LEVEL

INVENTORY

AD-A194 735

DTIC ACCESSION NUMBER

AFOSR·TR·88·0685

DOCUMENT IDENTIFICATION

14 MAR 87

This document has been approved
for public release and sale; its
distribution is unlimited.

DISTRIBUTION STATEMENT

ACCESSION FOR	
NTIS GRA&I	☒
DTIC TAB	☐
UNANNOUNCED	☐
JUSTIFICATION	

SOLD BY

BY
DISTRIBUTION /
AVAILABILITY CODES

DIST	AVAIL AND/OR SPECIAL

A-1 *21*

DISTRIBUTION STAMP

JOHN WILEY & SONS INC
605 THIRD AVE
NEW YORK, NY 10158

DTIC
SELECTED
JUL 0 5 1988
E

DATE ACCESSIONED

DTIC
COPY
INSPECTED

DATE RETURNED

88 6 29 148

DATE RECEIVED IN DTIC

REGISTERED OR CERTIFIED NO.

DTIC FORM 70A
MAR 86

DOCUMENT PROCESSING SHEET

PREVIOUS EDITION MAY BE USED UNTIL
STOCK IS EXHAUSTED.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR·TR· 88-0685

6a. NAME OF PERFORMING ORGANIZATION University of Florida	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NC

6c. ADDRESS (City, State, and ZIP Code) Gainesville, Florida 32611	7b. ADDRESS (City, State, and ZIP Code) Bldg 410 Bolling AFB DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NC	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR_87-0111

8c. ADDRESS (City, State, and ZIP Code) Bldg 410 Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2303	TASK NO. B3	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

1987 Sanibel Symposia — Proceeding of the International Symposium on Quantum Biology and Quantum Pharmacology #14

12. PERSONAL AUTHOR(S)
Per-Olov Lowdin

13a. TYPE OF REPORT FINAL	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The 27th Annual Sanibel Symposia, which included the 14th meeting of the Symposium on Quantum Biology and Quantum Pharmacology, was held on March 12-14, 1987 at the University of Florida Marine Biology Laboratory at Marineland on the Atlantic Coast of Florida. These proceedings comprise contributions of both invited talks and poster presentations. Submitted papers were subjected to the ordinary refereeing procedures of the International Journal of Quantum Chemistry. Although some of the papers fall outside the ordinary scope of quantum chemistry they reflect the content of the symposium.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Major Larry P. Davis	22b. TELEPHONE (Include Area Code) (202) 767-4960	22c. OFFICE SYMBOL NC

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted.   SECURITY CLASSIFICATION OF THIS PAGE
All other editions are obsolete.

International Journal of QUANTUM CHEMISTRY

Quantum Biology Symposium No. 14

Proceedings of the
International Symposium on

Quantum Biology
and Quantum Pharmacology

Held at Marineland, Florida, March 12–14, 1987

Editor-in-Chief: Per-Olov Löwdin
Special Editors: N. Yngve Öhrn, John R. Sabin and
Michael C. Zerner

Contents

Introduction

The 27th annual Sanibel Symposia, which included the 14th meeting of the Symposium on Quantum Biology and Quantum Pharmacology, was held on March 12–14, 1987 at the University of Florida Whitney Marine Biology Laboratory at Marineland on the Atlantic Coast of Florida.

These proceedings comprise contributions of both invited talks and poster presentations. Submitted papers were subjected to the ordinary refereeing procedures of the *International Journal of Quantum Chemistry*. Although some of the papers fall outside the ordinary scope of quantum chemistry they reflect the content of the symposium.

The organizers acknowledge support for the symposia from:

<div align="right">
P. O. Löwdin
N. Y. Öhrn
J. R. Sabin
M. C. Zerner
</div>

List of Participants

JOSE ALVES
Depto de Fisica, ICEx-UFMG
Cidade Universitaria 30000
Belo Horiozonte, Brazil

RAYMOND E. ANDREW
Department of Physics
223 Williamson Hall
University of Florida
Gainesville, FL 32611

GUSTAVO ARTECA
Department of Chemistry
Univ. of Saskatchewan
Saskatoon, Canada S7N 0W0

RODNEY J. BARTLETT
Quantum Theory Project
381 Williamson Hall
University of Florida
Gainesville, FL 32611

GOKHAN BAYKUT
Department of Chemistry
Leigh Hall
University of Florida
Gainesville, FL 32611

DAVID L. BEVERIDGE
Chemistry Department
Wesleyan University
Middletown, CT 06457

ERKKI BRANDAS
Dept. of Quantum Chemistry
Box 518, S-751 20
Uppsala, Sweden

RICHARD BRANDT
Dept. of Biochemistry
Med. College of VA/VCU
Richmond, VA 23219

PAUL CHUN
J-245, Biochem. & Molec. Bio.
College of Medicine
University of Florida
Gainesville, FL 32601-0245

PIOTR CIEPLAK
University of California
Dept. of Pharmaceutical Chem. S-926
San Francisco, CA 94943

MICHAEL CONRAD
Dept. of Computer Science
520 Mackenzie Hall
Wayne State University
Detroit, MI 48202

ROBIN DAVIES
Welsh School of Pharmacy, UWIST
P.O. Box 13
Cardiff, CF1
Wales, UK

JANET DEL BENE
Youngstown State University
Department of Chemistry
Youngstown, OH 44555

JOSEPH DELHALLE
Fac. Notre-Dame le la Paix
Dept. de Chimie
61, rue de Bruxelles
B-5000 Namur, Belgium

MICHAEL DUPUIS
IBM Corporation
Dept. 48B/428
Neighborhood Road
Kingston, NY 14201

ABBAS FARAZDEL
Amherst College
Department of Chemistry
Amherst, MA 01002

GEORGE FORD
Southern Methodist University
Department of Chemistry
Dallas, TX 75275

SIDNEY FOX
University of Miami
Institute for Molecular & Cellular
 Evolution
521 Anastasia
Coral Gables, FL 33134

MAXIMO GARCIA-SUCRE
IVIC, Centro de Fisica
Apartado 1827
Caracas 1010A, Venezuela

IVAR GIAEVER
GE Research & Development Center
Bldg. K1-3C36
Schenectady, NY 12301

MARVIN GOODGAME
Eastman Kodak Company
Kodak Research Labs, Bldg. B2
1999 Lake Avenue
Rochester, NY 14650

ANTE GRAOVAC
R. Boskovic Institute
IRB, YU-41001
P.O.B. 1016
Zagreb, Yugoslavia

HAROLD HANSON
U.S. House of Representatives
Committee on Space,
Science and Technology
Washington, D.C. 20515

BARRY HONIG
Columbia University
Department of Biochemistry
and Molecular Physics
630 W. 168th Street
New York, NY 10032

JOHN HURD
Floating Point Systems
3601 Murray Blvd.
Beverton, OR 97005

WILLIAM T. HUTCHENS
Baylor College of Medicine
Reproductive Research Lab.
5510 Greenbriar
Houston, TX 77005

AKIRA IMAMURA
Hiroshima University
Department of Chemistry
Higashisenda-machi 1-1-89, Naku-ku
Hiroshima-city, 730, Japan

