either ethylene or acetylene. Our attempt to perform a cycloaddition reaction between electron rich 1,3,4-oxadiazoles such as 2-methyl-5-

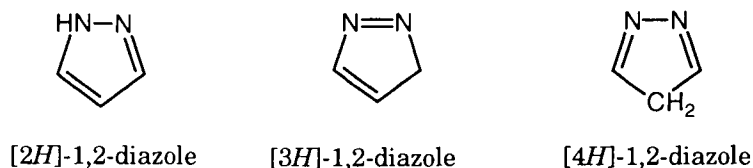
thiobutyl-1,3,4-oxadiazole and maleic anhydride or diethyl acetylenedicarboxylate was in vain. It has been demonstrated that the cycloaddition reactions with LUMO activated 1,3,4-oxadiazoles and exceptionally active dienophiles such as cyclopentene, *cis*-cyclooctyne, and *N,N*-dimethylaminoacetylene can occur. Although cycloadducts were formed, there was a problem with their isolation due to elimination of molecular nitrogen. In the case of the cyclopropene addition, derivatives of γ -pyrans were obtained, while in the case of cyclooctyne, a biscycloadduct was isolated as a final product.

6. CYCLOADDITION REACTIONS WITH ACTIVATED HETEROCYCLES THAT HAVE TWO OR THREE HETEROATOMS

As our computational results presented above demonstrate, it is highly unlikely that heterocycles would be good dienes for Diels-Alder reactions if formation of one or two C-N bonds were involved in the course of the reaction. This automatically eliminates some tautomeric forms of five-membered heterocycles with heteroatoms in 1 and 2 positions as well as five-membered heterocycles with heteroatoms in 1,2,3 and 1,2,5 positions. A major reason for the low reactivity of the heterocycles is because of their high aromaticity. It is obvious that diminishing or eliminating the aromaticity in these heterocycles would make them better dienophiles for Diels-Alder reactions.

6.1. Activation of 1,2-diazole as a diene for Diels-Alder reaction

There is no experimental evidence supporting the employment of *2H*-pyrazoles as dienes for the Diels-Alder reactions as far as we know. The report of their participation in a [4+2] cycloaddition has been shown to be incorrect [61]. There are three possible tautomers (Scheme 12). The first is aromatic while the two



Scheme 12. Structures of different tautomer structures of 1,2-diazoles

others are non-aromatic tautomers. It is reasonable to assume that [2H]-1,2-diazole is the most stable tautomer for 1,2-diazole. According to the AM1 calculations, the 1,2-diazole stability follows the order: [2H]-1,2-diazole (0.0 kcal/mol), [4H]-1,2-diazole (-6.1 kcal/mol), and [3H]-1,2-diazole (-7.2 kcal/mol). The FMO energy gap with ethylene as a dienophile chose [4H]-1,2-diazole as the most reactive. The reaction was predicted to be LUMO diene controlled.

A more reliable study of the 1,2-diazole tautomer reactivity as a diene for the Diels-Alder reaction was carried out by the MP2/6-31+G(d) theory model. The predicted order of reactivity was similar to the one generated by the AM1 study: [2H]-1,2-diazole (0.0 kcal/mol); [4H]-1,2-diazole (-24.3 kcal/mol), and [3H]-1,2-diazole (-29.1 kcal/mol). Because the reaction is LUMO diene controlled by

protonation of the [4*H*]-1,2-diazole, an even more reactive dienophile should be obtained. That was perfectly demonstrated by the computed activation energies for the ethylene addition to this tautomer (Table 44). The activation barrier for ethylene addition computed with B3LYP/6-31+G(d) theory level was very close to the one obtained with the more computationally expensive MP4/MP2 theory level. In all calculations [4*H*]-1,2-diazole was better diene for Diels-Alder reactions than

Table 44. Activation barriers (kcal/mol) for ethylene addition calculate by using the 6-31+G(d) basis set

Diene	RHF	B3LYP	BLYP	MP2	MP3/MP2	MP4/MP2
butadiene	47.0	24.9	23.1	18.5	27.6	23.1
[4 <i>H</i>]-1,2-diazole	40.9	20.7	20.0	10.7	22.1	16.1
protonated [4 <i>H</i>]-1,2-diazole	24.1	6.9	6.3	0.4	8.3	3.8

was, for instance, butadiene. On the other hand, a cycloaddition reaction with protonated [4*H*]-1,2-diazole should occur at a low temperature because activation barriers should be around 6 kcal/mol.

We have previously suggested that the major reason for the low reactivity is aromaticity of the heterocycle systems. By eliminating the cyclic conjugation, the heterocycles should become viable synthetic sources for Diels-Alder dienes. That was perfectly demonstrated with [4*H*]-1,2-diazole. Let us confirm the usefulness of the ring bond order uniformity as a measure of reactivity for these heterocycles. The deviation from the ring bond order uniformity of different 1,2-diazole tautomers perfectly follows the order of their relative stability. The most stable and therefore the ring most uniform bond order distribution was computed for [2*H*]-1,2-diazole while the least stable was [3*H*]-1,2-diazole with the highest bond order deviation (Table 45). Considering the fact that in the reaction with [3*H*]-1,2-diazole, one

Table 45. The ring's bond orders and bond order deviation for different tautomers of 1,2-diazole

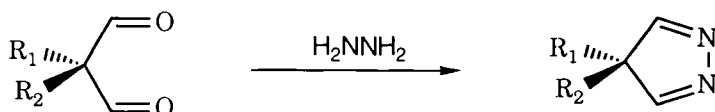
	BO ₁₂	BO ₂₃	BO ₃₄	BO ₄₅	BO ₅₁	SBO	ABO	BOD
cyclopentadiene	1.002	1.849	1.061	1.849	1.002	6.763	1.353	1.986
furan	1.104	1.670	1.190	1.670	1.104	6.738	1.348	1.290
[2 <i>H</i>]-1,2-diazole	1.172	1.195	1.559	1.262	1.580	6.768	1.354	0.864
[3 <i>H</i>]-1,2-diazole	1.927	0.967	0.995	1.854	1.014	6.757	1.351	2.156
[4 <i>H</i>]-1,2-diazole	1.097	1.832	0.984	0.984	1.832	6.729	1.346	1.945
H-[4 <i>H</i>]-1,2-diazole	0.998	1.658	1.026	0.987	1.839	6.507	1.302	1.788

BO_{X-Y} = bond order between atoms X and Y in the heterocycle ring; SBO = sum of ring bond orders; ABO = average bond order; BOD = sum of bond order deviation from the average ring bond order

energetically unfavorable C-N bond should be formed, the [4*H*]-diazole should be the most reactive as diene for the Diels-Alder reaction. The reaction was predicted to be LUMO diene controlled, and as a result, an acid catalyzed reaction.

Furthermore, protonation of [4*H*]-1,2-diazole should improve its reactivity. This was demonstrated by a considerably lower electron density on the aromatic ring of protonated (SBO = 5.507, Table 45) in comparison with unprotonated [4*H*]-1,2-diazole (SBO = 6.729, Table 45).

Of course, higher energy tautomers will not be engaged in cycloaddition reactions with dienophiles. Usually, it will be the lower energy tautomer that is responsible for the reactivity of the aromatic heterocycle. It is therefore necessary to lock the most reactive tautomer in its form and then the cycloaddition reaction cannot be conducted with this tautomer. The simplest organic synthetic way to prepare these tautomers is by adding two substituents in the 4 position of the 1,2-diazole ring. For example, Adam and coworkers [62] used a reaction between hydrazine and derivatives of malonic aldehydes to prepare a wide variety of 4,4-disubstituted[4*H*]-1,2-diazoles (Scheme 13). Here we are exploring activation barriers for the reaction with 4,4-dimethyl-[4*H*]-



4,4-disubstituted malonaldehyde

4,4-disubstituted-[4*H*]-1,2-diazoleScheme 13. Preparation of "frozen" 4,4-disubstituted-[4*H*]-1,2-diazoles tautomer

1,2-diazoles. The aromatic character of the 1,2-diazole ring is totally eliminated so one might expect very low activation barriers for the Diels-Alder reactions with such a diene. Before presenting our results for the estimation of its activation energy, let us first examine the two isomeric transition state structures for the cyclopropene addition to 4,4-dimethyl-[4*H*]-1,2-diazole. There are two possible isomeric transition state structures, for an *exo* and an *endo* cyclopropene addition (Figure 11). The transition state structures are for the synchronous formation of

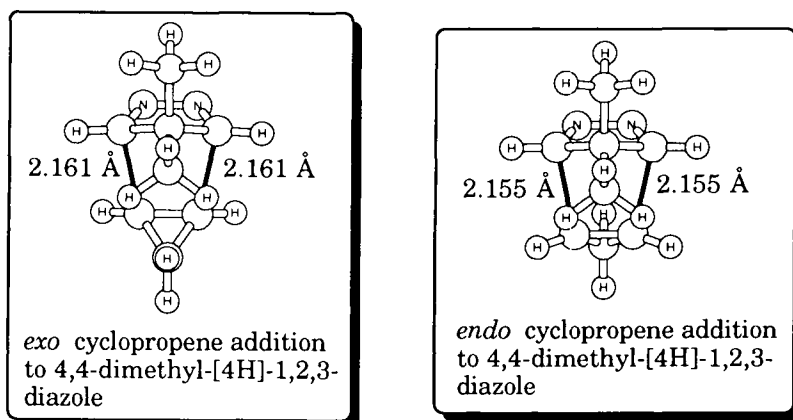


Figure 11. The AM1 computed transition state structures cyclopropene addition to 4,4-dimethyl-[4*H*]-1,3,4-oxadiazole.

both CC bonds. One can expect considerable repulsion interactions between the methyl group of the 4,4-dimethyl-[4*H*]-1,2-diazole moiety and the methylene group of the cyclopropane moiety in the *exo* transition streetcars. This results in a slightly longer CC bond in formation. In many of our previous cycloaddition reactions, we have found formation of the *exo* cycloadduct to be dominant because of secondary orbital interactions that help to stabilize the *exo* transition state structures. In this case, only H-*p* secondary orbital interactions exist in the *endo* transition state structure. One might expect that the *endo* transition state structure might have a substantially lower activation barrier.

The interactions in two isomeric transition state structures must be also responsible for change of the aromaticity of the six-membered ring of two isomeric transition state structures (Table 46). Both transition state structures have exceptionally low deviation from an ideal distribution of bond order in an six-membered transition state structure. The transition state structure with this heterocycle is more aromatic than the corresponding transition state structure for furan as diene. It is actually comparable in reactivity to cyclopentadiene, although for the *endo* transition state structure it was predicted to be the most aromatic of all studied transition states here (Table 46). Therefore should have the lowest activation barrier.

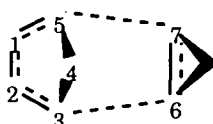


Table 46. Bond order uniformity for six-membered transition state structure of an *exo* and *endo* cyclopropene addition to 4,4-dimethyl-[4*H*]-1,2-diazole computed with the AM1 method.

transition state	BOC ₁₂	BOC ₂₃	BOC ₃₆	BOC ₆₇	BOC ₅₇	BOC ₁₅	SBOD
<i>exo</i> cyclopentadiene	1.390	1.433	0.362	1.482	0.362	1.433	0.234
<i>endo</i> cyclopentadiene	1.382	1.436	0.372	1.472	0.372	1.436	0.213
<i>exo</i> furan	1.530	1.286	0.378	1.471	0.378	1.286	0.449
<i>exo</i>	1.489	1.361	0.405	1.445	0.405	1.361	0.225
<i>endo</i>	1.446	1.402	0.379	1.473	0.379	1.402	0.184

Average bond order deviation is computed from formula $6X + 4 = \text{sum of the ring's bond orders}$ as explained in Scheme 4; BOC_{X.Y} = bond order change for bonds between atoms X and Y in the heterocycle ring required to achieve transition state structures; SBOD = sum of the bond order deviations from an average TS ring bond order

To confirm this assumption, we have computed activation barriers for acetylene, ethylene and cyclopropene addition to 4,4-dimethyl-[4*H*]-1,2-diazole. To our delight, the B3LYP/6-31G(d)/AM1 computed activation barrier for ethylene addition to 4,4-dimethyl-[4*H*]-1,2-diazole is almost identical (Table 47) to the value obtained with full B3LYP/6-31+G(d) calculation on [4*H*]-1,2-diazole as dienophile (Table 44). The activation barrier for the acetylene addition is 22.3 kcal/mol indicating that this reaction should be also experimentally feasible. As indicated

on the basis of the bond order uniformity of the transition state structures (Table 46) the *endo* transition state structure should have much lower activation than the isomeric *exo* transition state structures. That was confirmed by computing the activation barrier for the *endo* addition of cyclopropene to 4,4-dimethyl-[4*H*]-1,2-diazole to be only 16.4 kcal/mol. The energy difference between *exo* and *endo* transition state structures is 8 kcal/mol indication that formation of the *exo* cycloadduct should be excluded. In this way our computational studies suggest that 4,4-disubstituted-[4*H*]-diazoles are exceptionally good dienes for the Diels-Alder reactions with wide variety of dienophiles what was confirmed also experimentally [62].

Table 47. Activation barriers for ethylene, acetylene, and cyclopropene addition to 4,4-dimethyl-[4*H*]-1,2-diazole

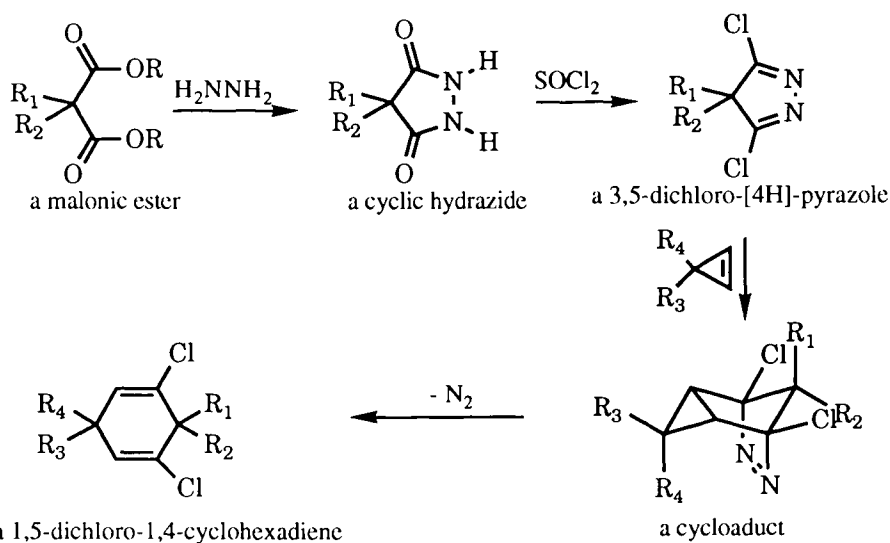
added dienophile (TS)	HOF	E	ΔE_I	ΔE_{II}
acetylene	156.3	-382.070459	38.5	22.3
ethylene	111.1	-383.336753	31.6	19.4
<i>exo</i> cyclopropene	178.2	-421.359413	40.4	24.4
<i>endo</i> cyclopropene	169.1	-421.371941	31.3	16.4

HOF = heat of formation computed by AM1; E = total energy (a.u.) computed with B3LYP/6-31G(d)/AM1; ΔE_I = activation barrier (kcal/mol) computed by AM1; ΔE_{II} = activation barrier (kcal/mol) computed with B3LYP/6-31G(d)/AM1.

6.2. Transformation of cyclic malonohydrazides into the Diels-Alder reactive 1,3-diazole

From a synthetic chemist's point of view, it is not sufficient to demonstrate that "frozen" anti-aromatic tautomer of 1,2-pyrazole can be a good diene for the Diels-Alder reaction. It is of equal importance to illustrate that we can prepare functionalized 1,2-diazole, which can be used as starting materials for the preparation of many different classes of organic compounds. For instance, it is possible to synthesize vinyl chloride as a valuable organic synthetic intermediate by taking advantage of the chemical transformations outlined in Scheme 14.

The reaction transformation should start by reacting malonic ester and hydrazine to prepare cyclic malonic hydrazide. It is well known that thionyl chloride can transfer amides to vinyl chloride, but in our case we used the pyrazole tautomer. We believe that this tautomer must be a sufficiently reactive diene for a Diels-Alder cycloaddition reaction because of the findings of one experimental study performed by Adam and coworkers [62] and one computational study presented above. There are at least two key factors that must be addressed before this reaction scheme can be considered seriously: the likelihood that 3,5-dichloro-[4*H*]-pyrazole will participate in a cycloaddition reaction and the stability of the formed cycloadduct. One can also question the stability and the availability of cyclic hydrazides of malonic ester. These compounds are unstable but they can



Scheme 14. The proposed mechanism for the transformation of cyclic malonic hydrazide into substituted 1,5-dichloro-1,4-cyclohexadiene.

be easily synthesized from their starting materials [63]. We hope that with this study the following questions will be answered: will AM1 and B3LYP/AM1 select this scheme as synthetically viable and what is the difference in energy stability between malonic cyclic hydrazide and its structural isomer, hydantoin. The last question is easily answered and even the AM1 semiempirical method, without any energy correction, gave a reliable energy. AM1 computed the heat of formation for cyclic malonic hydrazide to be -31.3 kcal/mol while for its isomer, hydantoin, it was predicted that heat of formation was -56.6 kcal/mol. Therefore hydantoin was by 25.3 kcal/mol more stable than our desirable starting material, cyclic malonic hydrazide. However, this energy difference has to be taken with caution because it was computed with the AM1 semiempirical method. The B3LYP/6-31G(d) energies also preferred hydantoin by 32.3 kcal/mol.

Table 48. The frontier molecular orbitals and frontier molecular orbital energy changes (eV) for the cyclopropene addition to cyclopentadiene and to 3,5-dichloro-[4H]-1,2-diazole

diene in the TS	HOMO	LUMO	ΔE_I	ΔE_{II}	SDE
<i>exo</i> cyclopentadiene	-8.418	0.434	0.661	-0.608	1.269
<i>endo</i> cyclopentadiene	-8.632	0.622	0.447	-0.420	0.867
<i>exo</i> 3,5-dichloro-[4H]-1,2-diazole	-9.526	-0.914	0.292	0.147	0.438
<i>endo</i> 3,5-dichloro-[4H]-1,2-diazole	-9.622	-0.882	0.196	0.179	0.375

$\Delta E_I = \text{HOMO}_{\text{TS}} - \text{HOMO}_{\text{cyclopentadiene}}$; $\Delta E_{II} = \text{LUMO}_{\text{TS}} - \text{LUMO}_{\text{diene}}$; $\text{SDE} = \Delta E_I + \Delta E_{II}$.

As in the case of unsubstituted [4*H*]-1,2-diazole, 3,5-dichloro-[4*H*]-1,2-diazole is a LUMO diene controlled cycloaddition reaction. Necessary FMO changes for reactants in the *exo* and *endo* addition of cyclopropene to 3,5-dichloro-[4*H*]-1,2-diazole to adapt the transition state structures are presented in Table 48. The required energy is too small to be compared with the other heterocycles presented in this study (for example see Table 41). Furthermore, the *endo* addition of cyclopropene should be favored over the *exo* cycloaddition.

To confirm these findings, we have computed activation barriers for the acetylene, ethylene and cyclopropene additions to 3,5-dichloro-[4*H*]-1,2-diazole. The computed activation barriers (Table 49) were even slightly lower than the activation barriers for same reaction with 4,4-dimethyl-[4*H*]-1,2-diazole as a diene (Table 47). This is a reasonable observation because the cycloaddition reaction is LUMO heterocycle (diene) controlled. The LUMO energy of 3,5-dichloro-[4*H*]-1,2-diazole was substantially lower than the LUMO energy of 4,4-dimethyl-[4*H*]-1,2-diazole. Subsequently, with a modest activation barrier of 20.1 kcal/mol, even poor dienophiles such as acetylene should be capable to react with 3,5-dichloro-[4*H*]-1,2-diazole as a diene in the Diels-Alder reactions (Table 49).

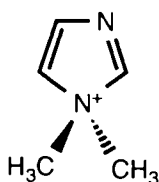
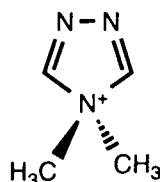
Table 49. Computed activation barriers (kcal/mol) of the Diels-Alder reactions of cyclopentadiene and 3,5-dichloro-[4*H*]-pyrazole as dienes

Transition state structure	ΔE_I	ΔE_{II}
3,5-dichloro-[4 <i>H</i>]-pyrazole + acetylene	35.2	20.1
3,5-dichloro-[4 <i>H</i>]-pyrazole + ethylene	27.3	18.8
<i>exo</i> 3,5-dichloro-[4 <i>H</i>]-pyrazole + cyclopropene	29.6	14.3
<i>endo</i> 3,5-dichloro-[4 <i>H</i>]-pyrazole + cyclopropene	27.4	12.3

ΔE_I = activation barrier computed by AM1; ΔE_{II} = activation barrier computed with the B3LYP/6-31G(d)//AM1 theory model.

6. 3. Quaternization of nitrogen atom as a way to activate 1,3-diazole, and 1,3,4-triazole as a diene for the Diels-Alder reaction

As we have demonstrated with the example of [4*H*]-1,2-diazole as a diene for the Diels-Alder addition, there are two critical prerequisites for any five-membered heterocycle to become a diene for the Diels alder reaction: the aromaticity of the heterocycle should be diminished as much as possible and the formation of two CC bonds is preferable over formation of a C-heteroatom bond. Because these reactions are LUMO diene controlled, it is helpful if electron-withdrawing substituents are attached to the heterocyclic ring, all of which may be accomplished if the nitrogen in position one of 1,2-diazole and 1,3,4-triazole is quaternized. We expect that these heterocycles should be exceptionally good dienes that promote the formation of the *endo* cycloadduct due to strong steric repulsion interactions between methyl groups of quaternized heterocycles and the dienophiles. To reinforce the existence of a high localization of the double bonds in

*N,N*-dimethyl-1,3-diazolium cation*N,N*-dimethyl-1,3,4-triazolium cation

Scheme 15. Structure *N,N*-dimethyl-1,3-diazolium and *N,N*-dimethyl-1,3,4-triazolium cations used as dienes for the Diels Alder reactions.

the heterocycle rings, we have computed the bond order deviation from uniformity for these two systems (Table 50). These calculations strongly favor both quaternized heterocycles as more reactive dienes over both furan and cyclopentadiene because of strong localization of single and double bonds. Both quaternized heterocycles are also favored because of the very high ring bond order deviation from uniformity.

Table 50. The ring bond orders and bond order deviation for cyclopentadiene, furan; two quaternized heterocycles computed with the AM1 semiempirical method

	BO ₁₂	BO ₂₃	BO ₃₄	BO ₄₅	BO ₅₁	SBO	ABO	BOD
A	1.002	1.849	1.061	1.849	1.002	6.763	1.353	1.986
B	1.104	1.670	1.190	1.670	1.104	6.738	1.348	1.290
C	0.835	1.918	1.038	1.838	0.915	6.544	1.309	2.277
D	0.853	1.890	1.052	1.890	0.853	6.530	1.308	2.330

A = cyclopentadiene; B = furan; C = 1,1-dimethyl-1,3-diazolium cation; D = 1,1-dimethyl-1,3,4-triazolium cation; BO_{X.Y} = bond order between atoms X and Y in the heterocyclic ring; SBO = sum of ring bond orders; ABO = average bond order; BOD = sum of bond order deviation from the average ring bond order.

Certainly a better insight into the reactivity of these two quaternized heterocycles can be obtained through evaluation of the activation barriers for cycloaddition (Table 51). It must be mentioned that transition state structures with these two quaternized heterocycles in reaction with acetylene, ethylene, and cyclopentene are very similar to the transition state structure with unquaternized heterocycles. As predicted on the basis of the bond order distribution in the heterocyclic ring, these two heterocycles were very reactive dienes for the Diels-Alder reactions. The computed energies were in full agreement with the bond order uniformity presented in Table 50. Both heterocycles had low π -density on the ring, but the *N,N*-dimethyl-1,3,4-triazolium cation was more electron deficient (SBO = 6.530, Table 50), with a higher π -orbital localization (BOD = 2.330, Table 50). Because the reaction is LUMO diene controlled, the activation barriers with the *N,N*-dimethyl-1,3,4-triazolium cation must be lower in comparison to activation

barriers with the *N,N*-dimethyl-1,2-diazolium cation. On the other hand, *endo* cyclopropene addition should be preferable due to a strong steric repulsion interaction present in the *exo* transition state structure.

All the qualitative evaluations of the Diels-Alder reactions with these two dienes are perfectly incorporated in the evaluated activation barriers. Even dienophiles with a low reactivity such as acetylene should add easily to both *N,N*-dimethyl-1,2-diazolium and *N,N*-dimethyl-1,3,4-triazolium cations. Later it becomes much more reactive. The *exo* cycloaddition had an activation barrier that was 9-10 kcal/mol higher, indicating that the *endo* cycloadduct should be the sole product of the reaction. In this way, we have demonstrated that cycloaddition reactions with quaternized heterocycles should be a very powerful method for utilizing heterocycles in organic syntheses.

Table 51. The AM1 and the B3LYP/6-31G(d) computed activation barriers for the acetylene, ethylene, and cyclopropene additions to *N,N*-dimethyl-1,3-diazolium and *N,N*-dimethyl-1,3,4-triazolium cations

reactant or reaction	HOF	E	ΔE_I	ΔE_{II}
Ia	328.4	-382.437697	33.7	16.0
Ib	282.5	-383.699914	26.1	15.6
Ic	347.6	-421.724309	32.9	19.4
Id	340.3	-421.744676	25.6	6.7
IIa	359.8	-398.439865	31.8	12.9
IIb	313.3	-399.704836	23.6	10.8
IIc	379.1	-437.728677	31.1	15.0
IId	370.8	-437.742736	22.8	6.2

I = TS with *N,N*-dimethyl-1,3-diazolium cation; II = TS with *N,N*-dimethyl-1,3,4-triazolium cation; a = for addition of acetylene; b = for addition of ethylene; c = for *exo* addition of cyclopropene; d = for *endo* addition of cyclopropene; HOF = heat of formation computed by AM1; E = total energy (a.u.) computed with B3LYP/6-31G(d)/AM1; ΔE_I = activation barrier (kcal/mol) computed by AM1; ΔE_{II} = activation barrier (kcal/mol) computed with B3LYP/6-31G(d)/AM1.

6. 4. Oxidation of a sulfur atom: a way to activate 1,3-thiazole and 1,3,4-thiadiazole as dienes for the Diels-Alder reaction

Elimination of the ring's π -orbital delocalization in five-membered heterocycles is a most efficient chemical manipulation used to synthesize highly reactive dienes for Diels-Alder reactions from heterocycles that would otherwise not readily participate in the Diels-Alder reaction. One way to accomplish this goal using heterocycles that contain sulfur atoms is through their oxidation to sulfone derivatives. We will demonstrate the usefulness of this approach by studying the reactivity of 1,3-thiazole 1,1-dioxide and 1,3,4-thiadiazole 1,1,-dioxide as dienes for Diels-Alder reactions with acetylene, ethylene, and cyclopropene.



Scheme 16. Structures of two oxidized sulfur containing heterocycles

Bond order deviations from an average ring bond order for both oxides were higher than for both furan and cyclobutadiene, denoting that both oxidized heterocycles presented in Scheme 16 are excellent dienes for the Diels-Alder reaction (Table 52). The reaction is a LUMO heterocycle (diene) controlled cycloaddition reaction; therefore, it is expected that by oxidizing 1,3-thiazole and 1,3,4-thiadiazole, not only do we localize the π -molecular orbitals, but the π -electron density on the heterocycle ring is also decreased. That was perfectly illustrated by the sum of the ring bond orders (Table 52). It was 6.738 for furan but only 5.978 and 6.017 for 1,3-thiazole and 1,3,4-thiadiazole 1,1-oxides respectively. That indicated that oxidized heterocycles might be even more reactive than the quaternized one (Table 51). It is interesting to point out that the bond orders for C-S bonds were exceptionally low, showing that the cycloadduct with some of the dienophiles might survive and could actually be transferred to other heterocycles. A slightly higher deviation of bond order uniformity for 1,3,4-thiadiazole 1,1-dioxide in comparison with 1,3-thiazole 1,1-dioxide suggested that the former should be more reactive in the Diels-Alder reaction.

Table 52. The ring bond orders and bond order deviation for cyclopentadiene, furan and the two quaternized heterocycles computed by the AM1 semiempirical method

	BO ₁₂	BO ₂₃	BO ₃₄	BO ₄₅	BO ₅₁	SBO	ABO	BOD
A	1.002	1.849	1.061	1.849	1.002	6.763	1.353	1.986
B	1.104	1.670	1.190	1.670	1.104	6.738	1.348	1.290
C	0.560	1.920	1.026	1.844	0.628	5.978	1.197	2.746
D	0.582	1.907	1.039	1.907	0.582	6.017	1.203	2.807

A = cyclopentadiene; B = furan; C = 1,3-thiazole 1,1-dioxide; D = 1,3,4-thiadiazole 1,1-dioxide; BO_{X.Y} = bond order between atoms X and Y in the heterocycle ring; SBO = sum of ring bond orders; ABO = average bond order; BOD = sum of bond order deviation from the average ring's bond order

All of these observations, concluded on the basis of the bond order distribution in the heterocycles, were confirmed by the computed activation barriers. The values were extremely low, suggesting that cycloaddition with almost any carbon dienophile should be experimentally feasible. In general, 1,3,4-thiadiazole 1,1-oxide is more reactive than 1,3-thiazole 1,1-dioxide. The computed activation barriers

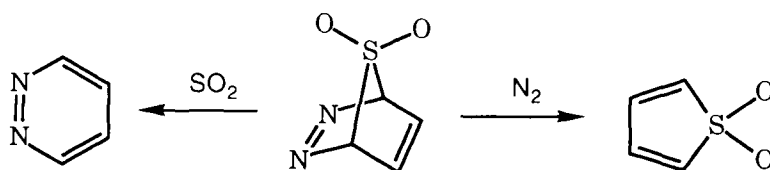
certainly suggest that the reaction should be carried out at room or even lower temperature (Table 53).

One may justifiably speculate that the produced cycloadducts in this reaction will not be sufficiently stable but might further engage in retro-cycloaddition. Let us, for instance, take into account the cycloadduct between 1,3,4-thiadiazole 1,1-dioxide and acetylene (Scheme 17). This cycloadduct can be decomposed by either eliminating nitrogen or sulfur dioxide, therefore producing new heterocyclic compounds. The B3LYP/6-31G(d)/AM1 computed an activation barrier for the elimination of nitrogen to be 17.9 kcal/mol, higher than the original activation barrier for the acetylene addition to 1,3,4-thiadiazole 1,1-dioxide. The other elimination of SO₂ from the cycloadduct transition state structure was not possible to locate. The reaction should occur spontaneously. This is not surprising because, by the elimination of SO₂, a new six-membered aromatic ring is formed (Scheme 17).

Table 53. The AM1 and the B3LYP/6-31G(d) computed activation barriers for the acetylene, ethylene, and cyclopropene additions to 1,3-thiazole and 1,3,4-thiadiazole 1,1-oxides

reactant or reaction	HOF	E	ΔE_I	ΔE_{II}
Ia	95.0	-796.653557	33.9	15.7
Ib	47.8	-797.920524	25.0	12.3
Ic	112.5	-835.951375	31.4	12.1
Id	106.5	-835.963172	25.4	4.7
IIa	121.4	-812.661377	30.3	15.5
IIb	73.8	-813.930271	21.0	10.9
IIc	139.8	-851.960880	28.7	10.9
IId	133.3	-851.970594	22.2	4.8

I = TS with 1,3-thiazole 1,1-dioxide; II = TS with 1,3,4-thiadiazole 1,1-dioxide; a = for addition of acetylene; b = for addition of ethylene; c = for *exo* addition of cyclopropene; d = for *endo* addition of cyclopropene; HOF = heat of formation computed by AM1; E = total energy (a.u.) computed with B3LYP/6-31G(d)/AM1; ΔE_I = activation barrier (kcal/mol) computed by AM1; ΔE_{II} = activation barrier (kcal/mol) computed with B3LYP/6-31G(d)/AM1.



Scheme 17. Two possible ways of decomposition of the cycloadduct between acetylene and 1,3,4-thiadiazole 1,1-dioxide.

7. CONCLUSION

There are many computational studies in this field generated by others, as well as us, that are not mentioned in this chapter. Our target was to demonstrate that simple theoretical approaches with modest computational requirements can give valuable results for experimentalists to use in planning further experiments. Four qualitative approaches and one quantitative approach were used to determine the reactivity of various five-membered heterocycles as dienes for the Diels-Alder reaction. Three dienophiles: acetylene, ethylene, and cyclopropene were selected because they perfectly represented extremely low reactivity (acetylene), low reactivity (ethylene), and high reactivity (cyclopropene) which can produce two isomeric (*exo* and *endo*) cycloadducts. We have demonstrated that a ring bond order uniformity is an excellent approach to determine the reactivity of heterocyclic compounds. One other approach, involving examination of isolated heterocycles, allows one to determine their relative aromaticity. The more uniform an aromatic system is, the less eager it will be to participate in cycloaddition reactions. Another approach is based on the principle that the more aromatic the transition state is, the lower the reaction barriers will be. Again, the aromaticity is judged through bond order uniformity of the Diels-Alder six-membered transition state ring. The last qualitative approach used here is Frontier Molecular Orbital (FMO) energy changes required for reactants to adapt electronic and conformational changes in order to become a part of the transition state structure. This approach has been shown to be more reliable than the traditional FMO gap between separated reactants.

In all our computational studies, we have established that these qualitative theoretical approaches, in combination with activation energies computed by B3LYP/6-31G(d) on AM1 geometries, are useful for evaluating the reactivity of aromatic five-membered heterocyclic compounds. In a few instances, we have demonstrated that B3LYP/6-31G(d) energies on AM1 geometries are just as reliable as the B3LYP/6-31G(d) energies on B3LYP geometries. Therefore, we suggest to use this approach for evaluation of the reactivity of large molecular systems. All computed results are in agreement with available experimental values or observations.

It has been proven that only five heterocyclic compounds can react as dienes for Diels-Alder reactions. The vast majority are unreactive. For a long time, experimental organic chemists have avoided their use in preparation of complex compounds such as natural products. Fortunately, with small chemical modifications, such as the introduction of electron-withdrawing substituents, methylation, or oxidation, otherwise unreactive heterocycles can become exceptionally good dienes for the Diels-Alder reaction. It is our hope that synthetic organic chemists will use the computational tools presented in this chapter to modify heterocycles to the required reactivity. In doing so, the desired organic transformations can be accomplished, which should, in turn, produce valuable organic materials.

REFERENCES

1. MOPAC version 6.0. Quantum Chemistry Program Exchange (QCPE), Program No. 455, 1990.
2. B. S. Jursic and Zdravkovski, *J. Mol. Struct. (Theochem)*, 333 (1994) 177.
3. A. D. Becke, *J. Chem. Phys.*, 98 (1993) 5648; C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 37 (1988) 785.
4. M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, *J. Chem. Phys.*, 77 (1982) 77.
5. Gaussian 94, Revision B.3, M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1995.
6. Z. Chen and M. L. Trudell, *Chem. Rev.*, 96 (1996) 1179 and references therein; R. R. Schidt, *Acc. Chem. Res.*, 19 (1986) 250; W. Adam and O. Delucchi, *Angew. Chem. Int. Ed. Engl.*, 19 (1980) 762; J. A. Mayoral and E. Pires, *J. Org. Chem.*, 26 (1996) 9479 and references therein
7. For general discussion of Diels-Alder reactions see: J. Sauer and R. Sustmann, *Angew. Chem. Int. Ed. Engl.*, 19 (1980) 779 and references therein.
8. For example see: G. Piancatelli, M. D'Auria, and F. D'Onofrio, *Synthesis*, (1994) 867.
9. For a review see: M. G. Block, *Reactions of Organosulfur Compounds*, Academic Press: New York, 1978; B. M. Trost and T. N. Salzman, *J. Org. Chem.*, 40 (1975) 40, 148, and references therein.
10. E. J. Corey and M. C. Desai, *Tetrahedron Lett.*, 26 (1985) 5747; L. D. Quin, J. Leimert, E. D. Middlemas, R. W. Miller, and A. T. McPhail, *J. Org. Chem.*, 44 (1979) 3496; K. Mori, *Tetrahedron*, 33 (1977) 289.
11. M. J. Glukhovtsev, *J. Chem. Ed.*, 74 (1997) 132; P. J. Garrat, *Aromaticity*; J. Wiley, New York, NY, 1986; D. Lewis, D. Peters, *Facts and Theory of Aromaticity*; Macmillan: London, 1975; A. T. Balaban, M. Bancin, and V. Ciorba, *Annulenes, Benzo-, Hetero-, Homo-Derivatives and Their Valence Isomers*, CRC, Boca Raton, FL, 1987; Vol. 1.
12. A. Streitwieser, Jr., *Molecular Orbital Theory for Organic Chemists*; J. Wiley, New York, NY, 1961; F. A. Carey and R. J. Sundberg, *Advanced Organic Chemistry, Part A: Structure and Mechanism*; 3rd ed., Plenum, New York, NY, 1990; N. Isaacs, *Physical Organic Chemistry*; 2nd ed., Longman, London, 1995.
13. L. Pauling, L. O. Brockway, and J. Y. Beach, *J. Am. Chem. Soc.*, 57 (1935) 2705; W. G. Penney, *Proc. Roy. Soc. A*, 158 (1937) 306; R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Oxford University Press, Oxford, 1990.
14. P. v. R. Schleyer, P. K. Freeman, H. Jiao, and B. Goldfuss, *Angew. Chem. Int. Ed. Engl.*, 34 (1995) 337; B. S. Jursic, *J. Heterocyclic Chem.*, 33 (1996) 1079.

15. A. R. Katritzky and R. Taylor, *Adv. Heterocyclic Chem.*, 47 (1990) 87; A. R. Katritzky, M. Karelson, and N. Malhotra, *Heterocycles*, 32 (1991) 127; T. L. Gilchrist, *Heterocyclic Chemistry*, Pitman, Marshfield, MA, 1985; A. R. Katritzky, *Handbook of Heterocyclic Chemistry*, Pergamon Press, New York, NY, 1985.
16. For reactivity of *S*-methylthiophenium ion in Diels-Alder reactions see: B. S. Jursic, Z. Zdravkovski, and S. L. Whittenburg, *J. Chem. Soc. Perkin Trans 2*, (1996) 455.
17. For reactivity of thiophene 1-oxide and thiophene 1,1-dioxide in Diels-Alder reactions, see: B. S. Jursic, *J. Heterocyclic Chem.* 32 (1995) 1445, and references therein.
18. For an *ab initio* study of low reactivity of thiophene in Diels-Alder reactions see: B. S. Jursic, Z. Zdravkovski, and S. L. Whittenburg, *J. Phys. Org. Chem.* 8 (1995) 753 and references therein.
19. K. Fukui and H. Fujimoto, *Bull. Chem. Soc. Japan*, 40 (1967) 2018; K. Fukui and H. Fujimoto, *Bull. Chem. Soc. Japan*, 42 (1969) 3399; K. Fukui, *Fortschr. Chem. Forsch.*, 15 (1970) 1; K. Fukui, *Acc. Chem. Res.*, 4 (1971) 57; K. N. Houk, *Acc. Chem. Res.*, 8 (1971) 361; K. Fukui, *Angew. Chem. Int. Ed. Engl.*, 21 (1982) 801; I. Fleming, *Frontier Orbitals and Organic Chemical Reactions*, J. Wiley, London, 1976, Chap. 4.
20. R. B. Woodward and R. Hoffmann *The Conservation of Orbital Symmetry*, Verlag Chemie, Weinheim/Bergstr, 1970; A. P. Marchand and R. E. Lehr (eds.), *Pericyclic Reactions*, Vol 1 and 2, Academic Press, New York, NY, 1977.
21. How to obtain, verify, and optimize transition state structures with semiempirical, *ab initio* and density functional theory methods, see: B. S. Jursic, *Computing Transition State Structures with Density Functional Theory Methods in, Recent Developments and Applications of Modern Density Functional Theory*, J.S. Seminario (ed.), Elsevier, Amsterdam, 1996; B. S. Jursic, *J. Chem. Educ.*, submitted.
22. B. S. Jursic, *Can. J. Chem.* submitted; B. S. Jursic, *Tetrahedron*, submitted.
23. K. B. Wiberg and W. J. Barley, *J. Am. Chem. Soc.*, 82 (1960) 6375; R. Breslow and M. Oda, *J. Am. Chem. Soc.*, 94 (1972) 4787; M. Oda, and R. Breslow, *J. Pecoraro, Tetrahedron Lett.*, 13 (1972) 4419.
24. D. L. Boger and S. N. Weinreb, *Hetero Diels-Alder Methodology in Organic Synthesis*, Academic Press, New York, NY, 1987; A. R. Katritzky and C. W. Rees, *Comprehensive Heterocyclic Chemistry*, Pergamon Press, New York, NY, 1985.
25. B. S. Jursic, *Tetrahedron Lett.*, 38 (1997) 1305 and references therein.
26. L. Pauling, *The Chemical Bond*, Cornell University Press, Ithaca, New York, NY, 1967.
27. For transition state theory see: K. B. Wiberg, *Physical Organic Chemistry*, Wiley, New York, 1964; L. P. Hammett, *Physical Organic Chemistry*, 2nd ed., McGraw Hill, New York, NY, 1970; D. L. Bunker, *Acc. Chem. Res.*, 7 (1974) 195; F. K. Fong, *Acc. Chem. Res.*, 9 (1976) 433; B. S. Jursic, *Computing Transition State Structures with Density Functional Theory Methods*, in:

- Recent Developments and Applications of Modern Density Functional Theory, J. M. Seminario (ed.), Elsevier, Amsterdam, 1996, p. 709.
28. B. S. Jursic and Z. Zdravkovski, *J. Mol. Struct. (Theochem)*, 309 (1994) 249.
 29. B. S. Jursic, *Can. J. Chem.*, 74 (1996) 114 and references therein.
 30. B. S. Jursic, *J. Mol. Struct. (Theochem)*, 358 (1995) 139; B. S. Jursic, *J. Mol. Struct. (Theochem)*, 370 (1996) 85.
 31. B. S. Jursic and Z. Zdravkovski, *J. Chem. Soc. Perkin Trans. 2*, (1995) 1223.
 32. K. N. Houk, Y. Li, and J. E. Evanseck, *Angew. Chem. Int. Ed. Engl.*, 31 (1992) 682; J. E. Eksterowicz and K. N. Houk, *Chem. Rev.*, 93 (1993) 2439; K. N. Houk, J. Gonzalez, and Y. Li, *Acc. Chem. Res.*, 28 (1995) 81 and references therein.
 33. K. Torrsel, *Acta Chem. Scand., Ser. B*, 30 (1976) 353; B. Iddon and R. M. Scrowston, *Adv. Heterocyclic Chem.*, 11 (1970) 177; P. Grieco, *Synthesis*, (1975) 67; W. Grime, K. Pohl, J. Wortmann, and D. Frowein, *Liebigs Ann.*, (1996) 1905, and references therein.
 34. There are many computational studies with some experimental references included in many of our previous studies of this problem: B. S. Jursic, *J. Chem. Soc., Perkin Trans. 2*, (1995) 1217; B. S. Jursic and Z. Zdravkovski, *Z. J. Mol. Struct. (Theochem)*, 331 (1995) 215; B. S. Jursic and Z. Zdravkovski, *J. Mol. Struct. (Theochem)*, 331 (1995) 229; B. S. Jursic and Z. Zdravkovski, *J. Mol. Struct. (Theochem)*, 332 (1995) 39; B. S. Jursic and D. Coupe, *J. Heterocyclic Chem.*, 32 (1995) 483; B. S. Jursic and Z. Zdravkovski, *J. Mol. Struct. (Theochem)*, 337 (1995) 9; B. S. Jursic and Z. Zdravkovski, *J. Phys. Org. Chem.*, 7 (1994) 641; B. S. Jursic and Z. Zdravkovski, *J. Heterocyclic Chem.*, 31 (1994) 1429; B. S. Jursic, and Z. Zdravkovski, *J. Chem. Soc. Perkin Trans. 2*, (1994) 1877; B. S. Jursic and Z. Zdravkovski, *J. Org. Chem.*, 59 (1994) 3015, and references therein.
 35. J. D. Roberts and H. E. Simmons, Jr., L. A. Carlsmith, and C. W. Vaughan, *J. Am. Chem. Soc.*, 75 (1953) 3290; J. D. Roberts, D. A. Semenow, H. E. Simmons, Jr., and L. A. Carlsmith, *J. Am. Chem. Soc.*, 78 (1956) 601; J. D. Roberts, C. W. Vaughan, L. A. Carlsmith, and D. A. Semenow, *J. Am. Chem. Soc.*, 78 (1956) 611.
 36. G. Wittig and K. Niethammer, *Chem. Ber.*, 93 (1960) 944; G. Wittig, H. Härle, E. Knauss, and K. Niethammer, *Chem. Ber.* 93 (1960) 951.
 37. For instance see: T. Matsumoto, T. Hosoya, and K. Suzuki, *J. Am. Chem. Soc.*, 114 (1992) 3568; A. Cobas, A. Guitián, and L. Castedo, *J. Org. Chem.*, 58 (1993) 3113; A. Menzek, M. Krawiec, W. H. Watson, and M. Balci, *J. Org. Chem.*, 56 (1991) 6755; D. G. Batt, D. G. Jones, S. La Greca, *S. J. Org. Chem.*, 56 (1991) 6704.
 38. R. McCulloch, A. R. Rye, and D. Wege, *Tetrahedron Lett.*, (1969) 5231.
 39. T. F. Spand, H. M. Garraffo, M. W. Edwards, and J. W. Daly, *J. Am. Chem. Soc.*, 114 (1992) 3475.
 40. For instance see: Z. Chen and M. L. Trudell, *Chem. Rev.*, 96 (1996) 1179, and references therein.

41. K. H. Gluesenkamp, E. Jaehde, W. Drosdziok, and M. Rajewsky, Ger. Offen. DE 4,295,306 1993; Chem. Abstr., 120 (1994) 134539n; K. Akasaka, T. Kimura, M. Senaga, and Y. Machida, Jpn. Kokai Tokyo Koho JP 10,878 [95 10,878] 1933; Chem. Abstr., 122 (1994) 291257c.
42. B. S. Jursic and Z. Zdravkovski, J. Mol. Struct. (Theochem), 332 (1995) 39; B. S. Jursic and Z. Zdravkovski, J. Heterocyclic Chem., 31 (1994) 1429.
43. U. Pindur, G. Lutz, and C. Otto, Chem. Rev., 93 (1993) 741.
44. N. E. Heard and J. Turner, J. Org. Chem., 60 (1995) 4302.
45. O. Diels, K. Alder, H. Winckler, and E. Peterson, Justus Liebigs Ann. Chem. 498 (1932) 1; O. Diels, K. Alder, H. Winckler, and E. Peterson, Justus Liebigs Ann. Chem., 490 (1931) 267.
46. R. M. Acheson, A. R. Hand, and J. M. Vernon, Proc. Chem. Soc., (1961) 164; R. M. Acheson and J. M. Vernon, J. Chem. Soc., (1961) 457.
47. C. K. Lee, C. S. Hahn, and W. E. Noland, J. Org. Chem., 43 (1978) 3727.
48. R. Kitzing, R. Fuchs, M. Joyeux, and H. Prinzbach, Helv. Chim. Acta, 51 (1968) 888.
49. W. E. Noland and C. K. Lee, J. Org. Chem., 45 (1980) 4573, and references therein.
50. N. W. Gabel, J. Org. Chem., 27 (1962) 301.
51. L. F. Fieser and M. J. Haddadin, J. Am. Chem. Soc. 86 (1964) 2081; L. F. Fieser and M. J. Haddadin, Can. J. Chem., 43 (1965) 1599.
52. W. Friedrichsen, Adv. Heterocyc. Chem., 26 (1980) 135.
53. U. E. Wiersum and W. J. Mijs, J. Chem. Soc., Chem. Com., (1972) 347.
54. For example see: M. J. S. Dewar, The Molecular Orbital Theory of Organic Chemistry, McGraw-Hill, New York, NY, 1969; M. J. S. Dewar, Angew. Chem. Int. Ed. Engl., 10 (1971) 761, and references therein.
55. G. S. Hammond, J. Am. Chem. Soc., 77 (1955) 344; W. J. LeNoble, A. R. Miller, and A. D. Hamann, J. Org. Chem., 42 (1972) 338; A. R. Miller, J. Am. Chem. Soc., 100 (1978) 1984.
56. L. E. Saris and M. P. Cava, J. Am. Chem. Soc., 98 (1976) 868, and reference therein.
57. G. Ya. Kondrateva, Khim. Nauka Pro., 2 (1957) 666 [Chem. Abstr., 52 (1958) 6345]; G. Ya. Kondrateva, Izv. Akad. Nauk SSSR, Org. Khim. Nauk, (1959) 484; M. Ya. Karpeiskii and V. L. Florentev, Russ. Chem. Rev., 38 (1969) 540; B. S. Jursic and Z. Zdravkovski, Bull. Chem. Technol. Macedonia, 13 (1994) 55, and references therein.
58. W. Finnegan, R. A. Henry, and R. Lofquist, J. Am. Chem. Soc., 80 (1958) 3908; B. E. Huff and M. A. Staszak, Tetrahedron Lett., 50 (1993) 8011; B. S. Jursic and B. LeBlanc, J. Heterocyclic Chem. submitted.; B. S. Jursic and Z. Zdravkovski, Synth. Com., 24 (1994) 1575; B. S. Jursic and Z. Zdravkovski, J. Mol. Struct. (Theochem), 309 (1994) 241.
59. For example see: G. Piancatelli, M. D'Auria, and F. D'Onofrio, Synthesis, (1994) 867.
60. P. S. Engel, Chem. Rev., 80 (1980) 99.

61. J. Elguero, in: *Comprehensive Heterocyclic Chemistry*; K.T. Potts, (ed.), Pergamon: London, 1984; Vol. 5, p 247.
62. For example see: W. Adam, H. M. Harrer, W. M. Nau, and K. Peters, *J. Org. Chem.* 59 (1994) 3786; W. Adam, H. M. Harrer, W. M. Nau, and K. Peters, *J. Org. Chem.* 59 (1994) 7069.
63. *Weygand/Hilgertag Preparative Organic Chemistry*, G. Hilgertag and A. Martini,(eds.), J. Wiley, New York, NY, 1972.

This Page Intentionally Left Blank

Triplet Photoreactions; Structural Dependence of Spin-Orbit Coupling and Intersystem Crossing in Organic Biradicals

M. Klessinger

Organisch-Chemisches Institut, Westf. Wilhelms-Universität,
Corrensstraße 40, D-48149 Münster, Germany

1. INTRODUCTION

Triplet biradicals or biradicaloids are frequently intermediates in organic photoreactions, in particular in reactions initiated from the $n\pi^*$ state of photo-excited carbonyl compounds [1,2]. In the course of the reaction the triplet species has to experience intersystem crossing (ISC) to the singlet state before proceeding to the final singlet products. Since product formation on the ground-state surface (S_0) is very fast and does not allow major conformational changes, the triplet state geometry that is most favorable for ISC determines the ratio of the products in the overall reaction.

There are three mechanisms of spin flipping: solvent-induced spin relaxation (spin-lattice relaxation), spin-orbit coupling (SOC) and hyperfine coupling (HFC) [2,3]. While the first of these mechanisms is quite slow in the absence of paramagnetic impurities, HFC is important in biradicals in which the two radical centers are relatively far apart (1,6-biradicals and longer) [4], and SOC dominates in short biradicals, which are observed as intermediates in numerous photochemical reactions. For these systems, the order of magnitude of the SOC element is about $0.1\text{-}5\text{ cm}^{-1}$, which is much larger than that of the typical HFC, which is about 0.001 cm^{-1} . We will therefore concentrate on the SOC mechanism only.

The rate constant k_{ISC} for the nonradiative ISC step from the triplet state T_1 can then be estimated by Fermi's Golden Rule [5]

$$k_{\text{ISC}} = \frac{2\pi}{\hbar} |\langle T_1 | \hat{H}^{\text{SO}} | S_0 \rangle|^2 \rho(E), \quad (1)$$

where \hat{H}^{SO} is the spin-orbit coupling operator and $\rho(E)$ the density of states in the final electronic state. $\rho(E)$ accounts for the fact that the T_1 - S_0 energy gap has to be small for the SOC coupling mechanism to become efficient [6,7]. The SOC elements as well as the T_1 - S_0 energy splitting depend on the molecular geometry, and both effects have to be considered simultaneously, since a small T_1 - S_0 energy splitting E_{ST} combined with small SOC as well as a large SOC combined

with a large E_{ST} will both lead to negligible ISC rates. Furthermore, ISC usually occurs at geometries different from the minimum geometry of the initial state [8]. For an understanding of SOC effects in photoreactions it is therefore necessary to determine potential energy surfaces (PES) to locate the accessible areas and the T_1-S_0 gap as well as to evaluate SOC for each geometry of interest.

Since the *ab initio* determination of the structural dependence of these quantities is feasible only for small systems, we present here in Section 2 a formalism for a straightforward determination of SOC within the context of configuration interaction (CI) calculations [9]. This formalism has been implemented for the semiempirical MNDOC-CI method [10] and allows for routine determinations of PE surfaces and SOC surfaces at the same level of theory. Some results for 1,1-, 1,2-, 1,3- and 1,4-biradicals will be given in Section 3 and compared to *ab initio* results available from the literature. In Section 4 some simple models will be discussed, which allow for a rationalization of the structural dependence of SOC in the biradicals discussed in the previous sections, and finally, in Section 5 some general conclusions are presented concerning the mechanisms of organic triplet photoreactions.

2. BASIC THEORY

2.1 Wave functions and operators

The main effect of taking into account spin-orbit coupling in excited state calculations will be an admixture of triplet character to the singlet states S_k and of singlet character to the triplet states T_j . We will confine the following discussion to this situation and use the configuration interaction (CI) approach to describe singlet and triplet wave functions

$${}^{1,3}\Psi_k(1,\dots,N) = \sum_I C_{Ik} {}^{1,3}\Phi_I(1,\dots,N). \quad (2)$$

The spin-adapted configurational state functions (CSF's) ${}^{1,3}\Phi_I(1,\dots,N)$ may in general be expressed in terms of Slater determinants

$$|\Psi_1(1)\dots\Psi_n(N)| = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(x_1) & \dots & \psi_1(x_N) \\ \vdots & & \vdots \\ \psi_N(x_1) & \dots & \psi_N(x_N) \end{vmatrix} \quad (3)$$

built up from spin orbitals

$$\psi_i(x) = \phi_i(\mathbf{r}) \xi(s) = \phi_i(\mathbf{r}) \alpha(s) \text{ or } \phi_i(\mathbf{r}) \beta(s), \quad (4)$$

the space part $\phi_i(\mathbf{r})$ of which is given as linear combination of atomic orbitals χ_μ ,

$$\phi_i(\mathbf{r}) = \sum_{\mu} c_{\mu i} \chi_{\mu}(\mathbf{r}). \quad (5)$$

The spin-orbit coupling strength

$$\begin{aligned} SOC_{lk} &= \left[\sum_m |\langle T_{l,m} | \hat{H}^{SO} | S_k \rangle|^2 \right]^{1/2}, \quad m = -1, 0, 1 \\ &= \left[\sum_u |\langle T_{l,u} | \hat{H}^{SO} | S_k \rangle|^2 \right]^{1/2}, \quad u = x, y, z, \end{aligned} \quad (6)$$

which can be thought of as the length of a spin-orbit coupling vector \mathbf{SOC}_{lk} with components $\langle T_{l,u} | \hat{H}^{SO} | S_k \rangle$, can then be expressed in terms of matrix elements of the SOC operator \hat{H}^{SO} between MOs ϕ_i or basis functions χ_μ .

The operator

$$\begin{aligned} \hat{H}^{SO} &= \hat{H}_1^{SO} + \hat{H}_2^{SO} \\ &= \sum_i \hat{h}_1^{SO}(i) + \sum_{i \neq j} \hat{h}_2^{SO}(i,j) \end{aligned} \quad (7)$$

describes the weak interactions between singlet and triplet states and has the form

$$\hat{h}_1^{SO}(i) = g\beta_e^2 \sum_{\kappa} Z_{\kappa} |r_{\kappa}|^{-3} l_{\kappa}^{\kappa} \cdot s_i \quad (8)$$

$$\hat{h}_2^{SO}(i,j) = -g\beta_e^2 r_{ij}^{-3} [(\mathbf{r}_i - \mathbf{r}_j) \times \hat{\mathbf{p}}_i] \cdot (s_i + 2s_j), \quad (9)$$

where the sum runs over all nuclei κ with atomic number Z_κ , \hat{r}_κ and ℓ_κ are the position and the angular momentum operators with the nucleus κ taken as the origin, \hat{s} is the spin angular momentum operator, \hat{p} the linear momentum operator and $\beta_e = c\hbar/2mc$.

Numerical *ab initio* calculations for selected examples with polarized basis sets and CI of reasonably size confirmed that the size of the matrix elements H_2^{SO} within the active space matrix is negligible. In contrast, the elements of H_2^{SO} that involve both the active and inner shells are large, since H_2^{SO} is primarily due to the shielding of nuclei by inner-shell electrons [11]. It is therefore common practice in many semiquantitative applications, to account for the effect of the fixed-core electrons by replacing the factor $g\beta_e^2 Z_\kappa |\mathbf{r}_\kappa|^{-3}$ in H_1^{SO} by the empirical value of the atomic spin-orbit coupling constant ζ_κ for the valence p orbitals on atom κ [12].

The effective one-electron SOC operator then reads

$$\begin{aligned} H^{\text{SO}} &= \sum_i \hat{L}^{\text{SO}}(i) = \sum_i \sum_\kappa \zeta_\kappa \ell_\kappa(i) \cdot \hat{s}(i) \\ &= \sum_i \sum_\kappa \zeta_\kappa \left\{ \frac{1}{2} [\ell_\kappa^-(i) \hat{s}^+(i) + \ell_\kappa^+(i) \hat{s}^-(i)] + \ell_\kappa^z(i) \hat{s}^z(i) \right\} \end{aligned} \quad (9)$$

where $\ell^\pm = \ell_x \pm i\ell_y$ and $\hat{s}^\pm = \hat{s}_x \pm i\hat{s}_y$ are step-up and step-down ladder operators. Making use of the second quantization approach (SQA) and introducing two sets of creation and annihilation operators $\{\hat{a}_i^+, \hat{a}_i\}$ and $\{\hat{b}_i^+, \hat{b}_i\}$ relating to orthogonal orbitals with α and β spin, respectively, the SOC operator may be written as

$$H^{\text{SO}} = \frac{1}{2} \sum_{ij} \left\{ L_{ij}^- \hat{a}_i^+ \hat{b}_j + L_{ij}^+ \hat{b}_i^+ \hat{a}_j + L_{ij}^z [\hat{a}_i^+ \hat{a}_j - \hat{b}_i^+ \hat{b}_j] \right\}, \quad (11)$$

where

$$L_{ij}^\sigma = \langle \phi_i | \sum_\kappa \zeta_\kappa \ell_\kappa^\sigma | \phi_j \rangle, \quad \sigma = +, -, z. \quad (12)$$

In eq. (11) the operator chain in the last term of the sum corresponds to the spin-density operator \hat{Q}_{ij} [13].

2.2 Matrix elements between bonded functions

There are a number of different methods for the construction of spin eigen-

functions [14] but not all of the resulting functions are equally suited for computations involving spin-dependent operators. Matrix elements of a general spin-dependent one-electron operator between bonded functions [15], i.e. eigenfunctions of total spin constructed with the help of valence-bond or Rumer diagrams, have been derived by Manne and Zerner [16]. An alternative scheme for an efficient evaluation of matrix elements of the SOC operator is provided by an extension of the method of Golebiewski and Broclawik [17] to spin-dependent one-electron operators.

The bonded functions are defined by

$$|I\rangle = \Phi_I = \hat{A}(i_1 i_2)(i_3 i_4) \dots (i_{g-1} i_g)(i_{g+1}) \dots (i_N), \quad (13)$$

where

$$(ij) = 2^{-1/2} \phi_i(1)\phi_j(2) [\alpha(1)\beta(2) - \beta(1)\alpha(2)] \quad \text{for } i \neq j, \quad (14)$$

$$(ii) = \phi_i(1)\phi_i(2) \alpha(1)\beta(2) \quad (15)$$

and

$$(i) = \phi_i(1) \alpha. \quad (16)$$

The antisymmetrizer \hat{A} includes the conventional normalizing factor $(N!)^{-1/2}$. In order to be able to treat triplet states with $M_S \neq S$, the concept of bonded functions has been extended and the additional building units

$$(ij)_+ = 2^{-1/2} \phi_i(1)\phi_j(2) [\alpha(1)\beta(2) + \beta(1)\alpha(2)] \quad \text{for } i \neq j \quad (17)$$

and

$$(\bar{i}) = \phi_i(1) \beta(1) \quad (18)$$

have been defined [9].

The matrix element

$$H_{IJ}^{\text{SO}} = \langle I | \hat{H}^{\text{SO}} | J \rangle \quad (19)$$

between the extended bonded functions $|I\rangle$ and $|J\rangle$ is nonzero only if $|I\rangle$ and $|J\rangle$ differ at most in the occupation number of one orbital a in $|I\rangle$ and A in $|J\rangle$, say. Its evaluation may then be achieved in two steps. In the first step, the action of the operators $\{\hat{a}_i^+ \hat{a}_j - \hat{b}_i^+ \hat{b}_j\}$, $\hat{a}_i^+ \hat{b}_j$ and $\hat{b}_i^+ \hat{a}_j$ on $|J\rangle$ is considered. A single new extended bonded function $|J'\rangle$ results, which is not necessarily a spin eigenfunction. The second step then consists of calculating exactly one overlap integral $\langle I | J' \rangle$ between bonded functions in order to obtain the coefficient of the MO integral L_{ij}^σ in the matrix element of eq. (19). This overlap integral, which is zero if $|I\rangle$ and $|J'\rangle$ differ by one orbital or more, can be calculated by extending the contraction rules of Golebiewski and Broclawik to include all possibilities that arise from the extended functions. Tables of the operator rules and the contraction rules are given in the original paper [9].