ANDRZEJ JAWORSKI
University of Florida
Leigh Hall, Box 354
Gainesville, FL 32611

MU SHIK JHON
Korea Advanced Institute of
Science & Technology
P.O. Box 150
Cheong Ryang Ri
Seoul, Korea

WILLIAM JORGENSEN
Purdue University
Department of Chemistry
W. Lafayette, IN 47907

ISABELLA KARLE
Naval Research Lab.
Code 6030
Washington, D.C. 20375

JEROME KARLE
Naval Research Lab.
Code 6030
Washington, D.C. 20375

JOYCE KAUFMAN
The Johns Hopkins University
34th & Charles Streets
Baltimore, MD 21218

LEO KLASINC
The Rudjer Boskovic Institute
Bijenicka Cesta 54
YU-41001 Zagreb, Yugoslavia

VALENTINE J. KLIMKOWSKI
Polygen Corporation
200 Fifth Avenue
Waltham, MA 02154

CARLOS KUBLI-GARFIAS
Instituto Mexicano del Seguro Social
Costado de Atrio
Coyoacan Mexico, D.F. 04000 Mexico

JURAJ KUMICAK
Uppsala University, Box 518
Department of Quantum Chemistry
S-751 20 Uppsala, Sweden

JOSEF KWIATKOWSKI
N. Copernicus University
Institute of Physics
87-100 Torun, Poland

JANOS LADIK
University Erlangen-Nurnberg
Egerlandstr. 3
D-8520 Erlangen, West Germany

PIERRE LEBRETON
University of Illinois-Chicago
Depart. of Chemistry, P.O. Box 4348
Chicago, IL 60680

GILDA LOEW
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

PER-OLOV LOWDIN
University of Florida
QTP, Williamson Hall
Gainesville, FL 32611

BILL PARKINSON
University of Florida
QTP, Williamson Hall
Gainesville, FL 32611

MICHAEL T. MARRON
Program Manager of Molecular Biology
Office of the Chief of Naval Research
Arlington, VA 22217-5000

ZAIDA PARRA
University of Florida
QTP, Williamson Hall
Gainesville, FL 32611

LOU MASSA
CUNY
695 Park Avenue
New York, NY 10021

WILLIS PERSON
University of Florida
Chemistry Dept.
Gainesville, FL 32611

DAVID MASSON
University of Toronto
Mathematics Department
Toronto, Ontario, Canada M5S 1A1

TADEUSZ PLUTA
University of Florida
QTP, Williamson Hall
Gainesville, FL 32611

PAUL MEZEY
University of Saskatchewan
Departments of Chemistry & Math
Saskatoon, Canada S7N 0W0

JERKER PORATH
University of Uppsala
Institute of Biochemistry, P.O.B. 576
S-751 23 Uppsala, Sweden

FLORIAN MULLER-PLATHE
Max Planck Inst. fur Astrophys.
Karl-Schwarzchild-Strass 1
D-8046 Garching Bei Munchen,
West Germany

GEORGE D. PURVIS
University of Florida
QTP, Williamson Hall
Gainesville, FL 32611

WILLEM NIEWPOORT
University of Groningen
Lab. of Chem. Physics
Nijenborgh 16
97471G Groningen, Netherlands

JAMES RABINOWITZ
United States Environmental
Protection Agency
MD-68
Research Triangle Park, NC 27711

MILAN RANDIC
Ames Laboratory
3225 Kingman Road
Ames, IA 50010

HAROLD SCHERAGA
Baker Lab. of Chemistry
Cornell University
Ithaca, NY 14853-1301

ROBERT REIN
Roswell Park Memorial Inst.
666 Elm Street
Buffalo, NY 14263

PAUL SEYBOLD
Department of Chemistry
University of California
La Jolla, CA 92096

GRAHAM RICHARDS
Physical Chemistry Lab.
Oxford University
South Park Road
Oxford OX1 3QZ England

EDISON DA SILVA
Inst. de Fisica, UNICAMP
Cid. Univ. Barao Geraldo
Campinas, 13081 Brazil

JOHN R. SABIN
Quantum Theory Project
Williamson Hall
University of Florida
Gainesville, FL 32611

FULVIA STAMATO
Departamento de Quimica
Univ. Federal de Sao Carlos
13560 Sao Carlos, SP, Brazil

MUKTI SARMA
Department of Chemistry
SUNY
Albany, NY 12222

MUTTAIYA SUNDARALINGAM
Department of Biochemistry
University of Wisconsin
Madison, WI 58706

RAMASWAMY SARMA
Department of Chemistry
SUNY
Albany, NY 12222

KRYSTYNA SZCZEPANIAK
Department of Chemistry
University of Florida
Gainesville, FL 32611

STEVE SCHEINER
Department of Chemistry
Southern Illinois University
Carbondale, IL 62901

MARIAN SZCZESNIZK
Department of Chemistry
University of Florida
Gainesville, FL 32611

MARK THOMPSON
Quantum Theory Project
University of Florida
Gainesville, FL 32611

COLIN THOMSON
Department of Chemistry
University of St. Andrews
North Haugh, St. Andrews, Scotland

NENAD TRINAJSTIC
The Rudjer Boskovic Institute
P.O.B. 1016
41001 Zagreb, Croatia, Yugoslavia

DAN URRY
Laboratory of Molecular Biophysics
University of Alabama
P.O. Box 311, University Station
Birmingham, AL 35294

MATESH VARMA
Brookhaven National Laboratory
S&EP Division, Building 535A
Upton, NY 11973

DAN VASILESCU
Lab. de Biophysique
Univ. de Nice-Parc Valrose
06034 Nice Cedex, France

HAREL WEINSTEIN
Department of Physiology
and Biophysics
Mt. Sinai School of Medicine
New York, NY 10029

J. S. YADAV
Department of Chemistry
New Jersey Institute of Technology
Newark, NJ 07102

MICHAEL C. ZERNER
Quantum Theory Project
University of Florida
Gainesville, FL 32611

XUEHE ZHENG
Quantum Theory Project
University of Florida
Gainesville, FL 32611

In Memory of Professor Albert Szent-Györgyi

SIDNEY FOX

Institute for Molecular and Cellular Evolution,
University of Miami, Coral Gables, Florida 33134, U.S.A.

In paying tribute to Albert Szent-Györgyi, one can call on a wealth of recorded observation. The main details are well known and others are documented. Albert was a regular participant at the Sanibel conference; this tribute stems from that special relationship.

Since my days as a graduate student in the late 1930s I recall Szent-Györgyi as a teacher, although not in the classroom but through scientific literature. I know that numerous others have likewise perceived Albert as one of their most remembered non-classroom teachers.

The extent of this teaching relationship was emphasized for me on a global scale while in China for the month of October. Visiting a number of institutes in Beijing and Shanghai, I found quotations in (only) two of the entrance foyers. One of these was translated for me from the Chinese by my host. I could neither translate it nor transcribe the rapid translation I heard. The other was another mural quotation of Szent-Györgyi also, but in English, "Research is to see what everyone else has seen and to think what no one else has thought."

Szent-Györgyi's influence reached far. His warm human qualities are illustrated by a personal quotation of what he said to Mrs. Fox and me outside a meeting room. We three spotted the lovely Mrs. (Marcia) Szent-Györgyi approaching in the corridor. Albert said, "There is the best discovery I ever made."

Szent-Györgyi's contributions to, and suggestions in, science were marked by the outstanding ability to reduce concepts to simple terms. He was unmatched in his ability to place his feet in one area of science, and to look elsewhere, see what few others have seen, and "to think what no one else has thought."

It is painful to say goodbye to Albert Szent-Györgyi. The pain is however alloyed with joy when we recall that we are all better off for the fact that this "man of the century" lived in our time.

Albert Szent-Györgyi's Impact on Theoretical Biophysics[†]

JÁNOS J. LADIK

Institute for Theoretical Chemistry and Laboratory of the National Foundation for Cancer Research at the Friedrich-Alexander-University Erlangen-Nürnberg, D-8520 Erlangen, Egerlandstr. 3, West Germany

Introduction

Among the many ideas with which Albert Szent-Györgyi has inspired biophysics research, three examples should be discussed here in detail.

In 1941 he explained certain biological phenomena, including that under certain conditions, proteins can become conductors [1]. Later it was found, for instance, that if light is absorbed by chromophore A situated at the end of a polypeptide chain, the energy of the excited electron is emitted in the form of fluoroescence at the other end of the chain (200–300 Å from the first) by another chromophore, B [2]. (For further biological examples which make electronic conduction in proteins probable see Ref. 3). With modern quantum theory of solids it was possible to show that this can happen under two conditions:

First, if there are free charge carriers in the proteins (which very probably can be generated under biological conditions with the help of charge transfer, CT).

Second, despite the strong side chain disorder of proteins there is the possibility of non-negligible conduction due to hopping. The conditions of the occurrence of such hopping will be discussed in detail.

Szent-Györgyi has predicted also that proteins can form higher structures only if they are conductors [4]. Again with the help of modern quantum theory it was possible to show that this is the case, because the van der Waals forces are much larger in the interacting conducting chains than between insulator chains.

Szent-Györgyi often emphasized that easy energy and charge transport in DNA and proteins are necessary for normal cell functioning. He pointed out that if the flow of electric charges in these macromolecules is hindered this can lead to a cancerous state [5]. It will be shown below that to have an oxygen metabolism (or efficient photosynthesis in plants) it is really necessary to have a non-negligible conduction of proteins which was originally Szent-Györgyi's main argument in his early papers [1] in which he has assumed electronic conduction in proteins.

[†]In memorium of my Great Teacher, and Respected Friend, October 22, 1986.

Hopping Conduction in Disordered Proteins

Szent-Györgyi's suggestion [1] that proteins can become semiconductors was generally rejected in its time. The two main counterarguments were (1) that the fundamental energy gap in proteins is too large, and (2) that, due to the disorder caused by the different side chains, proteins do not have continuous regions of allowed energies but only very narrow peaks in the density of states (DOS) of allowed energy levels.

To answer the first counterargument one can point out that due to the large probability of charge transfer (CT) *in vivo*, free charge carriers can be generated quite easily. As Szent-Györgyi has suggested [5] and we have calculated [6], there is a possibility that in compounds with two carbonyl groups (like methylglyoxal) a small amount (0.03 e) of CT may occur between methylglyoxal and a peptide group [6] (the latter being the electron donor). *In vivo* there seems to be a much more effective CT transfer mechanism from the negatively charged PO_4^- groups of DNA to the positively charged side groups of a polypeptide chain (to the guanidium end groups of arginine, to protonated lysine side chains or to histidine cations) in nucleoproteins. It should be noted that one of the most important nucleoproteins is nucleohistone, which is rich in arginine, and further, in DNA (as all quantum mechanical calculations show [7, 8]) there is an internal CT from the sugar residue to the phosphate group of ∼0.05 e. In this way the actual charge on a phosphate group of a polynucleotide is −1.05 e instead of 1.0 e which facilitates strongly the postulated CT from DNA to proteins. Therefore, the gap problem of the conduction in proteins seems to be eliminated by the described generation of free charge carriers.

Detailed calculations in which four-component (glycine, serine, asparagine, and cysteine) nonperiodic polypeptide chains with random sequences were treated using the theory of disordered systems (for the applied so-called negative factor counting (NFC) technique see Ref. 9) showed that the disorder results in a very strong broadening of the allowed energy states both in the valence and conduction band regions [10]. This means that if free charge carriers are obtained (especially in the originally empty conduction band region of the polypeptide) in these regions, with CT from DNA or otherwise, detailed calculations have shown [10] (despite the localization of the wave functions due to the disorder), that a strong hopping conduction can be expected in these systems. It is well known from linear algebra that for any arbitrary energy level the corresponding eigenvector (wavefunction) can be determined with great precision using the standard inverse iteration technique [11]. In this way, after determining the individual energy levels in the conduction band region of the investigated disordered (i.e., having a random sequence) four-component gly, ser, asp, and cys system in a 1:1:1:1 composition using the previously mentioned negative factor counting method [9], we were able to determine the wavefunctions belonging to these systems [10]. Having these wavefunctions it was easy to see that all states in this disordered chain are strongly localized on a single amino acid residue. The detailed calculations were done for the first 15 filled and the next 40 unfilled levels in the conduction band region. Since these calculations were performed on the *ab initio* Hartree-Fock level (all electrons and all interactions between them were taken into account), the energy level spacing between consecutive levels is only 0.003 eV

(smaller than the thermal energy at 300°K, $k_B \cdot 300 = 0.03$ eV. Detailed considerations have shown that to get one electron of a chain of $N = 300$ units from one end to the other, the average energy for hopping from a unit to its first neighbor is $\Delta E_{i\rightarrow j} = 0.25$ eV. One can substitute this value in the expression of the primary jump rate [12] (the number of jumps per second from unit A to unit B)

$$P_{A\rightarrow B} = \nu_{\text{phonon}} e^{-\Delta E_{i\rightarrow j}/\beta_B T}(\sum_{r,s} < \chi_r^A | \chi_s^B > c_{i,r} c_{j,s})^2 \qquad (1)$$

Here ν_{phonon} is the acoustic phonon frequency which in polypeptides corresponds to the vibrations of the side chains relative to each other, $c_{i,r}$ and $c_{j,s}$ are the dominant coefficients of those LCAO wavefunctions with energies ε_i and ε_j, respectively, which are localized on atoms A and B, respectively, and finally χ_r^A is an atomic orbital (AO) localized on the amino acid residue A. Taking $\nu_{\text{phonon}} = 10^{12}\text{s}^{-1}$ (which is a typical value for acoustic phonons) one obtains for $P_{A\rightarrow B} \approx 10^5$ for first neighbors, 10^9 if the transition occurs within the same residue and 10^3 for second neighbor jump rates. These values are qualitatively in good agreement with $P_{A\rightarrow B}$ values found in different disordered solids which conduct through variable range hopping (for details and references see Ref. 10).

These results mean that if free charge carriers are generated *in vivo* by CT from DNA to protein or *in vitro* by doping by electron donors, proteins can become, despite their aperiodicity, quite good conductors by a hopping mechanism. In this way Szent-Györgyi's 45-year-old prediction [1] could be proven with the help of modern theoretical solid-state physical methods.