2.3 Evaluation of spin-orbit integrals

For the evaluation of the MO integrals L_{ij}^σ defined in eq. (12) we introduce the LCAO approximation $\phi_i = \sum_{\mu} c_{\mu i} \chi_{\mu}$ and obtain

$$L_{ij}^\sigma = \sum_{\mu} \sum_{\nu} c_{\mu i} c_{\nu j} \langle \chi_{\mu} | \sum_{\kappa} \zeta_{\kappa} \ell_{\kappa}^{\sigma} | \chi_{\nu} \rangle. \quad (20)$$

Within the ZDO approximation only one-center contributions are considered and all neglected terms are lumped into the empirical spin-orbit coupling parameters ζ_{κ} . Then,

$$L_{ij}^\sigma = \sum_{\kappa} \zeta_{\kappa} \sum_{\mu(\kappa)} \sum_{\nu(\kappa)} c_{\mu i}^{\kappa} c_{\nu j}^{\kappa} \langle \chi_{\mu}^{\kappa} | \ell_{\kappa}^{\sigma} | \chi_{\nu}^{\kappa} \rangle, \quad (21)$$

and from the action of the angular momentum operators on s and p orbitals [6], which are the only ones to be considered here, one has

$$L_{ij}^+ = \sum_{\kappa} \zeta_{\kappa} \{ [c_{p_z i}^{\kappa*} c_{p_x j}^{\kappa} - c_{p_x i}^{\kappa*} c_{p_z j}^{\kappa}] + i [c_{p_z i}^{\kappa*} c_{p_y j}^{\kappa} - c_{p_y i}^{\kappa*} c_{p_z j}^{\kappa}] \} \quad (22)$$

$$L_{ij}^- = \sum_{\kappa} \zeta_{\kappa} \left\{ [c_{p_x i}^{\kappa*} c_{p_z j}^{\kappa} - c_{p_z i}^{\kappa*} c_{p_x j}^{\kappa}] + i [c_{p_z i}^{\kappa*} c_{p_y j}^{\kappa} - c_{p_y i}^{\kappa*} c_{p_z j}^{\kappa}] \right\} \quad (23)$$

$$L_{ij}^z = \sum_{\kappa} \zeta_{\kappa} \left\{ i [c_{p_y i}^{\kappa*} c_{p_x j}^{\kappa} - c_{p_x i}^{\kappa*} c_{p_y j}^{\kappa}] \right\}. \quad (24)$$

As the bonded functions of both states are built from the same set of real MOs $\{\phi_k\}$, $L_{ii}^{\sigma} = 0$, and thus H_{IJ}^{SO} is nonzero only if the bonded functions $|I\rangle$ and $|J\rangle$ differ exactly in one MO, i.e. if $i = a$ and $j = A$.

While in semiempirical valence-electron methods based on the ZDO approximation the spin-orbit coupling integrals reduce to overlap integrals [18], in *ab initio* calculations they can be expressed in terms of differentiated nuclear-attraction and electron-repulsion integrals [19,20]. Corresponding methods for computation of one- and two-electron spin-orbit integrals over Gaussian-type basis functions have been developed by King and Furlani [20], and are included in the GAMESS program package [21]. Effective nuclear charges for SOC calculations were determined on the basis of MCSCF calculations [22].

In order to evaluate the Cartesian components of the SOC vector, the principal magnetic axes have to be determined by considering the spin-spin dipolar coupling [23]

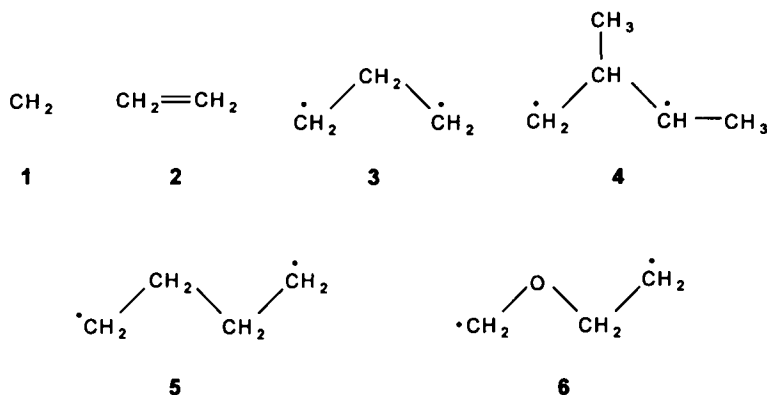
$$\hat{H}^{SS}(i,j) = g^2 \beta_e^2 \left\{ \hat{s}(i) \hat{s}(j) / r_{ij}^3 - 3 [\hat{r}_{ij} \cdot \hat{s}(i)] [(\hat{r}_{ij} \cdot \hat{s}(j)) / r_{ij}^5] \right\}. \quad (25)$$

If a molecule has symmetry C_{2v} or higher, the space integrals vanish by symmetry and the positions of the magnetic axes can be taken along the geometrical axis of the system. If the symmetry is lower, the molecular axes have to be chosen as the principal axes of the spin-spin dipolar coupling tensor. In semiempirical calculations based on the ZDO approximation, the point-charge approximation introduced by McWeeny [24] may be used, while in *ab-initio* methods one-center expansions [25] and integral evaluations over STOs and GTOs have been described [23].

3. SPIN-ORBIT COUPLING AND INTERSYSTEM CROSSING IN BIRADICALS

The theory of spin-orbit coupling in organic biradicals essentially started in 1972 with the analysis by Salem and Rowland [7] performed for the 2-in-2 model of biradical electronic structure (active space: two electrons in two orbitals, also known as the 3×3 CI model). A fair amount of theoretical work on SOC in

biradicals and biradicaloids followed, which confirmed and elaborated the Salem-Rowland rules [26]. Computational results have been obtained for 1,1-, 1,2-, 1,3- and 1,4- biradicals, but combined analyses of SOC and potential energy (PE) surfaces for the S_0 and T_1 state, which allow for a discussion of ISC processes and triplet state reactivity are scarce.



In this section, literature data will be briefly reviewed and PE and SOC surfaces based on semiempirical MNDOC-CI calculations and the formalism for calculating SOC effects given in Section 2 will be presented for carbene (1), ethylene (2), trimethylene (3), and tetramethylene (5). Stereoselectivity of triplet photoreactions is discussed using 1,2-dimethyltrimethylene (4) as an example, and the effect of heavy atoms is demonstrated by comparing the results for tetramethylene with those for 2-oxatetramethylene (6), which plays an important role in the Paterno-Büchi reaction. In these calculations CI is based on half-electron MOs and includes all single and double excitations within a given active space, which in general consists of the three highest occupied and the four lowest unoccupied MOs of the closed-shell configuration (3-4' CI); for triplet states the same active space is used but single and double excitations are defined with respect to the open-shell configuration. Thus, 91 singlet and 183 triplet CSFs are obtained except for carbene with only three unoccupied MOs (3-3' CI), where 55 singlet CSFs and 96 triplet CSFs result [9].

3.1 Carbene

Ab initio determinations of SOC in carbene (1) have been carried out for three bond angles ($\theta = 90, 112$ and 135°) by McKellar *et al.* [27] using SCI and SDCI wave functions and by Vahtras *et al.* [19] for the 3B_1 equilibrium geometry employing a multiconfiguration linear response (MCLR) approach. Both sets of calculations are based on the full coupling operator $\hat{H}^{\text{SO}} = \hat{H}_1^{\text{SO}} + \hat{H}_2^{\text{SO}}$, with the contributions from the two-electron operator \hat{H}_2^{SO} reducing the total SOC value to about half the one-electron contribution. Increasing the size of the CI or the active space considerably reduces the calculated SOC values, typical values being 19.80 cm^{-1} , 10.35 cm^{-1} and 7.89 cm^{-1} for SCF, 2-in-2 CAS (two electrons in an active space of two orbitals, i.e. HOMO-LUMO CI) and full CI wave functions.

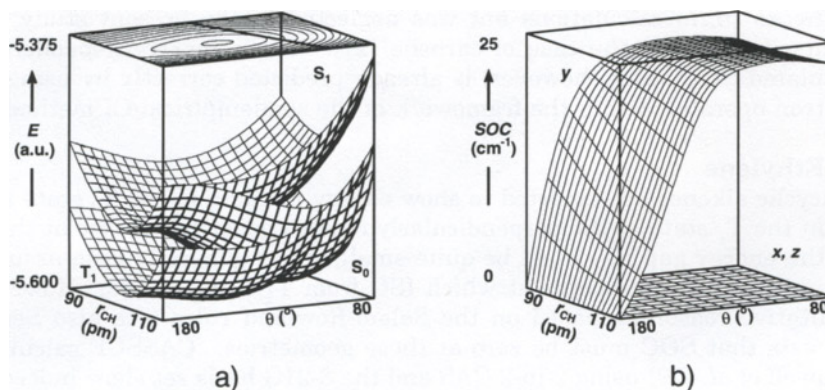


Figure 1. MNDOC-CI results for carbene: (a) state energies including contour map for T_1 and isopotential line for $E_{ST} = 0$; (b) SOC values.

MNDOC-CI results are given in Figure 1. Potential energy surfaces are plotted as a function of the CH distance r_{CH} and the HCH angle θ in Figure 1a, while Figure 1b gives the SOC_x , SOC_y and SOC_z surfaces plotted in the same way. Nonzero values are obtained only for SOC_y , which is nearly independent of the CH distance and increases from zero for $\theta = 180^\circ$ to a maximum value of about 23 cm^{-1} near $\theta = 110^\circ$ and stays more or less constant for smaller angles θ . Figure 1a also gives the contour-line diagram of the T_1 state together with the intersection line between the T_1 and S_0 states ($E_{ST} = E_S - E_T = 0$), which lies only 5.5 kcal/mol and 2 kcal/mol above the minima on the S_0 and the T_1 surface, respectively. The large SOC value at these geometries suggests a nearly unimpeded ISC transition, in agreement with experimental findings for carbene in the gas phase and for diphenylmethylene in solution [28], that demonstrate the existence of a thermodynamic equilibrium between singlet and triplet carbene.

Table 1
Carbene SOC values in cm^{-1} from different methods

θ	MNDOC-CI ^a	MCCI ^b	MCLR ^c
90	22.9	11.7 - 13.1	
112	23.4	11.7 - 13.2	
135 ^d	22.6	10.2 - 12.5	7.9 - 11.1

^a $r_{CH} = 110 \text{ pm}$ [9]. ^b Multiconfiguration - CI, $r_{CH} = 109.6 \text{ pm}$ [27].
^c Multiconfiguration linear response, $r_{CH} = 108.2 \text{ pm}$ [19]. ^d Exact value: 135.0 (MNDOC-CI), 135.1 (MCCI), and 132.4 (MCLR).

In Table 1 *ab initio* values are compared to our results, which are seen to be some 10 cm^{-1} too large. The reason for this discrepancy can be seen in the fact that the two-electron part of the spin-orbit coupling operator, which was retained

in the *ab initio* calculations but was neglected in the present study, is extraordinarily large in the case of carbene [27]. The geometry dependence of the calculated *SOC* value, however, is already predicted correctly by using the one-electron operator within the framework of the semiempirical CI method.

3.2 Ethylene

Acyclic alkenes are expected to show energy maxima in the S_0 state and minima in the T_1 state at the perpendicularly twisted geometry. Since at this geometry the energy gap E_{ST} must be quite small, it has generally been assumed that it represents the geometry at which ISC from T_1 to S_0 occurs. However, even qualitative reasoning based on the Salem-Rowland rules (see also Section 4.1) suggests that *SOC* must be zero at these geometries. CASSCF calculations by Caldwell *et al.* [29] using 2-in-2 CAS and the 3-21G basis set show indeed, that on twisting maximum *SOC* occurs in ethylene (2) at $\phi \approx 50^\circ$ ($SOC = 0.6 \text{ cm}^{-1}$), where the energy gap E_{ST} is substantial. But C_{2v} pyramidalization affords large *SOC*, and it was therefore suggested, that both twisting and pyramidalization contribute and provide a reasonable range of geometries for ISC of unconstrained alkene triplets [29]. Minaev *et al.* [30] calculated the 90° -twisted optimized geometry of the T_1 state to be slightly pyramidalized with a corresponding nonzero *SOC* value of 0.13 cm^{-1} . Using the same CASSCF technique as for ethylene, the heavy atom effect in vinyl chloride was shown by Caldwell *et al.* [31] to be quite large, increasing the *SOC* values approximately by a factor of 18 independent of the twist angle. Replacement of H by a CH_3 group to yield propene, on the other hand, causes no significant change in *SOC*.

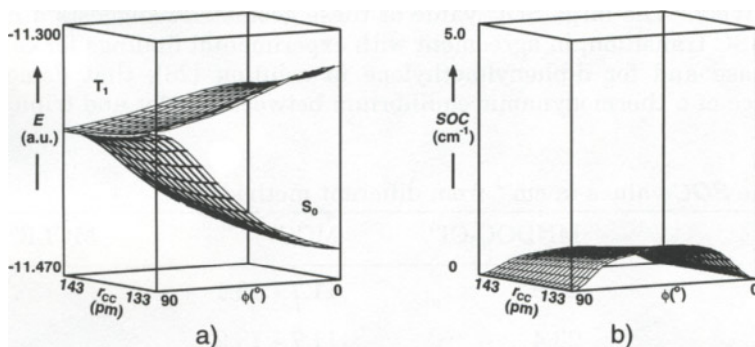


Figure 2. MNDOC-CI results for ethylene: (a) state energies and (b) *SOC* values as a function of r_{CC} and ϕ . The remaining geometry parameters r_{CH} and $\angle HCH$ are fixed to 109 pm and 117° , respectively.

These findings are well reproduced by semiempirical MNDOC-CI calculations, as illustrated by the data collected in Table 2. In Figure 2 energies E of the S_0 and T_1 states as well as *SOC* values are plotted as a function of ϕ and r_{CC} ; the maximum values are 0.411 cm^{-1} for $r_{CC} = 143 \text{ pm}$ at $\phi = 43^\circ$ and 0.747 cm^{-1} for $r_{CC} = 133 \text{ pm}$ at $\phi = 40^\circ$. A comparison of Figure 2a and b also shows that the triplet minimum occurs at a value of the twist angle ($\phi = 90^\circ$) appreciably diffe-

Table 2
Ethylene SOC values in cm^{-1} from different methods for torsion (ϕ) of the double bond

ϕ	MNDOC-CI ^a	MCSCF ^b	MCQR ^c
0	0.00	0.00	—
45	0.440		
50	0.437	0.6	—
90	0.003	0.0	0.13 ^d

^a Geometry taken from ref. [29]. ^b 2-in-2-CASSCF, 3-21G basis [29]. ^c Multiconfiguration quadratic response, triple zeta basis [30]. ^d Pyramidalization $\omega = 9.1^\circ$.

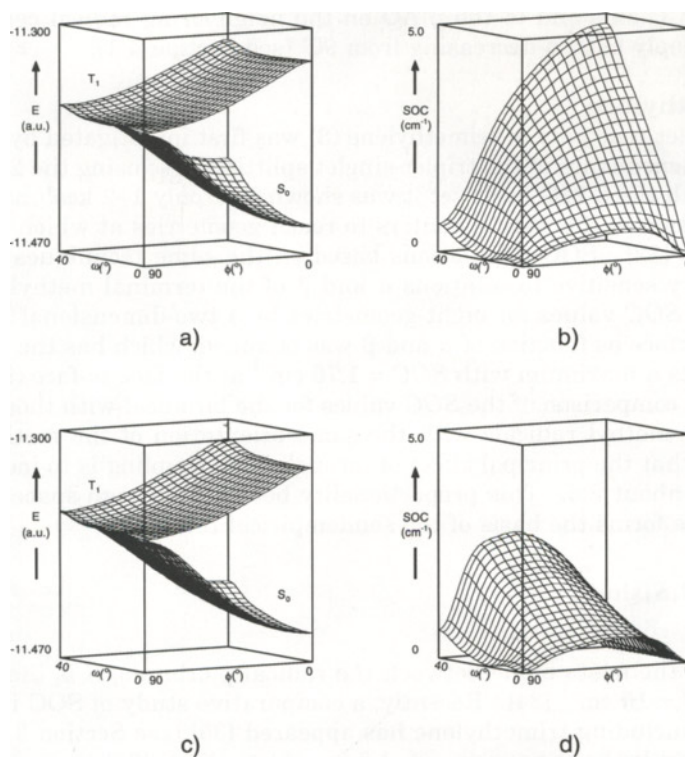


Figure 3. MNDOC-CI results for torsion (ϕ) and pyramidalization (ω) in ethylene: (a) state energies and (b) SOC values for C_{2v} pyramidalization; (c) state energies and (d) SOC values for C_{2h} pyramidalization. Geometry parameters as in Figure 2 with r_{CC} varied linearly with ϕ from 133 to 143 pm.

rent from that of maximum SOC ($\phi = 40\text{--}50^\circ$). At the triplet minimum the singlet-triplet separation E_{ST} is particularly small, but this is true also for the SOC values. 7 kcal/mol are necessary to reach geometries of maximum SOC , but at these geometries E_{ST} is nearly 40 kcal/mol.

In Figure 3 the dependence of the S_0 and T_1 energies and of the SOC value on the twist and pyramidalization angles ϕ and ω is shown, where ω is the acute angle between the CC bond direction and the bisector of the HCH angle. Results are shown only for the simultaneous pyramidalization of both methylene groups, either in the same sense (C_{2v} , $\omega' = \omega$; Figure 3a,b), or in the opposite sense (C_{2h} , $\omega' = -\omega$; Figure 3c,d). For pyramidalization of just one methylene group the SOC surface has the same qualitative appearance as shown for $\omega' = \omega$ in Figure 3b, with the SOC values being about half the size.

Pyramidalization of one methylene group and C_{2v} pyramidalization of both methylene groups increases the SOC values linearly, the effect being very small for $\phi = 90^\circ$ and largest for $\phi = 0^\circ$. While the component of the angular momentum integral along the CC axis, which contributes to SOC in unpyramidalized ethylene, is hardly affected by pyramidalization, it is mainly the component perpendicular to the CC axis and to the p AO on the neighboring radical center which increases steeply with ϕ decreasing from 90° (see Section 4.1).

3.3 Trimethylene

The triplet reactivity of trimethylene (**3**) was first investigated by analyzing the geometry dependence of the triplet-singlet splitting E_{ST} ; using the 2-in-2 CASSCF method with the 3-21G basis set it was shown that only 1–2 kcal/mol are necessary for rotation of the radical centers to reach geometries at which T_1 and S_0 are degenerate [32]. SOC calculations based on the same techniques revealed that SOC is very sensitive to rotations α and β of the terminal methylenes [33]. By fitting the SOC values for eight geometries by a two-dimensional Fourier series an SOC surface as function of α and β was obtained, which has the shape of a hat and exhibits a maximum with $SOC = 1.76\text{ cm}^{-1}$ at the face-to-face ($90,90^\circ$) geometry [34]. A comparison of the SOC values for the biradical with those for a pair of interacting methyl radicals with the same orientation of the methylene groups suggested that the principal effect of through-bond coupling is to increase SOC by a factor of about 2.5. This proportionality between through-space and through-bond effects forms the basis of the semiempirical relation

$$SOC = B|S|\sin\phi, \quad (26)$$

where ϕ is the acute angle between the radical p orbitals, S is the overlap integral, and $B = 15\text{ cm}^{-1}$ [34]. Recently, a comparative study of SOC in various 1, n -biradicals including trimethylene has appeared [35] (see Section 3.5), and 2-in-2 CASSCF results have been reported for norbornadiene [36].

The *ab initio* results are again very well reproduced by MNDOC-CI calculations, as can be seen from Figure 4, which shows the SOC surface resembling a hat with the maximum $SOC = 1.91\text{ cm}^{-1}$ at the face-to-face orientation ($\alpha = \beta = 90^\circ$), and vanishing SOC if at least one of the CH_2 groups lies within the molecular plane ($\alpha = 0$ and/or $\beta = 0$).

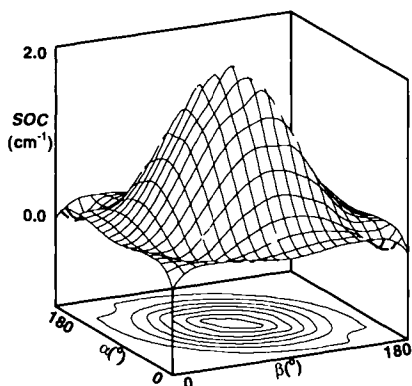


Figure 4. MNDOC-CI results for trimethylene: SOC values as a function of α and β (geometry taken from ref. [34]).

For all geometries considered in Figure 4 the T_1 surface is lower than S_0 by less than 2.5 kcal/mol and is rather flat. The energy differences between the minimum at $\alpha = \beta = 90^\circ$ and the maximum at $\alpha = \beta = 0$ is 1.5 kcal/mol. This is to say that the terminal CH_2 groups can rotate freely in the T_1 state and all geometries are easily accessible.

The main contribution to the total SOC value comes from the x component perpendicular to the molecular plane. The ℓ_x operator rotates the orbitals within the molecular (yz) plane, and since in the face-to-face conformation ($\alpha = \beta = 90^\circ$) both localized orbitals are located in this plane, this is the most favored conformation for spin-orbit coupling (cf. Section 4.1).