Interaction Between Insulator and Conductor Biopolymer Chains

During a personal discussion, Szent-Györgyi theorized that "proteins can form larger structures" only if they are conductors [4]. This is a good example of his legendary intuition. Of course there is (and was) no scientist in the world whose every intuitive idea has been proven correct. But even if only a small fraction of such intuitive ideas of one great scientist are right, such a person can be ranked as an extraordinary genius. There is no question that Albert Szent-Györgyi achieved that status.

If one looks into the theory of intermolecular (or interchain) interactions, one finds that besides the leading electrostatic term, the polarization and dispersion forces are major contributors to these interactions.

If one applies perturbation theory, one can see by the second order (what is also true in higher orders), that the dispersion energy between system A and B, respectively, has the form

$$\frac{C}{\Delta E^A + \Delta E^B} \qquad (2)$$

Here C stands for a fourfold summation of complicated integrals and ΔE^A and ΔE^B are excitation energies of systems A and B, respectively. If A and B are atoms or molecules ΔE^A and ΔE^B will give the energy difference between their highest filled and lowest unfilled energy levels [see Fig. 1(a)]. If many molecules A and B, respectively, are interacting forming the chains polyA and polyB, the corresponding energy levels broaden into so-called energy bands (continuous regions of allowed energy

(a)   (b)   (c)
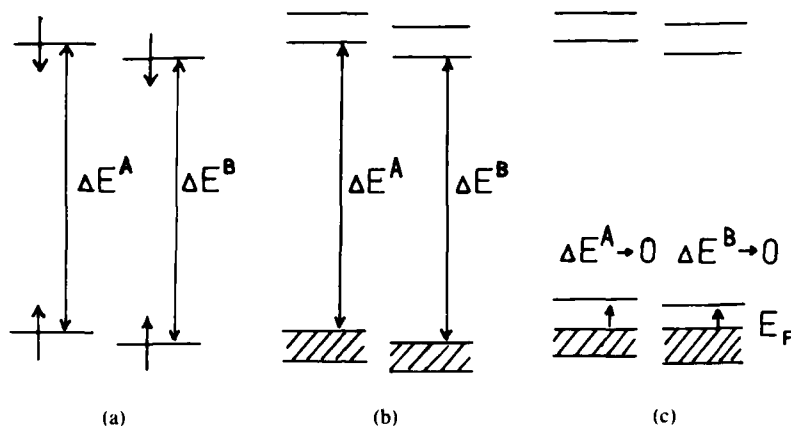
levels) from which one is completely filled and the other one completely empty [see Fig. 1(b)]. In this case both chains are insulators. If, on the other hand, due to CT both chains have only partially filled bands [see Fig. 1(c)], they become conductors. As we were able to show, together with the late Kàlmàn Laki [13], in this case the dispersion interaction (and in a similar way also the polarization interaction) becomes many orders of magnitude larger, because one can excite electrons within the partially filled bands ($\Delta E^A \rightarrow O$, $\Delta E^B \rightarrow O$). For chains with finite length (which is the realistic case) ΔE^A and ΔE^B will not be zero, but very small. Still if the denominator is very small the expression (2) will become large, and so a large increase of interaction energy takes place. This is also very important from the point of view of carcinogenesis, since CT caused by carcinogens can change the population of the valence or conduction bands, respectively, of both of DNA and proteins and therefore, change the dispersion and polarization forces between them. In this way they can influence the DNA–protein interaction (which controls genetic regulation) even if they are bound at some distant point along the DNA chain [14]. We can conclude that Szent-Györgyi's intuition was right in this case also.

Carcinogenesis

Szent-Györgyi has emphasized many times that for normal (noncancerous) cell functioning one needs an easy energy and charge transport [5]. To understand his idea one should think first of all on the respiratory cycle which is crucial for the O_2 metabolism of the normal cell (cancerous cells do not show O_2 metabolism, but instead of this fermentation takes place). One knows that in this respiratory cycle electrons have to be transported from one part to another part of different proteins. Though there are other possible mechanisms for electron transport in a protein (via H-bonds coupled to proton transport, via aromatic intermediates between different heme groups, etc.), the easiest way for electron (and energy) transport to occur in a protein is still via transport along the main polypeptide chain. In the section on Hopping Conduction in Disordered Proteins, one could see that despite the too large gap and disorder in proteins, this is quite possible through effective electron (or in an ex-

cited state) exciton hopping. Proteins (and DNA) which show an electron (exciton) conduction in this way also provide a much quicker and effective possibility for rapid signal transfer in these biopolymers, which certainly plays an important role in the self-regulation of the cell.

Acknowledgment

Bibliography

[1] A. Szent-Györgyi, Nature **148**, 157 (1941); A. Szent-Györgyi, Science **93**, 609 (1941).

[2] W. Lautsch, R. Pacedoy, L. Sommer, H. J. Julius, and N. Bodefeld, Chimia (Zürich) **13**, 129 (1959).

[3] J. Ladik, Kvantumbiokémia, Gendolat Kiandó, Budapest 1967, 7. fejezet.

[4] A. Szent-Györgyi, personal communication.

[5] A. Szent-Györgyi, Bioelectronics (Academic Press, New York-London, 1968); The Living State (Academic Press, New York-London, 1972); Electronic Biology and Cancer (Marcel Dekker Inc. New York-Basél, 1976).

[6] P. Otto, J. Ladik, and A. Szent-Györgyi, Proc. Natl. Acad. Sci. U.S.A. **75**, 3548 (1978).

[7] J. Ladik and S. Suhai, Int. J. Quant. Chem. QBS **7**, 181 (1980).

[8] P. Otto, E. Clementi, and J. Ladik, J. Chem. Phys. **78**, 4547 (1983).

[9] P. Dean, Rev. Mod. Phys. **44**, 122 (1972); R. S. Day and F. Martino, Chem. Phys. Lett. **84**, 86 (1981).

[10] A. K. Bakhshi, J. Ladik, M. Seel, and P. Otto, Chem. Phys. **108**, 203 (1986).

[11] See for instance: J. H. Wilkinson, The Algebraic Eigenvalue Problem (Clarendon Press, Oxford, 1965) p. 633.

[12] N. F. Mott and E. A. Davis, Electronic Processes in Non-Crystalline Materials (Clarendon Press, Oxford, 1981).

[13] K. Laki and J. Ladik, Int. J. Quant. Chem. QBS **3**, 51 (1976).

[14] J. Ladik and S. Suhai, Int. J. Quant. Chem. QBS **5**, 35 (1978).

Some Remarks on Certain Magnetic Properties of Water in the Study of Cancer

MU SHIK JHON* AND PER-OLOV LÖWDIN†

Quantum Theory Project, Departments of Chemistry and Physics, 362 Williamson Hall, University of Florida, Gainesville, Florida 32611, U.S.A.

Abstract

In the study of cancer as a complicated disease caused by many combinations of various factors, it may be of importance to consider also any possible changes of the water structure in the environment of the malignant cells. The occurrence of such changes has been established experimentally, for example, by a study of the magnetic properties of water by nuclear magnetic resonance (NMR) spectra. It has been found that the protons in the water surrounding malignant cells have a much longer spin-lattice relaxation time than the protons in the water around normal cells. This indicates that the water molecules in tumor cells are less structured and able to move more freely than in normal tissues, where, due to the effect of hydrogen bonding, water occurs mainly as five- or six-membered rings. This prolongation of the proton spin-lattice time may be an important factor in cancer, but further studies are necessary before one can decide with certainty whether it would be possible to use this effect to diagnose malignant transformations at an early stage. It is suggested that changes in the magnetic properties of water in a malignant tumor during chemotherapy and other treatments be monitored as control tools.