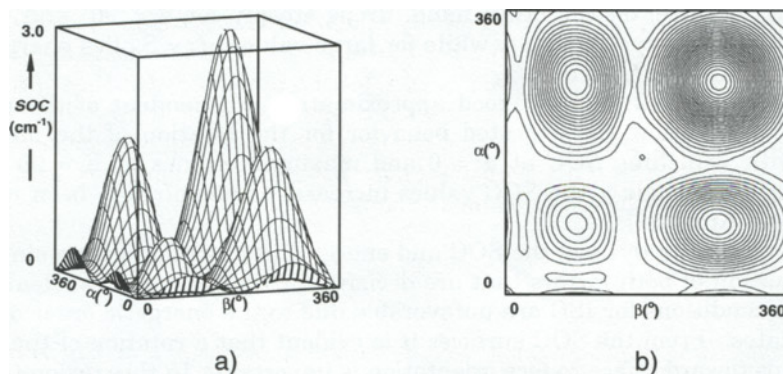


Figure 5. MNDOC-CI results for 20°-pyramidalized trimethylene: spin-orbit coupling as a function of α and β : (a) SOC surface and (b) contour map. The distance between contours is 0.1 cm^{-1} . For geometry parameters, cf. Figure 4.

In contrast to the situation for ethylene, pyramidalization of the radical centers has only a minor effect on SOC in the case of trimethylene. This is shown in Figure 5, where the rotational angles α and β now have been varied between 0°

and 360° . Due to pyramidalization different face-to-face conformations result, as indicated in Figure 6, and the *SOC* values for these conformations vary between 0.65 and 2.98 cm^{-1} .

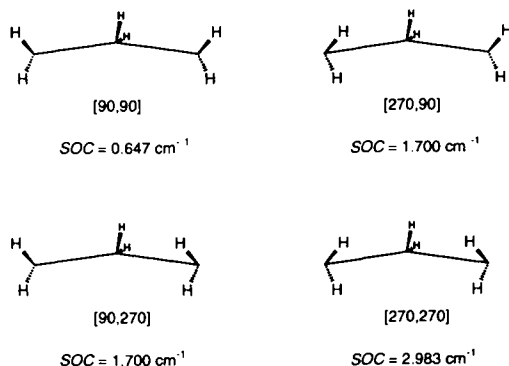


Figure 6. 20° -pyramidalized trimethylene: the four face-to-face geometries, and *SOC* values from MNDOC-CI calculations.

In addition to the rotational angles α and β the CCC bond angle γ is also important for the ISC process. It influences the energies of the singlet and triplet states and describes the reaction from the open-chain 1,3-biradical ($\gamma = 115^\circ$) to cyclopropane ($\gamma = 60^\circ$). Therefore, the potential energy surfaces as well as the *SOC* surface were calculated for a variation of γ and the conrotatory ($\alpha = \beta$) and disrotatory ($\beta = 180 - \alpha$) motion of the terminal methylene groups. From the results shown in Figure 7 for the conrotatory motion it is seen that the T_1 surface exhibits a shallow valley for $\gamma \approx 115^\circ$ and is strongly repulsive for small values of γ . The S_0 surface, on the other hand, drops steeply for $\gamma < 90^\circ$ and $\alpha > 45^\circ$ toward the product cyclopropane, while for large values of γ S_0 lies energetically above T_1 for all values of α .

The *SOC* surfaces are to a good approximation independent of the mode of rotation. They show the expected behavior for the rotation of the methylene groups with vanishing *SOC* at $\alpha = 0$ and maximum values at $\alpha = 90^\circ$. With decreasing CCC angle γ the *SOC* values increase, as has already been noted by Furlani and King [33].

The combined analysis of the *SOC* and energy surfaces allows for a simultaneous estimation of both factors that are decisive for the ISC process: Near the T_1 valley the conditions for ISC are unfavorable due to the energetic order of the S_0 and T_1 states. From the *SOC* surfaces it is evident that a rotation of the methylene groups toward a face-to-face orientation is important. In this region ($\alpha > 45^\circ$) a decrease of γ to values around 105° leads to a T_1 - S_0 intersection; only a few kcal/mol are required to reach this region where the geometries are most favorable for ISC. Here, the S_0 surface drops clearly toward the cyclopropane structure. This corresponds to a preferred formation of cyclic products, a preference postulated previously solely on the basis of the potential energy surfaces [32a]. The *SOC* surface, however, clearly emphasizes the importance of the *SOC* value for the ISC process.

3.4 1,2-Dimethyltrimethylene

In 1,2-dimethyltrimethylene (**4**) there are two different modes of conrotatory motion of the radical centers, leading to stereoisomeric cyclization products: rotation by positive ($\alpha > 0$) and negative ($\alpha < 0$) values of the rotational angle yields *cis*- and *trans*-dimethylcyclopropane, respectively.

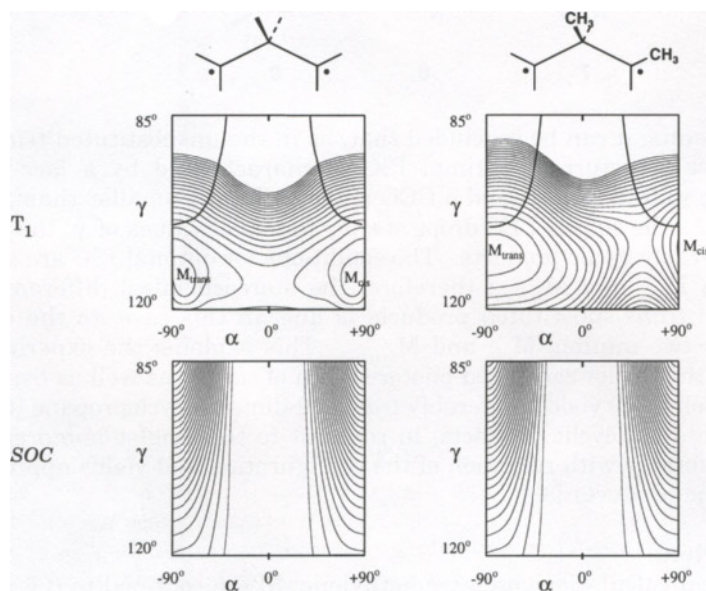
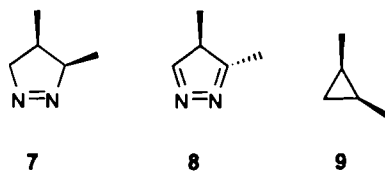


Figure 7. Triplet PE (top) and SOC surfaces (bottom) for the ring-closure reaction of trimethylene (left) and 1,2-dimethyltrimethylene (right), as a function of the CCC valence angle γ and the conrotatory motion ($\alpha = \beta$) of the radical centers. SOC values vary between 0 cm^{-1} in the middle of the diagram and 7.5 cm^{-1} at the upper right and left corners. T_1 - S_0 intersection $E_{ST} = 0$ is indicated by heavy lines.

T_1 and SOC surfaces are therefore shown in Figure 7 for the range $\alpha = -90^\circ$ to $\alpha = 90^\circ$ [37]. From a comparison of the results for the unsubstituted and the substituted trimethylene it is evident that to a first approximation the methyl substituents do not at all affect the SOC values, while steric effects appreciably change the appearance of the valley on the T_1 PES. For the *syn* mode of rotation a barrier of approximately 3.4 kcal/mol is found to separate the local minimum M_{cis} at $\alpha = 90^\circ$ from the planar structure ($\alpha = 0^\circ$) that is 1.0 kcal/mol lower in energy than the barrier. For the *anti* mode of rotation, however, the valley descends practically without a barrier toward the minimum M_{trans} at $\alpha = -90^\circ$, 5.0 kcal/mol lower in energy than M_{cis} . By decreasing the bond angle γ , geometries are accessible both from M_{cis} and M_{trans} at which S_0 and T_1 are degenerate ($E_{ST} = 0$) and SOC is appreciable ($SOC > 2 \text{ cm}^{-1}$), the necessary energies being smaller than 1 kcal/mol. For rotational angles $|\alpha| < 45^\circ$, however, much larger

energies (> 10 kcal/mol) are required and geometries with $SOC < 1$ cm⁻¹ are reached.



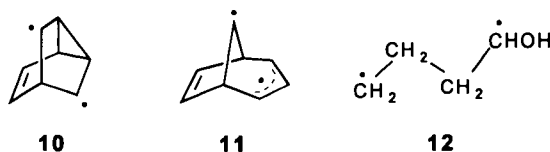
From these results, it can be concluded that, as in the unsubstituted trimethylene, the reactive structure of optimal ISC is characterized by a face-to-face orientation of the radical centers and a CCC angle γ slightly smaller than for the triplet minima. As the singlet PES drops steeply for small values of γ , the triplet state yields preferably cyclic products. The conditions for optimal ISC are similar for both minima M_{cis} and M_{trans} ; therefore, the stereochemical differentiation between *cis*- and *trans*-substituted products is due, in this case, to the energy difference of the two minima M_{cis} and M_{trans} . This explains the experimental observation that the triplet sensitized photoreaction of *cis*- (**7**) as well as *trans*-3,4-dimethyl-1-pyrazoline (**8**) yields preferably *trans*-1,2-dimethylcyclopropane (**9**), and negligible amounts of acyclic products, in contrast to the singlet photoreaction, which occurs preferably with retention of the configuration and yields appreciable amounts of acyclic products [38].

3.5 Tetramethylene

Most theoretical calculations on tetramethylene (**5**) are confined to the singlet biradical [39]. In view of the extremely flat potential it has been discussed, whether there exists a true minimum on the PE surface for this species or whether this flexible structure should rather be described as "twixtyl" [40], i.e. as operationally indistinguishable from a true minimum or as an entropy locked intermediate [41]. In contrast, more recent MCSCF/MP2 calculations with double zeta basis sets yield two minima, corresponding to a *trans* and a *gauche* structure. However, when zero-point vibrational energy corrections are taken into account, the *gauche* minimum disappears [42]. Calculations for the T_1-S_0 energy gap E_{ST} show that geometries with $E_{ST} = 0$ are easily encountered during the lifetime of the triplet biradical [43].

SOC calculations were recently reported for two bicyclic biradicals **10** and **11** derived from barrelene [44] and for 1,*n*-alkanediyls with $n = 3 - 8$, for which a unique alternation of SOC with the parity of the number of intervening bonds between the radical centers was found [35]. In addition, it was shown that omission of H_2^{SO} in eq. (7) leads to less serious errors than confining the active space to two MOs (HOMO-LUMO), and a dissection of SOC values into local hybrid orbital contribution was described [35].

1-Hydroxytetramethylene (**12**), the 1,4-biradical intermediate in the Norrish type II reaction, has been studied by 2-in-2 MCSCF calculations and all possible *gauche* and *trans* minima have been located on the singlet and triplet PE surfaces [45]. The singlet-triplet energy splitting and SOC values vary with the conformational geometry and have no correlation to the distance between the two radical



sites because of strong through-bond interactions. The ISC rate constant was calculated on the basis of molecular dynamics trajectories [45].

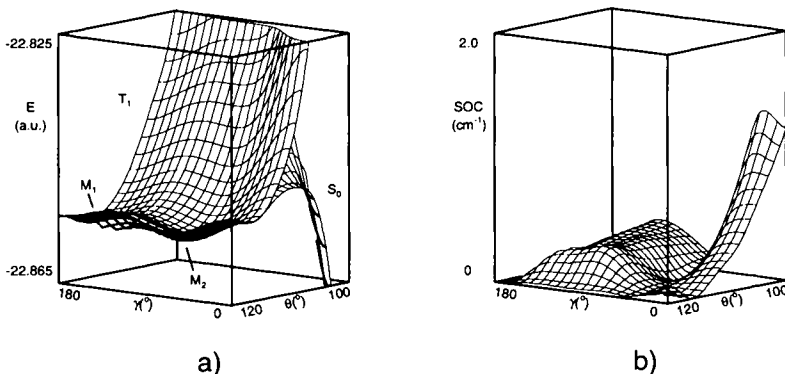


Figure 8. MNDOC-CI results for tetramethylene; a) state energies and b) *SOC* values as a function of the CCCC dihedral angle γ and the CCC valence angle θ (rotational angles α and β of the terminal methylene groups optimized for T_1).

For tetramethylene, the configurational space of interest is determined by the two torsional angles α and β and the $C_1C_2C_3C_4$ dihedral angle γ ; in addition, the CCC valence angle θ is important for the ring closure reaction to form cyclobutane. Optimization of the T_1 structure at the MNDOC-CI level yields two minima M_1 and M_2 corresponding to *anti* ($\gamma = 180^\circ$) and *gauche* ($\gamma = 62.3^\circ$) conformations of approximately the same energy, separated by a barrier of ~ 1.5 kcal/mol. PE and SOC surfaces in Figure 8 show that the T_1 energy rises with decreasing valence angle θ , while the S_0 energy descends for small values of γ and θ toward the cyclobutane minimum; SOC becomes particularly large at small values of θ and synperiplanar geometries ($\gamma \leq 30^\circ$) [46].

In Figure 9 T_1 and SOC surfaces are shown as a function of the rotational angles α and β of the terminal CH_2 groups for different values of the dihedral angle γ . They demonstrate that the behavior of a triplet 1,4-biradical is determined by two opposing trends: SOC is large at *syn* geometries, but regions of T_1 - S_0 degeneracy are easiest to reach at *anti* geometries. These results account for the experimental results: Due to the shape of the T_1 surface the system will reach the region of *anti* geometries very quickly. Here, *SOC* values are only about one tenth of the maximum value, but the high probability of finding the system at these geometries together with the small E_{ST} for nearly all values of the rotational angles α and β will favor the ISC process, yielding mainly open-chain products [47]. Cyclic products, on the other hand, can result only from ISC

in the region of *syn* geometries. The low stereospecificity of the triplet ring-closure reaction [48] is then due to the fact that geometries with $E_{ST} \approx 0$ and large *SOC* values are found in this region to occur at face-to-edge conformations, i.e., at geometries at which just one of the terminal methylene groups has been rotated by approximately 90° .

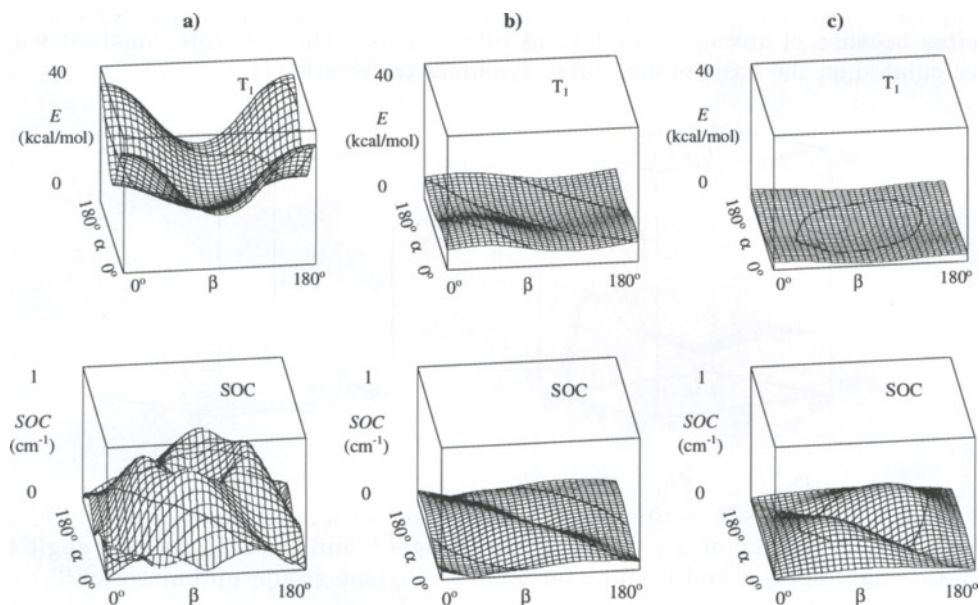
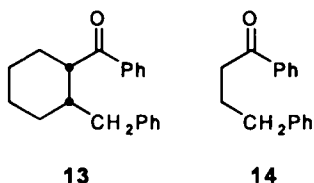


Figure 9. Triplet PE and SOC surfaces of tetramethylene for rotations α and β of the radical centers; a) *syn* conformation (CCCC dihedral angle $\gamma = 0^\circ$), b) *gauche* conformation ($\gamma = 60^\circ$) and c) *anti* conformation ($\gamma = 180^\circ$) (MNDOC-CI results for a CCC valence angle $\gamma = 105^\circ$).

The results also provide a rationalization of the rather surprising finding of Caldwell *et al.* [49], that the ISC rate constant for the Norrish type II reaction of **13**, which is conformationally fixed at $\gamma = 60^\circ$, and of the flexible analogue **14** are nearly the same. In **13**, rotation around one of the terminal CC bonds will lead to T_1 - S_0 degeneracy with $SOC = 0.28 \text{ cm}^{-1}$, which corresponds exactly to the situation in **14**, which will prefer the *anti* conformation where near S_0 - T_1 degeneracies are found for any α and β with $SOC < 0.3 \text{ cm}^{-1}$. This explains why other effects like the solvents affect the ISC rate to a larger extent than fixation of the biradical to $\gamma = 60^\circ$.



3.6 Oxatetramethylene

Similarly to tetramethylene two minima M_1 and M_2 of comparable energy and separated by a barrier of only 1 kcal/mol were located for the *anti* ($\gamma = 180^\circ$) and the *gauche* conformation ($\gamma = 78^\circ$) of 2-oxatetramethylene (**6**) [50] in agreement with the results of CASSCF calculations using a 6-31G* basis set [51]. *SOC* values for *syn* conformations are larger than for tetramethylene by a factor of approximately five, while typical SOC values at the triplet minima of *anti* conformations are 0.0 - 0.2 cm^{-1} .

The triplet PE and SOC surfaces for rotations α and β of the radical centers are shown in Figure 10 for conformations with $\gamma = 0^\circ$ (*syn*), $\gamma = 30^\circ$, $\gamma = 60^\circ$ (*gauche*) and $\gamma = 180^\circ$ (*anti*). In contrast to tetramethylene, the *SOC* values for *syn*-oxatetramethylene exhibit two pronounced maxima with $\text{SOC} = 5.49 \text{ cm}^{-1}$ for $\alpha = 90^\circ$ and $\beta = 45^\circ$ or 135° . With increasing γ , one of the maxima decreases and disappears, while the other one is shifted to larger β values until for $\gamma = 60^\circ$ only one maximum with $\text{SOC} = 2.85 \text{ cm}^{-1}$ at $\beta = 180^\circ$ is left over. Near $\gamma = 90^\circ$ the two maxima show up again with more or less unchanged α and β angles and nearly constant $\text{SOC} = 1.65 \text{ cm}^{-1}$.

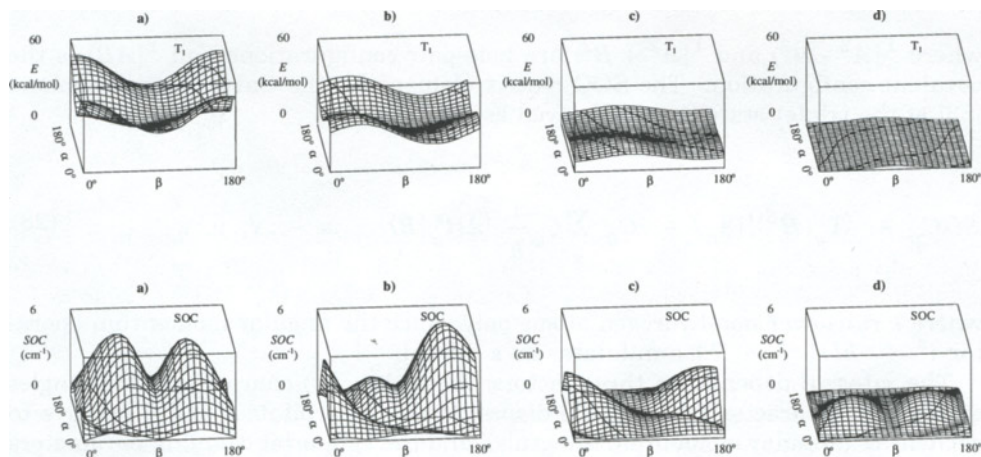


Figure 10. Triplet PE and SOC surfaces of 2-oxatetramethylene for rotations α and β of the radical centers; a) *syn* ($\gamma = 0^\circ$), b) $\gamma = 30^\circ$, c) *gauche* ($\gamma = 60^\circ$) and d) *anti* ($\gamma = 180^\circ$) conformation (MNDOC-CI results for $\angle\text{CCO} = \angle\text{COC} \approx 105^\circ$).

The energetically most favorable points on the T_1 - S_0 intersection line ($E_{\text{ST}} = 0$) are for *syn* and *anti* geometries equally far from the SOC maxima. This is due to the fact that for large γ T_1 minima for oxatetramethylene are found at $\alpha \approx \beta \approx 0^\circ$, while for tetramethylene one has $\alpha \approx \beta \approx 90^\circ$. Barriers to rotation for reaching the $E_{\text{ST}} = 0$ line are for all values of γ approximately 1 kcal/mol larger than for tetramethylene.

4. MODELS FOR SPIN-ORBIT COUPLING

The present basis for qualitative understanding of the structural dependence of SOC at biradicaloid geometries is an analysis by Salem and Rowland performed in the 2-in-2 model of biradical electronic structure [7]. This model has been extended and a more rigorous version of the Salem-Rowland rules has been derived by Michl [11]. On the basis of this model the results of the previous section will be discussed, and finally, it will be shown how symmetry determines the essential features of SOC in 1,*n*-biradicals.

4.1 The 2-in-2 model

Within the two-electron two-orbital model of Michl and Bonačić-Koutecký [52], i.e. confining the treatment of the biradical electronic structure to the two fully localized radical-carrying orbitals *A* and *B*, the singlet ground state (S_0) wave function may be written as

$$S_0 = C_{0,-} |A^2 - B^2\rangle + C_{0,+} |A^2 + B^2\rangle + C_{0,0} |AB\rangle, \quad (27)$$

where $|A^2 - B^2\rangle$ and $|A^2 + B^2\rangle$ are hole-pair configurations and $|AB\rangle$ is the covalent configuration. The SOC matrix element for the Cartesian components [53] of the triplet state is then derived as [54]

$$SOC_u = \langle T_u | \hat{H}^{SO} | S_0 \rangle = C_{0,+} \sum_{\kappa} \zeta_{\kappa} \frac{i}{\hbar} \langle A | \ell_u^{\kappa} | B \rangle, \quad u = x, y, z \quad (28)$$

where κ runs over non-hydrogen atoms only, since the angular momentum operator $\ell^{\kappa} = \hbar/i(\hat{r}^{\kappa} \times \nabla)$ annihilates an *s* orbital.

The integral depends on three factors: the $|A^2 - B^2\rangle$ character of the singlet state ("ionic character"), the spatial disposition of the orbitals *A* and *B* relative to each other (angular momentum integrals), and the spin-orbit coupling parameters ζ_{κ} (heavy atom effect).

Expressing the orbitals *A* and *B* in terms of an AO basis yields

$$SOC_u = C_{0,+} \sum_{\kappa} \zeta_{\kappa} \sum_{\mu, \nu} c_{A\mu} c_{B\nu} \langle \mu | (\hat{r}^{\kappa} \times \nabla)_u | \nu \rangle, \quad (29)$$

and allows for a discussion of the SOC vector in terms of atomic vector contributions provided by each atom κ .

Based on these results, Michl [11] derived the following revised formulation of the Salem-Rowland rules for large SOC between T_1 and S_0 :

(1) The most localized orthogonal orbitals A and B singly occupied in T_1 must either interact covalently through a non-zero resonance integral and/or must be sufficiently different in energy for one to have electron occupancy near two in S_0 .

(2) The biradical must contain one or more high- Z atoms at which one p orbital contributes strongly to A and another to B .

(3) These p orbitals must enter into A and B in a manner such that the contributions on all such atoms add rather than cancel.

The following examples drawn from the original paper by Michl [11], where more details will be found, will illustrate these rules.

Carbene. In linear carbene (valence angle $\theta = 180^\circ$) the orbitals A and B are represented by the degenerate and non-interacting p_x and p_y AOs on carbon; carbene is a perfect biradical with the lowest two singlet states S_0 and S_1 degenerate and the triplet state T_1 the ground state. For $\theta < 180^\circ$ the orbitals A and B still have a zero resonance integral, but their energies are different since one contains an admixture of s character. Within the 2-in-2 model this can be treated as a heterosymmetric perturbation δ that causes the hole-pair states S_1 and S_2 to interact, so the lower one drops immediately below the covalent state and eventually crosses the T_1 state and becomes the ground state. Already a small perturbation is sufficient to guarantee that both conditions (1) and (2) are fulfilled, while condition (3) is irrelevant. SOC is therefore expected to increase strongly as the HCH valence angle decreases from 180° , as shown in Figure 1.

Ethylene. 90° -twisted ethylene is a perfect biradical; distortion toward planarity ($\phi < 90^\circ$) leads to an interaction $\gamma_{AB} \neq 0$ of the localized orbitals A and B , yielding a homosymmetric biradicaloid, and the coefficient $C_{0,+}$ of the hole-pair configuration in the singlet ground state increases with increasing γ . That is to say, ethylene violates condition (1) but satisfies condition (2) when it is orthogonally twisted, and satisfies condition (1) but violates condition (2) when it is planar. In partially twisted ethylene, however, conditions (1), (2) and (3) are fulfilled. Therefore, SOC is expected to vanish for $\phi = 0$ and $\phi = 90^\circ$ and to have its maximum value for $\phi \approx 45^\circ$, as has been pointed out first by Caldwell *et al.* [29], and is apparent from Figure 3.

Pyramidalization affects SOC mainly through the in-plane component of the SOC vector, which increases particularly for small values of ϕ , since the increasing p_z part of the localized hybrid B is rotated by ℓ_y into the x direction and yields a large overlap with the localized hybrid A . In the case of C_{2h} pyramidalization, however, the hybrids A and B remain parallel and SOC becomes practically zero for $\phi = 0$, since the coupling through ℓ_y is not possible in this case.

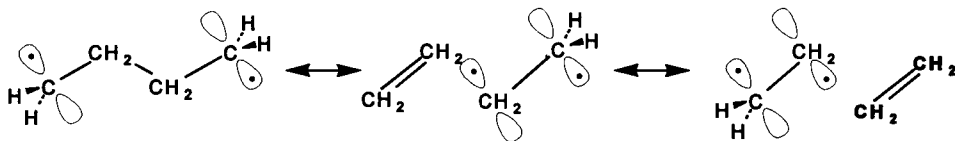
Trimethylene. If both terminal methylene groups are perpendicular to the molecular plane ($\alpha = \beta = 90^\circ$), the energies of A and B in trimethylene are equal, and condition (1) can be satisfied only in the presence of a non-zero resonance integral between them, which is provided by direct through-space interaction; through-bond coupling provides an additional opportunity for covalent interaction. Condition (2) is satisfied on both terminal carbon atoms. On one, A has a large amplitude on the in-plane p orbital of the radical center and B has some, albeit small, amplitude on the p orbital participating in the formation of the CC bond.

On the other terminal carbon, the roles of *A* and *B* are interchanged. The direction of the atomic vectorial contribution is along the direction perpendicular to the plane of the carbon atoms. If these out-of-plane vectors add, condition (3) will be satisfied; if they cancel, SOC will vanish. Working out the directions from the detailed expression for the vectorial contribution shows that the two contributions add, and SOC will not vanish. This conclusion can also be reached quite easily using symmetry arguments. (See Section 4.2).

In the planar conformation ($\alpha = \beta = 0^\circ$, both p orbital axes perpendicular to the CCC plane) condition (1) is satisfied, but condition (2) is not, since *A* and *B* are both of π symmetry and cannot each comprise a different p orbital on any other center. At all partially twisted symmetric conformations ($\alpha = \beta$ and $\alpha = -\beta$, corresponding to conrotatory and disrotatory motions, respectively) conditions (1), (2) and (3) are all satisfied. Among the less symmetrical configurations, the orthogonal geometry ($\alpha = 0$, $\beta = 90^\circ$) fails to satisfy condition (1), since the resonance integral vanishes for one orbital of π and the other of σ symmetry. At partial twist angles ($\alpha = 0$, $\beta \neq 0, 90^\circ$ or $\alpha \neq 0, 90^\circ$, $\beta = 90^\circ$), however, all three conditions are satisfied.

Summarizing these considerations, the expected dependence of the SOC values on rotations of the terminal methylene groups agrees very well with the "hat" shape of the calculated SOC surface shown in Figure 5.

Tetramethylene. With increasing chain length, the number of rotational degrees of freedom increases and it becomes more and more difficult to discuss the structural dependence of SOC. However, some qualitative results may be obtained by following Michl [11] and viewing tetramethylene as "ethanologous" triplet ethylene (15) for the purpose of understanding the delocalization of the radical electrons. In order for condition (2) to be fulfilled, the ethylene substructure must not be planar.



15

At the $90,90^\circ$ conformation, *anti*-tetramethylene ($\gamma = 180^\circ$) satisfies condition (1) and (2), but fails condition (3), and at the $0,0^\circ$ conformation, condition (1) is satisfied, but not condition (2). SOC therefore vanishes. At partially twisted conrotatory conformations with parallel *A* and *B* orbital axes, condition (3) is not met; in contrast, at partially twisted disrotatory conformations, all three conditions are satisfied. As in the case of trimethylene the orthogonal geometry ($\alpha = 0$, $\beta = 90^\circ$) does not satisfy condition (1), while at partial twist angles ($\alpha = 0$, $\beta \neq 0, 90^\circ$ or $\alpha \neq 0, 90^\circ$, $\beta = 90^\circ$) all three conditions are satisfied. Thus, the general features of the SOC surface in Figure 9c with two maxima and vanishing SOC along the edges and the diagonal corresponding to disrotatory twisting of the radical centers is easily verified.

4.2 Symmetry considerations

The use of group theory in the evaluation of SOC in molecules [6] in general and in biradicaloids [55] in particular has been common. The overall symmetry of the three possible triplet functions can be derived by considering the space part of T_1 and the three possible spin parts $\Theta_x, \Theta_y, \Theta_z$, which transform like the rotations R_x, R_y and R_z . Only those of the three triplet functions that belong to the same irreducible representation as S_0 can be mixed with it by the action of the totally symmetric spin-orbit coupling operators. For others, SOC vanishes.

Thus, it follows easily from symmetry, that the out-of-plane atomic vectorial contributions in the face-to-face conformation of trimethylene do not cancel but rather add: In the C_{2v} point group, the symmetry-adapted orbitals $A + B$ and $A - B$ belong to a_1 and b_2 , respectively, the space part of T_1 therefore belongs to B_2 , and the overall symmetry of the three components T_{1x}, T_{1y} and T_{1z} belong to A_1, A_2 and B_1 , respectively. Since S_0 transforms like A_1 , symmetry permits SOC between T_{1x} and S_0 , but SOC with the other two triplet sublevels vanishes [11].

Similarly, for the face-to-face conformation of *anti*-tetramethylene in the point group C_{2h} , T_{1x}, T_{1y} and T_{1z} transform like A_u, A_u and B_u , respectively and cannot couple with S_0 , which transforms according to A_g . Finally, at partially twisted disrotatory geometries of C_2 symmetry, T_{1z}, T_{1y} and T_{1x} transform like A, A and B , respectively, so the first two spin-orbit couple to \bar{S}_0 , which also transforms like A .

4.3 The "through-space" vector model

Although numerical calculations show that through-bond interactions normally dominate SOC in saturated biradicals [11], the results for ethylene and trimethylene indicate that both the through-space and the through-bond contributions exhibit a similar dependence on the rotation of the radical centers. This suggests the use of a "through-space" model [46] for a rationalization of this structural dependence, much in the spirit of the Salem-Rowland model [7] that was developed at a time when numerical results were not available.

For this purpose we introduce a space-fixed coordinate system in order to discuss the Cartesian components of the SOC vector, although the individual components may have no physical significance, unless the axes are unambiguously defined either by symmetry or by diagonalization of the spin dipole-dipole coupling tensor. Furthermore, for simplicity of the argument we will use the same coordinate frame for all systems to be discussed, independent of the conventional orientation of the coordinate axes according to the symmetry point group.

Based on eq. (28) the dependence of the components SOC_u on the orientation of the localized orbitals A and B will be discussed by representing the radical p AOs by unit vectors in spherical coordinates ϑ and φ and by approximating the coefficient $C_{0,+}$ by the overlap integral $\langle A|B \rangle$. According to Figure 11 the angular momentum and overlap integrals are then obtained by simple trigonometric arguments as

$$\langle l_x \rangle = \cos\omega [S_\sigma + S_\pi] \{\cos\vartheta\sin\vartheta' - \sin\vartheta\cos\vartheta'\}$$

$$\langle l_y \rangle = \sin\omega [2S_\pi] \{\cos\vartheta\sin\vartheta' + \sin\vartheta\cos\vartheta'\} \quad (30)$$

$$\langle l_z \rangle = -\sin 2\omega [S_\sigma + S_\pi] \{\sin\vartheta\sin\vartheta'\}$$

$$\langle A | B \rangle = [\cos^2 \omega S_\sigma - \sin^2 \omega S_\pi] \sin \vartheta \sin \vartheta' + S_\pi \cos \vartheta \cos \vartheta' \quad (31)$$

where ϑ and ϑ' are the angles (α, β) by which the orbitals are rotated in the plane perpendicular to the bond axis, which forms an angle $\varphi = 90^\circ - \omega$ and $\varphi' = 90^\circ + \omega$, respectively, with the x axis. 2ω is the angle between the two CC bonds of the radical centers, i.e. in ethylene, trimethylene and tetramethylene $2\omega = 180^\circ, 112^\circ$ and 24.5° , respectively.

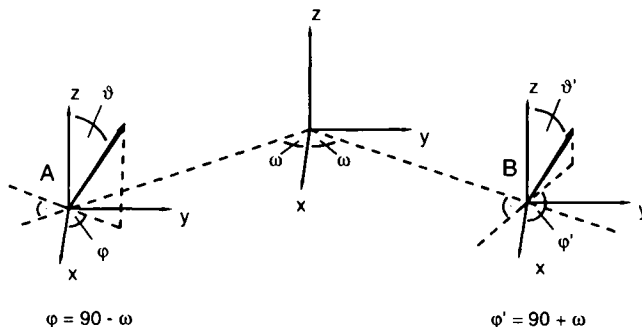


Figure 11. Spherical coordinates used in the "through-space" vector model. 2ω is the angle between the two CC bonds carrying the radical centers, and φ the angle between x axis and the rotational plane perpendicular to this CC axis.

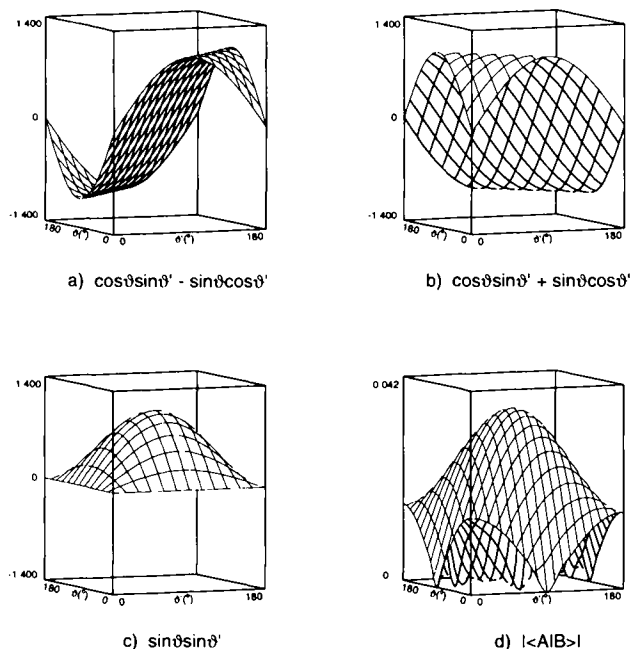


Figure 12. The "through-space" vector model. Cartesian components (a-c) of the orientational factors in the angular momentum integrals and (d) the overlap integral $\langle A | B \rangle$.

The angular momentum integrals depend on three factors: an orientation dependent factor given in braces, a distant dependent factor in brackets and a structural factor dependent on ω that describes the particular biradical. The orientation dependent factors are depicted in Figure 12a-c. The diagrams for the x and y components differ only by a rotation through 90° ; they are equal to zero along one of the diagonals, while the z component has its maximum at the center of the diagram. Figure 12d shows the absolute value of the overlap integral $\langle A|B\rangle$, which also assumes its maximum value for $\vartheta = \vartheta' = 90^\circ$.

Some general conclusions may be drawn from these results: If both methylene groups lie in the molecular plane ($[0,0]$ conformation, i.e. $\alpha = \beta = 0$), SOC is zero because all three components of the angular momentum vanish; for the $[0,90]$ conformation, however, with one radical center in the molecular plane and the other one perpendicular to the plane, SOC is zero because the overlap $\langle A|B\rangle$ vanishes. The x component of the angular momentum vanishes for conrotatory $[\vartheta, \vartheta]$ and the y component for disrotatory $[\vartheta, -\vartheta]$ motions of the radical centers. Both components are proportional to $\sin 2\vartheta$ and assume their maximum value for $\vartheta = 45^\circ$, while the z component is proportional to $\sin^2 \vartheta$ and therefore has its maximum at $\vartheta = 90^\circ$.

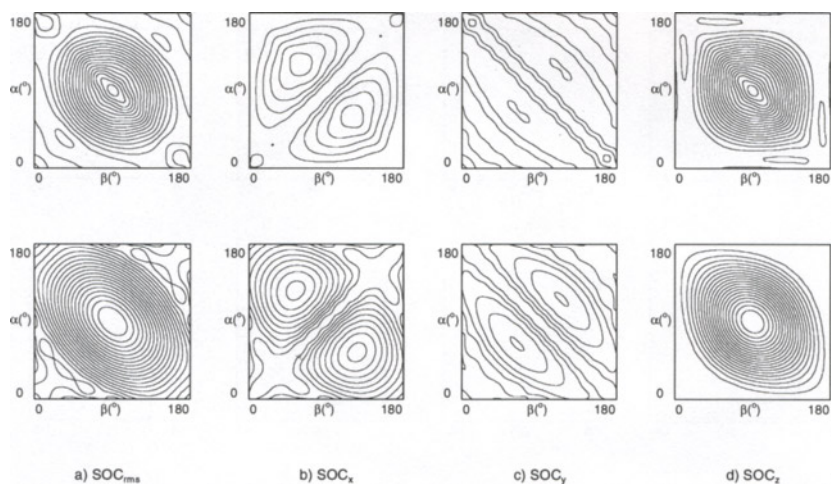


Figure 13. SOC surfaces for trimethylene: (a) total SOC value and (b-d) Cartesian components SOC_x , SOC_y , SOC_z , comparison of MNDOC-CI results (upper row) with the through-space vector model (lower row).

The difference between the various $1,n$ -biradicals is due to the first factor in eq. (30), the structural factor. For ethylene with $\omega = 90^\circ$, these factors become 0, 1 and 0 for the x , y and z component, respectively, and together with the $\sin 2\vartheta$ proportionality of the y component, this completely describes the computational SOC results shown in Figure 2b. For trimethylene ($\omega \approx 60^\circ$) the structural factors of all three component are nonzero; since the overlap integral $\langle A|B\rangle$ exhibits a similar dependence on the rotational angles as the z component of the angular momentum, while the x and y component are zero at the $[90,90]$ conformation

where $\langle A|B \rangle$ has its maximum, the z component dominates and yields the "hat"-like SOC surface of Figure 4. The detailed comparison of calculated data with results from the simple "through-space" vector model shown in Figure 13 emphasizes the surprisingly good performance of this model.

The situation is a little bit more complicated in the case of tetramethylene. Although the angular momentum components reflect the correct dependence on the rotational angles, their relative weights lead to an SOC surface that shows only a rough similarity with the computational results. Furthermore, due to the large distance of the radical centers the overlap integral $\langle A|B \rangle$ and therefore also SOC should vanish for *anti*-conformations, which is not the case. In this situation $\langle A|B \rangle$ is a very bad approximation for the hole-pair contribution $C_{0,+}$, and it is quite obvious that the covalent interaction required to produce a homosymmetric biradicaloid is due to through-bond interactions. If this as well as the 1,3-delocalization of the orbitals A and B is taken into account in an appropriate way [9, 46], good agreement with the calculated SOC surfaces is obtained also for tetramethylene, as can be seen from Figure 14. Thus, although in actual calculations the one-center terms dominate SOC, the "through-space" vector model mimicks the correct dependence of SOC on twisting of the radical centers and may be quite useful in rationalizing the structural dependence of SOC.

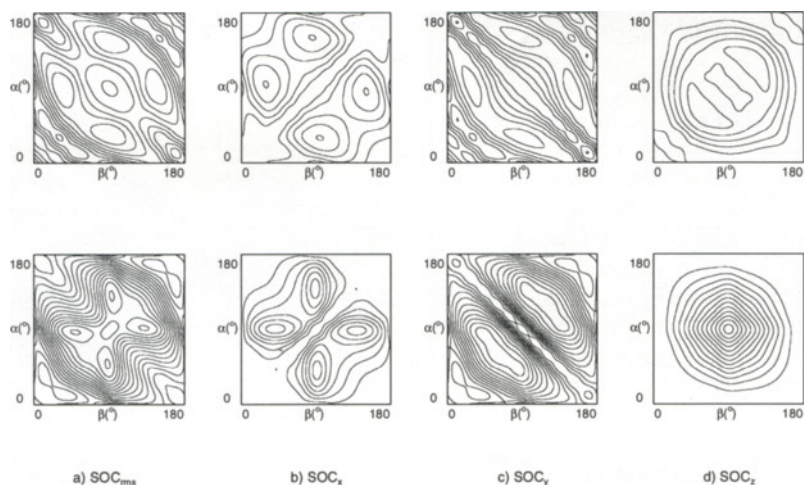


Figure 14. SOC surfaces for tetramethylene: (a) total SOC value and (b-d) Cartesian components SOC_x , SOC_y , SOC_z , comparison of MNDOC-CI results (upper row) with the through-space vector model (lower row).