This paper provides some remarks prepared for the panel discussion on "Models of Carcinogenesis" at the 1987 Sanibel Symposium on Quantum Biology and Quantum Pharmacology, and it deals briefly with the role of the water environment in cancer.

According to Boyland [1], a rough estimate of the most important sources causing human cancer indicates that less than 5% have physical origins such as radiation damage, that less than 5% are caused by external viruses, and that more than 90% may come from chemicals in our environment. More recent studies [2], have found that some of our living habits such as tobacco smoking, drinking of alcohol, the combinations of these two factors, or excess consumption of fats, represent a very large portion of the causes in the various forms of cancer. Most of these carcinogenic chemicals enter the body through breathing, eating, and drinking, and, even though many of them come in water solution, one would not expect that water itself would play any major role in the study of cancer.

Cancer is a disease characterized by the fact that some cells in the body start replicating in an uncontrolled fashion. All experimental experience indicates that cancer is connected with some form of damage to the genetic code and its regulatory behavior,

---

Permanent affiliations:

  *Professor of Chemistry, Korea Advanced Institute of Science and Technology, Seoul, Korea.

  †Professor Emeritus of Uppsala University, Uppsala, Sweden; NFCR Senior Investigator.

namely to some change in the DNA-molecule as a nucleoprotein in an ordinary body cell. At the 1986 Sanibel panel discussion on this subject, the following models of carcinogenesis [3] were particularly discussed: the immunological model, the virus model, the somatic mutation model, the reading-error model, and the proton-tunneling model.

In the study of carcinogenesis, the presence of water is often taken for granted and therefore ignored, although water plays an essential role in all biological systems. It has been found, however, that in the environment of malignant cells, the system of water molecules shows a different structure than around normal cells. This may be seen, for instance, in the nuclear magnetic resonance (NMR) spectra of the water protons. If one measures the proton spin relaxation time (T_1) at 100 MHz in normal and malignant human tissues, the relaxation times in tumor tissues (T_{1tumor}) are, on the average, longer than the relaxation times in normal tissues $(T_{1normal})$ (see Ref. 4). This indicates that the system of water molecules in tumor tissues are less structured and the molecules are able to move more freely than in normal tissues.

In this connection, it is important to study the structure of water in the environment of biological molecules in general, and in normal cells or tissues in greater detail for comparison purposes. According to our recent work [5], the environmental water structures in solutions of biological molecules are investigated with particular attention to the ring structures formed by hydrogen bonding. It is shown that the most plausible forms of water structure are short-range six-membered connections. Next, according to our improved water model [6, 7], an equilibrium between six-membered structures and five-membered ring connections is established, and the cooperative changes between their structures are observed.

TABLE I. The values of ΔE_{nn} (kJ/mol)

Positive hydration		Negative hydration	
Li^+	27.2	K^+	-3.8
Na^+	3.3	Rb^+	-6.3
Ca^{2+}	32.2	Cs^+	-8.8
Sr^{2+}	10.5	Be^{2+}	-184.5
Ba^{2+}	1.3	Mg^{2+}	-8.8
Sc^{3+}	51.9	Al^{3+}	-313.4
Y^{3+}	41.0	Tl^+	-7.1
La^{3+}	23.0	F^-	-18.0
Ag^+	4.2	Cl^-	-7.5
Zn^{2+}	50.6	Br^-	-7.5
Mn^{2+}	42.3	I^-	-7.9
Fe^{2+}	48.5		
Co^{2+}	48.5		
Ni^{2+}	51.0		
Cu^{2+}	49.8		
Fe^{3+}	51.9		
Hg^{2+}	10.9		
Pb^{2+}	5.9		

In lowering the temperature from 0°C to −40°C under supercooling conditions, the form of short-range six-membered connections are increased dominantly. In this connection, it is also interesting to observe the existence of many medical and biological reports about the unusual growth of plants and animals in abnormally cold regions, where supercooled water is provided in large amounts, but many more studies are certainly necessary before any definite conclusions about this interesting phenomenon can be drawn.

We have also investigated the structural effect of various ions in dilute aqueous solutions. In particular, the quantitative prediction of the strength of the various ions in breaking and making structures can be illustrated by considering the interaction between water molecules in the solutions (ΔE_{ww}) [8]. In Table I, we have listed the values indicating the relative strength of structure-breaking and -making ions. It should be added here that research on the properties of magnetized water under the influence of different magnetic fields and various flow conditions of ionic solutions are in progress. The preliminary results of these experiments indicate that the structural effects on water observed by increasing the magnetic field are similar to the effects shown by the addition of structure-making ions or temperature lowering, that is, a favoring of the short-range six-membered connections, and such structure changes will also influence, for instance, the surface tension under an increase of applied magnetic fields. As an example, the surface tensions of $CuCl_2$ salt and Mohr's salt (a Fe^{2+} ion salt) have been measured at 25°C by a Fisher Autotensiomat with some results shown in Figure 1.

Finally, to make a preliminary test of the assumption that there is some connection between malignant cells and the water structure of their environment, we have carried
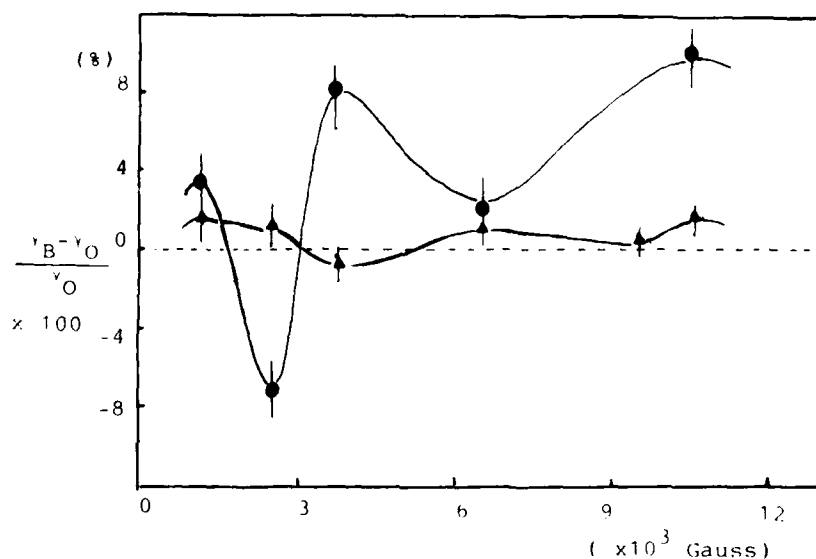
Figure 1. The change of surface tension according to magnetic field strength change in 0.025 M Mohr's salt solution in 0.5 N H_2SO_4 and pure water. (●) Mohr's salt; (▲) pure water.

out a study of a cell structure *in vitro*, in which the water structure was changed by the addition of a structure-making ion. In this experiment, cultured fibroblast 3T331 cell lines are used, and the material medium was prepared as Eagle's Modified Eagle Medium (MEM) plus 5% fetal bovine serum. In Figure 2, it is shown that the addition of 25 mM Ca^{2+} ion (one of the strong structure-making ions) obtained by an electrolytic process in $CaCl_2$ brings a remarkable decrease of the cell numbers from 2 million down to nearly 20,000 in a few days.

This decrease could, of course, also be an effect of the Ca ions themselves. However, when one added the Ca ions in the form of 25 mM $CaCl_2$ solution, which con-
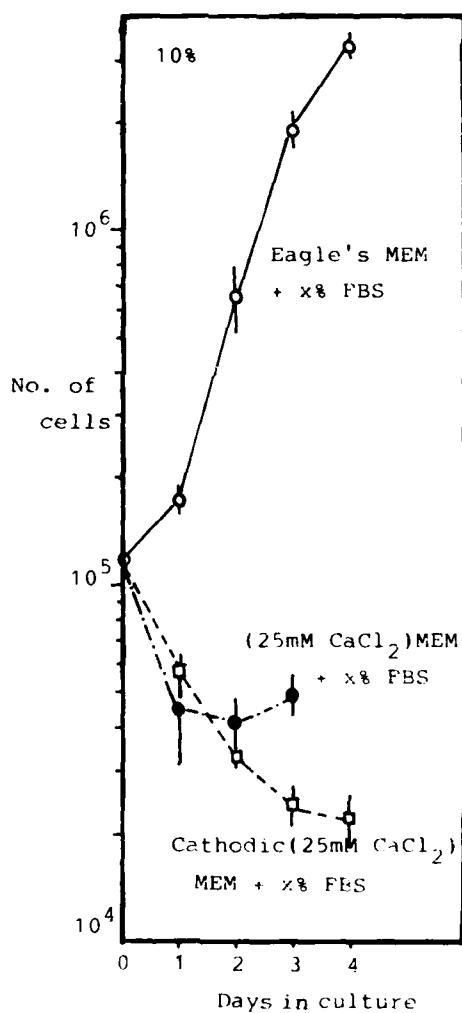
Figure 2.   Number of cells versus days.

tains a mixture of both structure-making and structure-breaking ions, one is unable to control the increased number of the cell lines. This experiment was carried out in collaboration with Professor H. Sequchi, Kochi Medical School, Japan. Even if this *in vitro* experiment is not conclusive for what will happen in a living system, it indicates that it may be worthwhile to continue studies along these lines and to try to influence the water structure in malignant tumors by various means, to see if such an approach has any influence on the disease.

The prolonged spin-lattice relaxation times of the protons in the water environment of malignant cells may be an important symptom of cancer, and investigations as to whether it may be used to diagnose any malignant changes at an early stage should continue. It should be remembered, however, that a similar effect may occur in connection with other diseases.

In fact, a similar prolongation seems to occur in the water environment of the beta cells in diabetes. It is somewhat surprising to learn that, on a cellular or subcellular level, there are certain similarities between the causes of cancer and the causes of diabetes. Among the causes of diabetes [9], the following factors are considered: (1) virus that destroy the beta cells or cause them to malfunction, (2) autoimmune reactions that damage the beta cells, (3) environmental chemicals that may destroy the beta cells or cause them to function improperly, (4) hereditary factors that reduce the production of insulin in pancreatic beta cells. In addition, we have now found a characteristic change in the water structure around malfunctioning beta cells in diabetes, and further studies may be worthwhile in this connection.

All the results reported at this panel discussion are somewhat preliminary, but further experiments are in progress. At this point, however, we would like to make the definite suggestion that one should follow the changes in the magnetic properties of water in a malignant tumor during chemotherapy and other treatments and to use it as an extra control tool.

It should finally be observed that the water structures discussed here and their changes are essentially due to the hydrogen bonding and its collective behavior. A hydrogen bond, in this instance, is a proton shared between two electron pairs, and the theoretical study of this simple system is still one of the most important problems of quantum biology.

Bibliography

[1] E. Boyland, *Proc. Israel Acad. Sciences and Humanities* (E. Bergmann and B. Pullman. Eds. Jerusalem, 1969). p. 203.

[2] R. Doll and R. Peto, *The Causes of Cancer: Quantitative Estimates of Cancer in United States Today* (Oxford Univ. Press. Oxford, 1981).

[3] P. O. Löwdin, Proc. Sanibel Symposia on Quantum Biology and Quantum Pharmacology. Int. J. Quantum Chem. S QB 13, 1986, 311.

[4] J. S. Waugh, *NMR Imaging in Biomedicine* (Academic Press, New York, 1982). p. 22.

[5] C. N. Yoon and M. S. Jhon, Int. J. Quantum Chem. 12, 33 (1986).

[6] M. S. Jhon and H. Eyring, Ann. Rev. Phys. Chem. 27, 45 (1976).

[7] M. S. Jhon and B. J. Yoon, *Ions and Molecules in Solution*. N. Tanaka et al., Eds. (Elsevier, Amsterdam, 1983).

[8] M. J. Moon and M. S. Jhon, Bull. Chem. Soc. Jpn. **59**, 1215 (1986).
[9] H. J. Sanders, Chem. Eng. News, March 2, 1981.

Intercalative Binding and Antitumor Activity of Bisantrene and Derivatives

KAI-XIAN CHEN, NOHAD GRESH, AND BERNARD PULLMAN

Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique associé au C.N.R.S. 13, rue Pierre et Marie Curie, 75005 Paris, France

Abstract

A theoretical exploration of the intercalative mode of binding with DNA of the antitumor drug bisantrene and the closely related but inactive drug LC 230487 indicates that both drugs show a preference for a major groove intercalation to GC sequences. In this mode of binding bisantrene is predicted, however, to have a greater affinity for DNA than LC 230487 due to a stronger and somewhat different network of H-bonding interaction with the double-stranded receptor. While this difference in the strength and pattern of the intercalative association could be considered as possibly related to the striking difference in their antitumor activity, the experimental observation that both compounds exhibit a practically identical affinity constant for binding to DNA, confronted with the theoretical evaluation of different affinities for intercalative association suggest that LC 230487 could possibly interact with DNA by a different mechanism which does not lead to antitumor activity. This could perhaps consist of an exterior binding.

Introduction

The restriction imposed on the use of anthracyclines as antitumor agents by their dose-dependent cardiotoxicity has prompted the design and synthesis of derivatives or analogues with the aim of diminishing or abolishing their toxic effect while maintaining or enhancing their chemotherapeutic activity. The origins of the cardiotoxicity are unknown. Subscribing to the findings of an important group of researchers headed by Lown [1–3], the cardiotoxic effects could be due to the engagement of the quinoid ring of the anthracyclines in a redox cycle, leading through intermediate semi-quinones, to the generation of reactive oxygen species capable to produce peroxidative injury to membrane lipids and DNA lesions. "Second-generation anthracyclines" have been prepared inspired by this concept, which are characterized by more negative reduction potentials and thus a lesser aptitude to produce the supposedly harmful oxygenated free radicals. Outstanding among these new compounds are *mitoxantrone* [3,4] and *anthrapyrazole* [5,6]. These molecules are found, indeed, to be less cardiotoxic than the classical anthracyclines; it seems, however, that they also have a reduced potency and spectrum of antitumor activity. This situation could be due to many factors, one possibility being that, as shown by theoretical exploration [7,8] their mode of binding to DNA, although also intercalative, shows significant differences with respect to that of the classical anthracyclines. A tempting conclusion is that the antitumor activity could well depend not only on the strength of binding of a drug to DNA but also on its detailed mode of association [9].