5. CONCLUSIONS

Efficient methods for estimating SOC and its structural dependence in organic $1,n$ -biradicals, which have become available during the last decade, have improved our understanding of ISC and triplet photochemical reactivity in these systems to a large extent. The development occurred along three interrelated lines which very

much influenced each other: *ab initio* calculations of SOC effects that yield quantitative results for specific situations, semiempirical methods that allow for systematic searches of large areas of PE as well as SOC surfaces, and simple models that are particularly suited to estimate and rationalize the structural dependence of SOC effects. Especially useful is the 2-in-2 model of Michl [11], which in connection with even qualitative MO results allows for a discussion of SOC effects on the basis of atomic vector contributions, e.g. in order to discuss the heavy atom effect in systems like 2-oxatetramethylene.

From the results available so far it is quite clear that ISC is most likely at geometries that do not correspond to minima on the T_1 PE surface. Thus, thermal activation of one or a few kcal/mol is in general required to reach the conformations that are most favorable for ISC. For 1,2-dimethyltrimethylene it was shown that stereodifferentiation is due to the relative stability of different triplet conformations. Further calculations will have to show whether or not that is a general rule in triplet photochemistry. Semiempirical methods like those described here will be most useful for such investigations, because they allow for a fast determination of large areas of PE and SOC surfaces and for the treatment of fairly large systems and are therefore applicable to molecules of chemical interest including all kinds of substituent effects. Once these surfaces are known, it will be fairly easy to get more quantitative results on the basis of sophisticated *ab initio* methods. Thus, the scope of semiempirical methods in SOC calculations will be similar to their use in singlet photoreactions and thermal reactions [56].

In contrast to singlet photoreactions, where the return to the S_0 surface usually occurs through conical intersections of more or less well defined geometry [57], the geometries favorable for ISC generally encompass large areas of the T_1 surface. In order to obtain quantitative data for ISC rate constants it is therefore indispensable to study the dynamics of the T_1 - S_0 transition. Investigations of this kind have been carried out by Morita and Kato [45] for the Norrish Type II reaction, and the methods introduced by Robb *et al.* [58] for studying singlet photoreactions might prove useful for triplet reactions as well. Thus, it is expected that calculations on PE and SOC surfaces as well as of the dynamics on these surfaces will very much increase our understanding of triplet photoreactions in the near future.

ACKNOWLEDGEMENT

I would like to thank my coworkers listed in the references for their valuable contributions. The work was generously supported by Deutsche Forschungsgemeinschaft, Bonn (Kl 170/20).

REFERENCES

1. (a) N. J. Turro, *Modern Molecular Photochemistry*, Benjamin/Cummings, Menlo Park, CA, 1978. (b) J. Kopecký, *Organic Photochemistry*, VCH, New York, NY, 1992.

2. M. Klessinger and J. Michl, *Excited States and Photochemistry of Organic Molecules*, VCH, New York, NY, 1995.
3. (a) N. J. Turro and B. Kraeutler, In *Diradicals*; W. T. Borden (ed.), Wiley, New York, NY, 1982. (b) C. Doubleday Jr., N. J. Turro and J.-F. Wang, *Acc. Chem. Res.*, 22 (1989) 199.
4. G. L. Closs, M. D. E. Forbes and P. Piotrowiak, *J. Am. Chem. Soc.* 114 (1992) 3285.
5. (a) M. Bixon, J. Jortner, *J. Chem. Phys.* 48 (1968) 715. (b) K. F. Freed, *Acc. Chem. Res.*, 11 (1978), 74. (c) K. F. Freed, *Adv. Chem. Phys.*, 47 (1981) 291.
6. S. P. McGlynn, T. Azumi and M. Kinoshita, *Molecular Spectroscopy of the Triplet State*, Prentice-Hall, Englewood Cliffs, 1969.
7. L. Salem and C. Rowland, *Angew. Chem. Int. Ed. Engl.*, 11 (1972) 92.
8. This is due to the fact that triplet-singlet intersections usually occur at geometries other than minimum geometry; cf. the results reported here.
9. M. Böckmann, M. Klessinger and M. C. Zerner, *J. Phys. Chem.* 100 (1996) 10570.
10. M. Klessinger, T. Pötter and C. v. Wüllen, *Theor. Chim. Acta*, 80 (1991) 1.
11. J. Michl, *J. Am. Chem. Soc.* 118 (1996) 3568.
12. D. S. McClure, *J. Chem. Phys.*, 17 (1949) 905.
13. M. Böckmann, *Diplomarbeit*, Münster, 1991.
14. R. Pauncz, *Spineigenfunctions*, Plenum Press, New York, 1979.
15. B. T. Sutcliffe, *J. Chem. Phys.*, 45 (1966) 235.
16. R. Manne and M. C. Zerner, *Int. J. Quantum Chem. Symp.*, 19 (1986) 165.
17. A. Golębiewski and B. Broclawik, *Int. J. Quantum Chem.*, 27 (1985) 613.
18. R. L. Ellis, R. Squire, H. H. Jaffe, *J. Chem. Phys.* 55 (1971) 3499.
19. O. Vahtras, H. Ågren, P. Jørgensen, H. A. Jensen, T. Helgaker and J. Olsen, *J. Chem. Phys.* 96 (1992) 2119.
20. H. F. King and T. R. Furlani, *J. Comp. Chem.* 9 (1988) 771.
21. M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery Jr., *J. Comp. Chem.* 14 (1993) 1347.
22. (a) S. Koseki, M. W. Schmidt and M. S. Gordon, *J. Phys. Chem.* 96 (1992) 10768. (b) S. Koseki, M. S. Gordon, M. W. Schmidt and N. Matsunaga, *J. Phys. Chem.* 99 (1995) 12764.
23. S. R. Langhoff, C. W. Kern in *Modern Theoretical Chemistry*, Vol. 4, F. Schaefer III (ed.), Plenum, New York, N.Y., 1977.
24. (a) R. McWeeny, *J. Chem. Phys.* 34 (1961) 399. (b) R. S. Hutton, H. D. Roth, *J. Am. Chem. Soc.* 104 (1982) 7395.
25. J. B. Lounsbury, G. W. Barry, *J. Chem. Phys.* 44 (1966) 4367.
26. (a) W. T. Borden (ed.), *Diradicals*, Wiley, New York, N.Y., 1982. (b) M. S. Platz (ed.), *Kinetics and Spectroscopy of Carbenes and Diradicals*, Plenum, New York, N. Y. 1990.
27. H. R. W. McKellar, P. R. Bunker, T. J. Sears, K. M. Evenson, R. J. Saykally and S. R. Langhoff, *J. Chem. Phys.*, 79 (1983) 5251.
28. (a) F. Lehman, *J. Am. Chem. Soc.*, 24 (1976) 2623. (b) K. B. Eisenthal, N. J. Turro, E. V. Sitzmann, I. R. Gould, G. Hefferson, J. Langan and Y. Cha, *Tetrahedron*, 41 (1985) 1543.
29. R. A. Caldwell, L. Carlacci, C. E. Doubleday Jr., T. R. Furlani, H. F. King,

- and J. W. McIver Jr., *J. Am. Chem. Soc.*, 110 (1988) 6901.
30. B. F. Minaev, D. Jonsson, P. Norman, H. Ågren, *Chem. Phys.*, 194 (1995) 19.
 31. R. A. Caldwell, L. D. Jacobs, T. R. Furlani, E. A. Nalley and J. Laboy, *J. Am. Chem. Soc.*, 114 (1992) 1623.
 32. (a) C. Doubleday Jr., J. W. McIver Jr. and M. Page, *J. Am. Chem. Soc.*, 104 (1982) 6533. (b) A. H. Goldberg, D. A. Dougherty, *J. Am. Chem. Soc.*, 105 (1983) 284.
 33. T. R. Furlani and H. F. King, *J. Chem. Phys.*, 82 (1985) 5577.
 34. L. Carlucci, C. Doubleday Jr., T. R. Furlani, H. F. King and J. McIver Jr., *J. Am. Chem. Soc.*, 109 (1987) 5323.
 35. H. E. Zimmerman and G. A. Kutateladze, *J. Am. Chem. Soc.* 118 (1996) 249.
 36. A. M. Helms, R. A. Caldwell, *J. Am. Chem. Soc.* 117 (1995) 358.
 37. M. Böckmann and M. Klessinger, *Angew. Chem. Int. Ed. Engl.*, 35 (1996) 2502.
 38. R. Moore, A. Mishra and R. J. Crawford, *J. Canad. Chem.*, 46 (1968) 3305.
 39. (a) G. Segal, *J. Am. Chem. Soc.*, 96 (1974) 7892. (b) C. Doubleday Jr., J. W. McIver Jr. and M. Page, *J. Am. Chem. Soc.*, 104 (1982) 3768.
 40. R. Hoffmann, S. Swaminthan, B. Odell and R. Gleiter, *J. Am. Chem. Soc.*, 92 (1970) 7091.
 41. (a) C. Doubleday Jr., R. N. Camp, H. F. King, J. W. McIver Jr., D. Mullay, M. Page, *J. Am. Chem. Soc.* 106 (1984) 447. (b) C. Doubleday Jr., M. Page and J. W. McIver Jr., *J. Mol. Struct. (Theochem)*, 163 (1988) 331.
 42. (a) F. Bernardi, A. Bottoni, M. A. Robb, H. B. Schlegel and G. Tonacchini, *J. Am. Chem. Soc.*, 107 (1985) 2260. (b) F. Bernardi, A. Bottoni, P. Celani, M. Olivucci, M. A. Robb and A. Venturini, *Chem. Phys. Lett.*, 192 (1992) 229.
 43. C. Doubleday, J. McIver and M. Page, *J. Am. Chem. Soc.*, 107 (1985), 7904.
 44. H. E. Zimmerman, A. G. Kutateladze, Y. Maekawa and J. E. Mangette, *J. Am. Chem. Soc.* 116 (1994) 9795.
 45. A. Morita and B. Kato, *J. Phys. Chem.*, 97 (1993) 3298.
 46. M. Böckmann, Ph.D. Thesis, Münster, 1995.
 47. P. G. Schultz and P. B. Dervan, *J. Am. Chem. Soc.*, 104 (1982) 6660.
 48. P. D. Bartlett and N. A. Porter, *J. Am. Chem. Soc.*, 90 (1969) 5317.
 49. R. A. Caldwell, N. S. Dhawan and T. Majima, *J. Am. Chem. Soc.*, 106 (1984) 6454.
 50. J. Mählmann, Ph.D. Thesis, Münster, 1997.
 51. I. J. Palmer, I. N. Ragazos, F. Bernardi, M. Olivucci and M. A. Robb, *J. Am. Chem. Soc.*, 116 (1994) 2121.
 52. (a) J. Michl and V. Bonačić-Koutecký, *Electronic Aspects of Organic Chemistry*, Wiley, NY, 1990. (b) V. Bonačić-Koutecký, J. Koutecký and J. Michl, *Angew. Chem. Int. Ed. Engl.*, 26 (1987) 170.
 53. The Cartesian and the spherical components of the triplet state are related by the unitary transformations

$$\begin{aligned} T_x &= -[T_1 - T_{-1}]/\sqrt{2} \\ T_y &= i[T_1 + T_{-1}]/\sqrt{2} \\ T_z &= T_0. \end{aligned}$$
 54. (a) J. Michl, in *Theoretical and Computational Models for Organic Chemistry*, S. J. Formosinho, I. G. Csizmadia and L. G. Arnaut (eds.), Kluwer, Dordrecht, 1991. (b) J. Michl, *J. Mol. Struct. (Theochem)*, 260 (1992) 299.

55. S. Shaik and N. D. Epiotis, *J. Am. Chem. Soc.*, 102 (1980) 122.
56. (a) J. Dreyer and M. Klessinger, *J. Chem. Phys.* 101 (1994) 10655. (b) J. Dreyer and M. Klessinger, *Chem. Eur. J.* 2 (1996) 335.
57. M. Klessinger, *Angew. Chem. Int. Ed. Engl.* 34 (1995) 549 and references given therein.
58. (a) B. R. Smith, M. J. Bearpark, M. A. Robb, F. Bernardi and M. Olivucci, *Chem. Phys. Lett.* 242 (1995) 28. (b) M. J. Bearpark, F. Bernardi, S. Clifford, M. Olivucci, M. A. Robb, B. R. Smith and T. Vreven, *J. Am. Chem. Soc.* 118 (1996) 169. (c) M. J. Bearpark, F. Bernardi, M. Olivucci, M. A. Robb and B. R. Smith, *J. Am. Chem. Soc.* 118 (1996) 5254. (d) S. Clifford, M. J. Bearpark, F. Bernardi, M. Olivucci, M. A. Robb and B. R. Smith, *J. Am. Chem. Soc.* 118 (1996) 7353.

INDEX

- ab initio* calculations 15, 96, 100, 197,
 245, 247, 478, 583, 587, 592, 607
 absolute proton affinities 203
 acetone 52-54, 56, 57
 acetylcholine receptor 432-433, 438
 membrane structure prediction
 433, 444
 transmembrane α -helices and
 β -strands 432
 acetylene 502, 522, 523, 542, 546, 549,
 552, 555
 accuracy of prediction 405-406, 410,
 418-419, 427-430, 435-441
 acidity 73-76, 192, 194, 196, 197, 203,
 204, 205, 213, 229
 acids 135
 acridines 244
 activation barriers 513, 518, 524, 527,
 538, 545, 549, 553, 555, 558, 564,
 567, 569, 571, 573
 activation energies 146, 148, 149, 150
 adenine 247
 adiabatic force constant 271
 adiabatic frequencies, experimental 302
 adiabatic frequency 271
 adiabatic internal modes 250, 267, 282,
 302
 adiabatic mass 271
 adiabatic mass amplitude 277
 adiabatic mode, intensity 312
 AGIBA effect 153
 algebraic structure count 43, 44
 algorithm
 DSSP 408-409
 Eisenberg's 410, 424, 430, 440
 neural network, *see* neural network
 algorithm
 PREF 410-411, 417, 428,
 439-440
 Rao and Argos 430, 440
 SPLIT 410-411, 413, 418, 425,
 440
 and overtraining 427-429
 and sensitivity 427-429
 comparisons with other
 methods 429-434
 alternating sp^2/sp^3 -hybridized carbon
 atoms 390
 AM1 method 15, 244, 503*ff*
 AMBER modeling 481
 amines 60, 64, 65, 67, 69, 70
 amino acid
 attributes 406, 418-420, 426,
 434-439
 scales 406, 411, 419-422,
 424-428, 431-439
 aminotetralins 370
 amphipathic
 helices 410, 421, 438, 440
 β -strands 410, 433
 angular momentum integral 592, 600,
 603-606
 operator 586
 aniline 222
 animal experiments 448
 anisole 174, 210
 annulenes 42, 43
anti conformation 330
 antiaromaticity 502
 aromatic character 180
 aromatic heterocycles 234
 aromaticity 180, 502
 defect 211, 213, 228
 artificial diamond 386
 asymmetric environment 332
 atomic dipole moment 313
 atomistic modeling 329

- atomistic molecular modeling 350
- automated search strategies 341
- automatic docking 350
- average local ionization energy
 - atoms 190
 - molecules 191, 195
- average localization energy
 - definition 61, 62
 - minima on surface ($I_{s,\min}$) 62-65, 70-72, 74
- azafullerenes 395
- azaheterocycles 193
- azines 192
- azulene 242
- B matrix 265
- bacteriorhodopsin 434
- Badger's rule 260, 308
- Bakhshiev equation 248
- β -barrel 409, 411
- barrelene 596
- bases 135
- basicity 71, 72, 203, 204, 205, 212, 213
- basicity scales 203
- bay-region
 - carbocation 453, 456*ff*
 - diol epoxide 452
 - methyl effect 460, 485
 - theory 453*ff*
- Bent-Walsh rule 152, 177
- benzene 6-9, 12, 18, 157, 192, 514, 515, 518, 530, 532
 - CNM analysis 292, 293
- benzenium cation 208
- benzo[*b*]furan 537, 538
- benzo[*c*]furan 531, 537, 538
- benzoic acid 73-75
- benzonaphthazepine 363
- benzonitrile 210
- benzo[*c*]phenanthrene 469
- benzo[*a*]phenothiazines 244
- benzo[*e*]pyrene 453, 467, 486
- benzo[*b*]pyrrole 538
- benzo[*c*]pyrrole 531, 538
- benzo[*b*]thiophene 538
- benzo[*c*]thiophene 531, 538
- benzyne 513-518, 529
 - m*-benzyne, CNM analysis 288, 290, 293, 295, 296
 - o*-benzyne, CNM analysis 288, 289, 292, 295, 296
 - p*-benzyne, CNM analysis 288, 192, 294-296
- benzynes, adiabatic frequencies 296
- biradical 581-582, 587-588, 592, 594, 600-601, 603, 606
- biradicaloid 581, 587-588, 600-601, 603, 606
- 6,6-bis-(*p*-chlorophenyl)fulvene 240
- 2,5-bis-(trifluoromethyl)-1,3,4-oxadiazole 558, 559
- 2,5-bis-(trifluoromethyl)thiophene 505*ff*
- bond
 - angles 102, 105, 118
 - characterization 198
 - dissociation energies 77-81
 - energies 143, 145, 155
 - lengths 102, 105, 118, 155
 - localization 36, 45, 46
 - orders 35, 45, 503, 504, 505, 515, 521, 530, 531, 535, 541, 544, 548, 551, 557, 564, 566, 570, 572
 - strain 198
- bonded function 584-585
- Born-Oppenheimer approximation 8, 19
- Brookhaven Protein Databank 371
- buckminsterfullerene 391
- butadiene 102, 108
- butane 330
- c vector modes 266
- caffeine 247
- calcium channel 424
- capacity factor 333, 359, 361
- capsules - carbon 393, 395
- carbene 588-590, 601
- carbocations 123-127, 456

- carbon
 - allotropes 381
 - cages with heteroatoms 395
 - capsules 393, 395
 - fibers 382
 - molecules 381
 - nanotubes with heteroatoms 395
 - nets 381
- carbon monoxide 240
- carbonyl sulfide 240
- carbyne 391, 395
- carcinogenesis - chemical 447
- carcinogenic potency 447
- carcinogenicity indices 450, 453
- CASSCF calculation 588, 590-592, 599
- catalyzed reactions 148
- CC bond lengths 302
- cellulose triscarbamate polymer 356
- CH bond lengths 298
- chaoite 391, 395
- characterization of normal modes 273
- charge
 - flux 313, 315
 - transfer 54, 61, 62, 68, 70-72, 74, 86, 87, 196
- charged amino acids
 - and DC constants 412
 - and filtering procedure 413
 - CHARGE-BREAK subroutine 413, 418
- charges - effective 313
- chemical carcinogenesis 447
- chemical reactivity 192, 199
- chiral chromatography 329
- chiral ligand exchange chromatography 370
- chiral stationary phases
 - type I 335, 336, 348
 - type II 335, 354
 - type III 335, 363
 - type IV 336, 370
 - type V 336, 371
- chiral stationary systems 335
- chlorine molecule 58, 59
- Chou-Fasman type preferences 436, 439
- CNDO methods 15, 244, 247
- comparative molecular field analysis
 - of enantioselectivity 353
- compressibility - linear 381
- computer modeling of carcinogenesis 463, 477ff
- conductivity
 - electrical 381
 - thermal 381, 385
- configuration interaction 10, 13, 14, 18, 37, 38, 582, 588
- conformational analysis 354
- conformational calculations 478ff
- conformational isomers 330
- conformers 330
- conical intersection 10, 17, 607
- conjugated circuit theory 38
- conjugated circuits 383
- conrotatory motion 594-595, 602
- constitutional isomers 330
- contraction rule 586
- coronene 392
- correlation diagram 16
- Coulson/Fischer orbitals 12, 15
- coumarins 244
- crossing 16, 17
- cross-validation technique 414
- crown ethers 363
- curvature
 - coupling coefficient 319
 - decomposition 321
- cyanobenzene 210
- cycloaddition 501, 563
- cyclobutane 597
- cyclobutene 108, 109
- cyclodextrin 332, 363, 364, 366
- cyclopentadiene 502, 519
- m*-cyclophane 182
- cyclopropane 594
 - adiabatic frequencies 281
 - CNM analysis 281
- cyclopropene 102, 105, 508, 535, 537, 542, 546, 549, 552, 554, 555
- cytochrome c oxidase 425
- cytosine 247