An interesting step along this line of research is the recent preparation and assay of *bisantrene* [1], a bisquanylhydrazone of anthracene-9, 10-dicarboxylaldehyde [10]. The compound possesses a planar anthracene ring which makes it a suitable candidate for intercalation into DNA but is devoid of the quinoid grouping and should thus not be subject to the redox cycle. In conformity with this situation it is found that bisantrene intercalates into double-stranded nucleic acids [11–13]. It manifests, however, also a second weaker type of interaction with this biopolymer, believed to consist of an exterior electrostatic binding to the anionic backbone of the DNA helix [13]. Bisantrene is said to be a low toxicity agent [11] and to manifest "good" antitumor activity against P-388 leukemia and B-16 melanoma in mice [14]. This drug seems thus akin to the "second-generation anthracyclines."

Interestingly, a close derivative of bisantrene, designated as CL 230487 [11], although binding to DNA with practically the same affinity constant as bisantrene [13] is inactive in the anticancer tests mentioned (Fig. 1).

This situation could possibly be related to the problem, outlined at the beginning of this paper, of the relationship between the *mode* of binding to DNA and antitumor activity of drugs. It is this question that we propose to investigate in this paper by trying to gain, through a theoretical exploration, more information about the details of interaction with DNA of I and II.
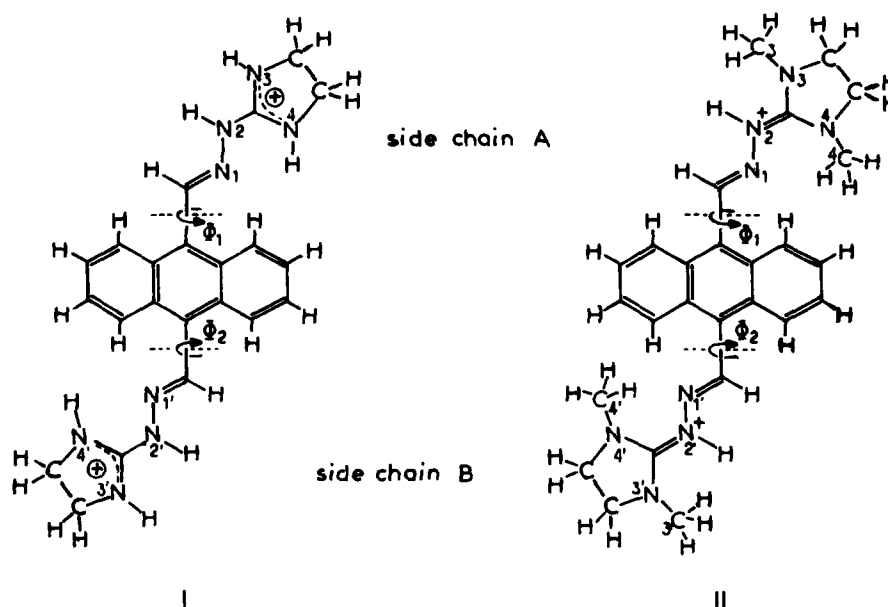
Figure 1.   Chemical formula and atom numbering of bisantrene I and CL 230487 II. (The dashed lines indicate the separation between constitutive fragments.)

Methodology

The binding mode was investigated with the help of model double-stranded tet-ramers $d(GCGC)_2$ and $d(ATAT)_2$ in which the intercalation is assumed to take place in the center of the oligomer, namely, between base pairs 2–3' and 3–2' (Fig. 2), that is, in a pyrimidine (3'–5') purine sequence. This choice is in keeping with the conclusions of our previous works on related compounds [7–9] where it was shown that the isomeric oligonucleotides $d(CGCG)_2$ and $d(TATA)_2$, with a purine (3'–5') pyrimidine sequence at the intercalation site, were significantly disfavored in terms of their binding affinities. The value of 17° has been adopted for the unwinding angle $(\Delta\alpha)$, in analogy to that of mitoxantrone, although a somewhat smaller value of this angle (14°) is possible [15]. The dihedral angles adopted for the tetranucleotide are those derived by Miller et al. for this value of $\Delta\alpha$ in the course of their study of inter-calative binding using the AGNAS procedure [16]. In line also with this procedure, a mixed sugar pucker pattern C3'endo–C2' endo was adopted at the intercalation site [17].

Similar to our previous studies [7, 8, 18–22], the interactions are computed be-tween rigid models of the oligonucleotides, fixed in the above-mentioned conforma-tion, but allow a large flexibility to the drug, considered as diprotonated on $N_1(N_{3'})$ and $N_4(N_{4'})$ of bisantrene, or $N_2(N_{2'})$ of CL 230487.

The variations of the conformational energy of the intercalator upon complex for-mation are computed with the SIBFA (sum of interactions between fragments com-puted *ab initio*) procedure [23]. Within this methodology, the intercalator is built of elementary constitutive fragments separated by single bonds and the variation of the intramolecular energy upon a conformational change is obtained as the variable part of the sum of interactions between the fragments expressed as:

$$\delta E_{conf.} = E_{MTP} + E_{pol} + E_{rep} + E_{disp} + E_{tor} \tag{1}$$

The intermolecular oligonucleotide-intercalator interaction energies are computed by the SIBFA 2 procedure [24] as a sum of five terms:
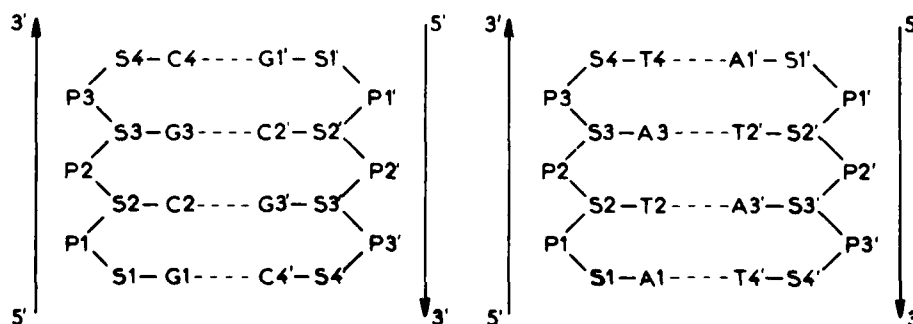
Figure 2. The model tetramers.

$$\Delta E_{inter.} = E_{MTP} + E_{pol} + E_{rep} + E_{disp} + E_{CT} \qquad (2)$$

In expressions (1) and (2), E_{MTP} and E_{pol} denote, respectively, the electrostatic and polarization contributions, computed using a multipolar expansion of the *ab initio* SCF molecular wave functions of the fragments, and E_{rep} and E_{disp} are the repulsion and dispersion contributions, respectively, E_{tor} is a torsional energy contribution, calibrated in [23] for elementary rotations along C-C and C-O bonds and E_{CT} is a charge-transfer contribution (see Refs. 23, 24 for details).