- Debye equation 235, 236
 decision constants 411-412, 418, 426, 432, 438-439
 degeneracy 10, 17
 dehydrobenzene 513-518, 529
 density functional theory (DFT) 95, 125, 501
 densities - solid-state 248
 density of states 384, 390
 detoxification 455, 472
 deuterium iodide 240
 2,5-diaminothiophene 505
 diamond 381, 385
 artificial 386
 by conversion from graphite 387
 graphitization 389
 diamond-graphite hybrids 390
 diastereomeric carcinogens 452, 459, 461*ff*, 478
 diastereomers 330
 diazoles 540, 541, 563, 567, 568
 dibenzo[*a,l*]pyrene 470
 dielectric constants of solvents 238
 Diels-Alder reaction 101-107, 502, 539, 549, 563
 differential enthalpies 334
 differential entropies 334
 difluorodioxirane, CNM analysis 286, 287
p-dimethoxybenzene 175-177
 7,12-dimethylbenz[*a*]anthracene 468
 1,2-dimethylcyclopropane 596
 3,4-dimethyl-1-pyrazoline 596
 2,5-dimethylthiophene 505
 1,2-dimethyltrimethylene 588, 595, 607
 dinitrogen tetroxide 199
p-dinitrosobenzene 176, 177
 dioxirane, CNM analysis 286, 287
 6,6-diphenylfulvene 240
 dipole derivatives 313
 dipole meter 237, 238
 dipole moment
 acridines 244
 azulene 242
 benzo[*a*]phenothiazines 244
 dipole moment
 coumarins 244
 definition 233
 direction 239
 fulvene 240, 242
 furan 234
 indoles 244
 merocyanine 540 244
 1-methylpyrrole 241
 phenazines 244
 phenothiazines 244
 pteridines 244
 purines 245, 247
 pyridine 242, 243
 pyrimidines 245, 247
 pyrrole 234, 241
 pyrrolidine 234
 quinazolines 250
 selenophene 234
 sign 239
 thiophene 234
 tetrahydrofuran 234
 tetrahydroselenophene 234
 tetrahydrothiophene 234
 triazenes 244
 dipole moments
 applications 234
 calculated 241, 247, 249, 250
 excited-state 245-247, 249, 250
 experimental 235-239, 245-247, 250
 ground-state 235-239, 241, 247, 250
 directed conformational change 481
 direction of dipole moment 239
 disrotatory motion 594, 602, 605
 dissociation energy 305, 307
 distribution coefficient 333
 DNA binding of carcinogens 461*ff*
 DNA catalyzed hydrolysis 476*ff*
 DNA-drug intercalation 344
 docking/minimization strategy 345
 donor-acceptor complexes 160
 dual graph 393
 effective charges 313

- electrical conductivity 381
- electron affinity 189
- electron pair 2, 8, 11
- electronegativity 135, 189, 190
- electronic density function 190, 191
- electronic energy 137
- electrophilic attack 138, 140, 192
- electrophilic aromatic substitution 203, 213, 228
- electrophilic superdelocalizabilities 352
- electrostatic potential
 - balance parameter 83
 - definition 51, 193
 - general 51-88
 - maxima on surface ($V_{s,max}$) 57, 58, 68, 69
 - measure of local polarity 82
 - minima on surface ($V_{s,min}$) 57, 67, 68
 - on molecular surface 55-58
 - polarization correction 60-61
 - polarization correction to spatial minima ($P_{v,min}$) 61-65, 70-72, 74
 - spatial minima (V_{min}) 52-57, 59, 61-65, 68, 70-72, 74-79, 85, 87, 88
 - variability 73, 83
- enantiomers 330, 333
- enantioselective binding 336, 346
- enantioselectivity 329, 340, 347, 353
- energies - activation 146, 148, 149, 150
- energy and hardness differences 140
- energy partitioning 343
- enthalpies
 - differential 334
 - protonation 196
- entropies - differential 334
- epidemiology 448ff
- ethanologous triplet ethylene 602
- ethylene 502, 517, 518, 542, 546, 549, 552, 555, 588-591, 593, 601, 602, 604-605
- ethylene, adiabatic frequencies 306
- Euler-Lagrange equations 264, 265, 274, 324
- exchange-repulsion energy 54, 56-58
- excited-state dipole moments
 - calculated 249
 - electrooptical methods 246
 - experimental 245, 247, 250
 - solvatochromic methods 246
 - solvent-shift methods 246
- face-to-edge conformation 598
- face-to-face conformation 592-594, 596, 606
- fenestrane sheet 386
- Fermi's golden rule 581
- filtering procedure 412-413
- fjord region 459, 469
- fluorine 192
- fluoroacetic acid 192
- fluorobenzene 213
- fluorobenzenes 219-221
- force constant - adiabatic 271
- fracture zones 382
- fragment modes 263
- frequencies - adiabatic 271, 302
- frequency-bond length relationships 302
- frequency shifts 64, 71
- frontier molecular orbital energies 103, 106, 114, 122, 143, 504, 508, 516, 517, 519, 523, 533, 534, 540, 543, 547, 550, 553, 556, 568
- frontier molecular orbital theory 135
- fukui function 135, 144-149
- fullerenes 391
 - isolated-pentagon 392
 - proper 392
- fullerynes 398
- fullerocoronand 395
- fulvene 162, 240, 242
- furan 105, 234, 502, 513, 516, 517
- gauche* conformation 330
- Gaussian theoretical models 96
- genotoxic vs. epigenetic carcinogens 449
- Gibbs-Helmholtz equation 333
- global interaction indices 81-88
- graph theory 34, 35, 38, 39, 43, 44
- graphene plane 381

- graphite 381
 graphitic cones 384
 graphitization of diamond 389
 grid search strategy 339, 344, 349, 357, 358
 ground-state dipole moments - calculated
 241, 245, 247, 250
 ab initio methods 245, 247
 empirical methods 241
 semiempirical methods 245
 ground-state dipole moments - experimental
 235
 dielectric constant methods 235
 electric resonance methods 239
 microwave methods 238
 molecular beam method 239
 Raman spectroscopy 239
 Stark effect method 239
 group V-VII hydrides 196
 guanine 247
 guest-host complex 333
 Guggenheim equation 237
 Halverstadt-Kumler method 236
 HAM3 method 244
 Hamiltonian circuit 393
 Hammett constants 192, 198
 Hammond postulate 97
 hardness 135
 hardness and softness theory 135
 heats of formation 155, 156
 heavy atom effect 588, 590, 600
 Hedestrand equation 236, 237
 Heisenberg model 8, 34, 37
 α -helix
 configuration 418, 423, 439
 conformation 406, 408, 414-415, 427, 434
 preferences 413-414, 419, 425
 heptafulvene 162
 Herndon-Simpson model 37, 38
 heterocycles - aromatic 234
 heterosymmetric perturbation 601
 HMO method 241
 holes bordered by heteroatoms 386
 HOMA index 153, 176, 180, 181, 182
 homosymmetric biradicaloid 601, 606
 HOSE model 153, 166
 Hubbard model 36-40, 42, 43
 Hückel model 4-6, 9, 13-15, 35, 36, 40, 41, 44, 241
 Hückel rule 41, 42
 hybridization 2, 11, 12, 211, 213, 229
 hydride affinity 123
 hydrogen bonding 53, 57, 58, 65-71, 75, 84-87, 161, 163-165, 479, 484
 hydrogen chloride 239
 hydrogen fluoride 53, 54, 58, 59, 239
 hydrolysis 459, 472, 475, 476
 hydrophobic moment 410-411, 414, 422, 431-434, 438-439
 hydrophobicity
 plots 405
 scales 420-421, 432, 434-438
 1-hydroxytetramethylene 596
 hyperfine coupling (HFC) 581
 Iball index 450, 458ff
 inclusion complexes 363
 independent substituent model 214, 223
 INDO methods 15, 244
 indoles 244
 infrared intensity 312
 intercalation 462, 466, 471
 intercalative model 354
 intermolecular hydrogen bonding 161
 internal coordinates 260, 261, 263, 273
 internal modes 260, 261, 262, 266, 273
 internal modes - adiabatic 250, 267, 282, 302
 intersystem crossing 581, 588-589, 594, 596-597, 606
 intramolecular radical addition 119
 intrinsic activity 460
 invariants 393
 inverse superatom 399
 ion-pair chromatography 370
 ionization energy 189, 190, 191, 195
 ipso protonation 217
 ISA model 203

- isolated CH stretching frequency 298
 isolated-pentagon fullerenes 392
 Jahn-Teller effect 10
 K,L,M theory 451
 Kahn-Prelog-Ingold stereochemical descriptors 334
 Kawski-Chamma-Viallet equation 248, 250
 Kekulé structures 36, 38, 41-44, 174
 kinetic energy 11
 Koopmans' theorem 190, 197
 Kyte-Doolittle
 algorithm 434
 and preference function 422
 hydropathy scale 406, 413-414, 420, 422, 424, 428-434, 438-441
 modified scale 420, 424, 438
 lattice, local defects
 diamond 389
 graphite 387
 LCAO approximation 583, 586
 leading parameter principle 261, 269
 length distribution of transmembrane segments 416-418
 ligand exchange chromatography 370
 light harvesting center 430
 linear compressibility 381
 linear solvation energy relationships (LSER) 66-68, 84-85
 lipophilicity parameter 362
 local defects in diamond lattice 389
 local defects in graphite lattice 387
 local ionization energy 190, 191, 195
 localization energy 61, 62-65, 70-72, 74
 long-range effects of carcinogens 467
 malonohydrazides 567
 mass - adiabatic 271, 277
 McKean correlation 286, 297, 298, 299
 McRae equation 248
 MCS model of cancer initiation 454ff, 487
 MCSCF calculation 587, 596
 mechanism-based structure-carcinogenicity relationships (MSCR) 449, 459
 6-mercaptopurine 247
 2-mercapto-4(3*H*)-quinazolinone 250
 merocyanine 540 244
meso form 331
 mesomerism 2, 8
 metabolic activation 451ff
 metabolic factor 455ff
 methanol 57, 71
p-methoxyanisole 175, 176, 177
 5-methylchrysene 470
 2-methylmercapto-4(3*H*)-quinazolinone 250
 1-methylpyrrole 241
 2-methyl-4(3*H*)-quinazolinone 250
S-methylthiophenium cation 504, 506ff
 MINDO/1,2,3 methods 244
 MNDO method 15, 244
 MNDOC-CI method 582, 588-593, 597-599, 605-606
 mode - adiabatic, intensity 312
 mode amplitude
 adiabatic 277
 force comparison 278
 mass comparison 278
 modeling enantioselective binding 336
 modeling - molecular 335, 367
 molecular dynamics
 calculations 597, 607
 simulations 365, 367
 molecular geometry 102, 126, 153ff
 molecular mechanics 367
 molecular modeling 335, 367
 molecular orbital theory 2-4, 9, 11-16, 33-47
 molecular surfaces 56, 191, 195, 196
 π -moment 241, 244
 σ -moment 241, 244
 moments method 384
 motif-based searching 337, 345
 Mulliken charges 313
 multiple alignment procedure 436, 440
 nanotubes - carbon 393
 naphthalenes 182, 183
 nested cages 399
 neural network algorithm
 accuracy 429-430

- neural network algorithm
 and secondary structure prediction
 405, 436
 computer time 439
 training process 429
 neutron capture 382
- neutron diffraction 367
- p*-nitroaniline 170
- nitrobenzene 169, 172, 195, 196
- nitrogen-containing heterocycles 193
- nitrosobenzene 211
- p*-nitrosophenolate anion 161
- norbomadiene 592
- Norinder-Hermansson models 373
- normal mode analysis 260
- normal modes 259, 260, 262, 273
- Norrish type II reaction 596, 598, 607
- nucleophilic attack 138, 139, 140
- nucleophilic substitution 99
- nucleophilic superdelocalizabilities 352
- octanol/water partition coefficients 83-87
- Onsager cavity radius 248
- overlap integral 603-606
- overlap of negative centers 482
- 1,3,4-oxadiazole 558
- 2-oxatetramethylene 588, 599, 607
- packing energies 354
- partition coefficients 83-87, 333
- Paterno-Büchi reaction 588
- Pauling-Wheland models 36-39, 44
- PED analysis 266
- Peierls distortion 44, 45, 46
- perfect biradical 601
- performance parameters 406, 409-411,
 418, 420, 427-430
- perylene 157
- pH-contour map of DNA 476
- phenanthrene 157
- phenanthrene triol carbocation 478*ff*
- phenol 209
- phenol-base complexation enthalpies 63,
 64, 69, 70, 87
- phenols 73, 76-81
- phenoxy radicals 77-81
- photosynthetic reaction center 409,
 430, 434-438
- physical complex DNA-carcinogen 463
- pillow 392, 393
- pK_a values - predicted 194
- PM3 method 15
- PMO theory 453*ff*, 457
- polar stabilization energies 79-81, 87
- polarity 70, 82, 84, 86, 87
- polarizability 189, 190, 191
- polarization 52, 54, 56, 60, 61, 65, 68,
 70-72, 74, 87, 196
- polyazafullerenes 395
- polybenzene 390
- polymers 44
- porin 406, 409, 431-432, 435-440
 from *Rhodobacter capsulatus*
 409, 431
- potential energy 12, 16-20
- potential energy surface (PES) 211,
 582, 588-591, 595-599
- PPP charge dispersal energy 456*ff*
- PPP method 14, 36-40, 42, 244, 247,
 250
- predicted pK_a values 194
- prediction
 in G-protein coupled receptors
 431
 in large eukaryotic proteins 430,
 431
 in membrane import machinery
 protein 433-434
 in mitochondrial carrier family
 433
 in nicotinic acetylcholine receptor
 431-433
 of transmembrane helices 406,
 412, 420, 429, 433, 436-439
 of transmembrane segments
 405-406, 414, 429-430, 434-437
 of transmembrane β -strands
 413-414, 434-440
 performance of 12 best scales
 426-427

- prediction
 - results in the SPLIT algorithm 418-419
 - tonb_ecoli protein 434
- preference functions
 - and extraction from data base 417
 - and homology 429
 - and Kyte-Doolittle scale 422
 - and overtraining 429
 - and representation of context dependence 434
 - best training procedure 423, 427
 - definition 406
 - evaluation 412, 414
 - for leucine and glycine 415
 - method and prediction 410-412
 - standard training procedure 411, 415, 426-428
 - testing procedure 411, 420, 428-429, 438
- propellane 198
- proper fullerenes 392
- propylene 590
- protein data bank
 - PDB 409
 - SWISS-PROT: *see* SWISS-PROT data base
- protein data base
 - BESTP 409
 - for false positive results 409
 - for training and testing 407-409
 - of β -class soluble proteins 409
 - PORINS 409
- protein phases 371
- proteins
 - integral membrane 405-411, 414, 417, 424
 - of well known structure 424-425
 - soluble 405-409, 414, 417
 - for false positive predictions 424-425
 - of β -class 425, 427, 437
- proton affinities 53, 63, 64, 71, 72, 87, 203, 204, 206, 211, 223, 472
- proton affinities
 - absolute 203
 - additivity 211, 215, 217, 225, 228
- proton transfer 203
- protonated diol epoxides 477
- protonation 52-54, 57, 58, 63-64, 71, 72, 74, 87
- protonation enthalpies 196
- protonation - ipso 217
- pteridines 244
- purines 245, 247
- pyracylene 393
 - automerization 394
- pyramidalization 590-594, 601
- pyrene 157
- pyridine 196, 242, 243
- pyrimidines 245, 247
- pyrrole 234, 241, 502, 513, 516, 518, 519, 521, 524*ff*
- pyrrolidine 234
- quantitative structure-activity relationships 352
- quantum mechanics 2-4, 8, 12, 18
- 2,4(*1H,3H*)-quinazolinone 250
- quinazolines 250
- 4(*3H*)-quinazolinone 250
- racemic mixture 330, 332
- radical addition - intramolecular 119
- radical attack 139, 140
- radical reactions 117-123
- radical stabilization energies 79-81, 87
- radicals 43, 77-81
- rate constants 99
- reaction barriers 100, 104, 107, 110, 112, 119, 121
- reaction energies 147
- reaction energy profile 97, 98, 101
- reaction mechanism 316, 318
- reaction path
 - curvature 317, 319
 - direction 316
- reaction valley 319
- reactions in solution 148
- reactivity parameters 136

- reactivity-selectivity 95, 96, 99
reduced dual 393
refractive indices of solvents 238
regioselectivity 452
resonance 2, 4, 7-9, 11, 15, 36, 37, 41, 46, 234
retention time 333
ring-closure reactions
 tetramethylene 597-598
 trimethylene 594-595
ring-energy content 157ff
ring-opening reactions 108-117
Rumer diagram 585
Salem-Rowland rules 587-588, 600-601, 603
sausage 392
SCF-MO method 13
second quantization 584
secondary orbital interactions 509, 510
secondary structure prediction 405, 434
selenophene 234
semempirical calculations 582, 587, 588, 607
semiregular planar lattice 383
separation factor 334, 359
shape selectivity 454, 460, 485
 β -sheet 406, 414, 423, 434, 437-438
 preferences 431-433
short-term test for carcinogenicity 448
sign of dipole moment 239
signal sequence 407-408, 426, 436-439
singlet 249, 250
singlet photoreactions 596, 607
singlet-triplet intersection 589, 594-599
singlet-triplet splitting (E_{ST}) 581, 590, 592, 596, 597
site-specific interactions 65-72
size factor 458
sliding window 411-412
 length 417, 428, 429
 method 405-406
Smith equation 237
smoothing procedure 412, 418
softness 135
solid-state densities 248
solitons 44, 45, 46
solubilities in supercritical fluids 83
spacer linkages 346
sphalerite 385
spin symmetry 36, 42, 43, 46-orbit
spin-orbit coupling (SOC) 581-607
 constant 584, 586, 600
 operator 581, 583-584, 603
 strength 583
 surface 582, 588, 589-599, 605-606
 value 589, 591-594, 595, 598-599
 vector 583, 600-601
spin-spin dipolar coupling 587, 603
spiral code 393
stabilization energies
 polar 79-81, 87
 radical 79-81, 87
statistically averaged interaction energy 342
statistically base interaction indices 81-88
stereochemical differentiation 343, 596, 607
stereodifferentiation 343
stereoisomers 330
stereoselectivity 454, 463, 484, 485, 588
steric effects 595
steric incompatibility 484
stilbene oxide 356, 357
Stone-Wales rearrangement 393
strain energies 126
strain-free oligoradical 384
 β -strand preferences 414, 437, 439
 β -strands
 membrane attached, *see* surface-attached β -strands
 membrane-embedded 406, 414, 422-423, 432-433, 438-439
 surface-attached 414, 432-433
stretching frequency, CH - isolated 298
structural factor 605

- structure count - algebraic 43, 44
 substituent constants 73-77, 80
 substituent effects
 acidities 73-76, 87
 bond dissociation energies 77-81, 87
 HOSE model 168
 importance of solvation 74-76
 molecular geometry 177
 substituted benzenes 206, 207, 211, 214-219
 substituted naphthalenes 223, 226, 227
 sulfoxides 352
 superconductor 393
 supercubane 391
 superdelocalizabilities
 electrophilic 352
 nucleophilic 352
 Suppan equation 248
 surfaces - molecular 56, 191, 195, 196
 SWISS-PROT sequence data base
 assignment 405-408, 4214, 430-431
 errors in assignment 437
 reference standard 433
 topological model 424
 symmetry 10, 15-19
 symmetry consideration 602-603
 Szentpály-Shamovsky model 487
 Taft constants 192, 196
 tautomerism 8
 tetrahydrofuran 234
 tetrahydrosephenone 234
 tetrahydrothiophene 234
 tetramethylene 588, 596-599, 602-604, 606
 thalidomide 332
 thermal conductivity 381, 385
 1,3,4-thiadiazole 571
 thiazoles 540, 541
 thiophene 234, 502, 513, 516-518
 thiophene-1-oxide 504, 506, 507
 thiophene-1,1-dioxide 504, 506, 507
 2-thiouracil 247
 three-stage model of carcinogenesis 449
 through-bond coupling 592, 597, 601, 606
 through-space coupling 592, 601, 603
 through-space vector model 603-606
 thymine 247
 topological invariants 393
 topological stereoisomers 395
 torus 395
 transition state 96ff, 149
 trapping of triol carbocation 474ff
 triazines 244
 tribenzopnenthrapentaphene 158
 1,3,4-triazole 569
 trichloromethyl radical 117
 1,3,5-tridiazabenzene 175, 176
 2,2,2-trifluoro-1-(9-anthryl)ethanol 339
 2,4,6-trimethoxy-*s*-triazine 174
 trimethylene 588, 592-595, 601, 603-605
 triphenylene 157
 triplet 6, 10, 251
 triplet photoreactions 581, 606-607
 trishomotropenylium cation, adiabatic frequency 302, 303
 truncated octahedron 386
 turn 406, 408
 conformation 406, 408, 438
 preference 413-414, 431, 437
 preference in filtering procedure 413
 two-electron two-orbital model 600-601
 uncertainty test of mode frequencies 302, 305
 unified reaction valley analysis 319, 321
 uracil 247
 valence bond theory 4, 6, 9, 11, 12, 14, 16, 33-47
 van der Waals
 complexes 58, 59
 contacts 486
 van't Hoff plots 334
 vibrational spectra, correlation 286
 vibrational spectroscopy 259, 263
 vinyl chloride 590

virial charges 313
weak complexes 58, 59
Wigner effect 382
Wilson's GF formalism 266

Wilson's G-matrix 265
wurtzite 386
ZDO approximation 586-587
Zeeman splitting effects 240