Drugs I and II are built out of constitutive subfragments, which are individualized in Figure 1 by dashed lines. The internal geometry of the fragments (bond length and valence angles) is taken to be the same as in the crystal structure of related compounds. The conformational energies of the intercalator, either free or in the complex, were minimized as a function of all its dihedral angles.

The DNA tetramers are built of elementary fragments, following a procedure developed in our laboratory for the computation of the molecular electrostatic potential of large biomolecules [25]. The fragments are the nucleic acid bases, the phosphodiester linkage and deoxyribose, with the same internal geometry as in standard B-DNA [26].

The *ab initio* SCF computations on the constitutive fragments necessary for the application of the SIBFA procedure [23, 24] are performed using our usual basis set with a *dzeta* exponent of 1.5 on the ammonium hydrogen [27].

The search for the optimal configurations for the oligonucleotide-intercalator complex are performed by energy minimization [28] of the sum of $\Delta E_{inter} + \delta E_{conf} + \delta E_{unstack}$, the last term measuring the energy expense necessary to unstack the base pairs at the intercalation site, computed also by the SIBFA procedure. The variables involved in the minimization are the conformational angles of the intercalator together with the intermolecular variables defining the relative orientation of the intercalator with respect to the tetramer.

Both major and minor groove intercalative binding, corresponding to the possible location of the side chains in both grooves, was explored for the two sequences studied.

Results and Discussions

The results of the computations are reported in Tables I and II. Table I lists the values of the intermolecular interaction energy ΔE_{inter} and of its components, the value of the conformational energy variations $\delta E_{conf.}$ of I and II with respect to their intrinsically preferred conformations taken as energy zero's, the energy of unstacking of the oligonucleotides $\delta E_{unstack}$, the resulting energy balance $\delta E = \Delta E_{inter} - \delta E_{conf.} - \delta E_{unstack}$, and the energy balance difference, δ, with respect to the most stable value taken as energy zero. Table II reports the most significant interatomic distances between H atoms of I or II and binding sites in the tetramers. In Figure 3 is given a representation of the optimized structures of the intercalator with $d(GCGC)_2$ and $d(ATAT)_2$.

In the optimized conformations of the drug-tetranucleotide complexes, the plane of the chromophore is parallel to the plane of the base pairs at the intercalation site with its long axis perpendicular to the long axis of these pairs. The two side chains of the

TABLE I. Values of the intercalative interaction energies of bisantrene and CL 230487 in the major and minor grooves of sequences d(GCGC)₂ and d(ATAT)₂ (see text for definitions) (energies in kcal/mol).

	Major groove				Minor groove			
	d(GCGC)$_2$		d(ATAT)$_2$		d(GCGC)$_2$		d(ATAT)$_2$	
	Bisantrene	CL 230487	Bisantrene	CL 230487	Bisantrene	CL 230487	Bisantrene	CL 230487
ΔE_{inter}	−499.2	−456.4	−490.4	−441.5	−444.4	−427.9	−450.2	−435.8
E_{MTP}	−468.5	−426.7	−456.7	−407.4	−402.1	−396.4	−409.2	−404.5
E_{pol}	−24.2	−21.6	−25.5	−22.0	−23.4	−22.4	−23.0	−22.1
E_{ct}	−3.3	−1.5	−3.6	−1.1	−3.8	−5.3	−3.2	−4.8
E_{disp}	−57.9	−49.0	−58.2	−49.7	−56.4	−69.7	−52.3	−64.6
E_{rep}	54.8	42.4	53.4	38.7	41.3	65.9	37.5	60.2
δE_{conf}	0.7	1.8	2.4	0.0	0.3	0.6	0.3	0.8
$\delta E_{unstack}$	16.9	16.9	18.0	18.0	16.9	16.9	18.0	18.0
δE	−481.6	−437.7	−470.1	−423.5	−427.2	−410.4	−431.9	−417.1
δ	0.0	43.9	11.5	58.1	54.4	71.2	49.7	64.5

TABLE II. Interatomic distances (in Angstrom units) of H atoms of side chain A of bisantrene and CL 230487 with binding sites on the unprimed strand of the tetramers. Identical distances are found between corresponding H atoms of side chain B and binding sites on the primed strand (not shown).

Major groove				Minor Groove			
d(GCGC)₂		d(ATAT)₂		d(GCGC)₂		d(ATAT)₂	
Bisantrene	CL 230487	Bisantrene	CL 230487	Bisantrene	CL 230487	Bisantrene	CL 230487
$H(N3)$—$O_1(P2)$ 1.80	$H_2(C_1)$—$O_1(P2)$ 1.68	$H(N_1)$—$O_1(P2)$ 1.77	$H_2(C_1)$—$O_1(P2)$ 1.72	$H(N_1)$—$O_1(S\text{-}C2)$ 2.96	$H(N_2)$—$O_1(S\text{-}G3)$ 2.40	$H(N_1)$—$O_3(S\text{-}T2)$ 2.?6	$H(N_2)$—$O_3(S\text{-}A3)$ 2.41
$H(N2)$—$O_1(P2)$ 2.53	$H_1(C_1)$—$O_6(G3)$ 2.64	$H(N_2)$—$O_1(P2)$ 2.50	$H_1(C_1)$—$N_6(A3)$ 3.02	$H(N_2)$—$O_1(S\text{-}G3)$ 2.16	$H_2(C_1)$—$O_1(S\text{-}C2)$ 2.16	$H(N_2)$—$O_1(S\text{-}A3)$ 2.16	$H_2(C_1)$—$O_3(S\text{-}T2)$ 2.17
—$N_7(G3)$ 3.48	—$N_7(G3)$ 2.80	—$N_7(A3)$ 3.45	—$N_7(A3)$ 3.55	—$O_1(S\text{-}C2)$ 3.33	—$O_4(S\text{-}G3)$ 3.18	—$O_1(S\text{-}T2)$ 3.33	—$O_3(S\text{-}A3)$ 3.18
$H(N4)$—$N_7(G3)$ 3.47	$H_1(C_1)$—$O_1(P2)$ 3.17	N_1—$H(N_6\text{-}A3)$ 3.08	$H_1(C_1)$—$O_1(P2)$ 3.25	—$O_1(S\text{-}G3)$ 3.52	$H_1(C_1)$—$O_1(S\text{-}C2)$ 3.58	—$O_3(S\text{-}A3)$ 3.52	$H_1(C_3)$—$O_3(S\text{-}T2)$ 3.58
	$H_1(C_1)$—$O_1(P2)$ 3.36		$H_1(C_1)$—$O_1(P2)$ 3.14		$H_1(C_1)$—$O_1(S\text{-}G3)$ 3.48		$H_3(C_1)$—$O_1(S\text{-}A3)$ 3.50
	$H(N_2)$—$O_1(P2)$ 3.38		$H(N_2)$—$O_1(P2)$ 3.22		$H_1(C_1)$—$O_3(C2)$ 3.32		$H_1(C_4)$—$O_2(T2)$ 3.33

[a] In the present notations, e.g., $O_1,(S\text{-}C2)$ denotes O_1, oxygen of the sugar linked to base C2, $H(N_6\text{-}A3)$ the 6-amino hydrogen of base A3, etc.

<div align="center">(a)   (b)</div>
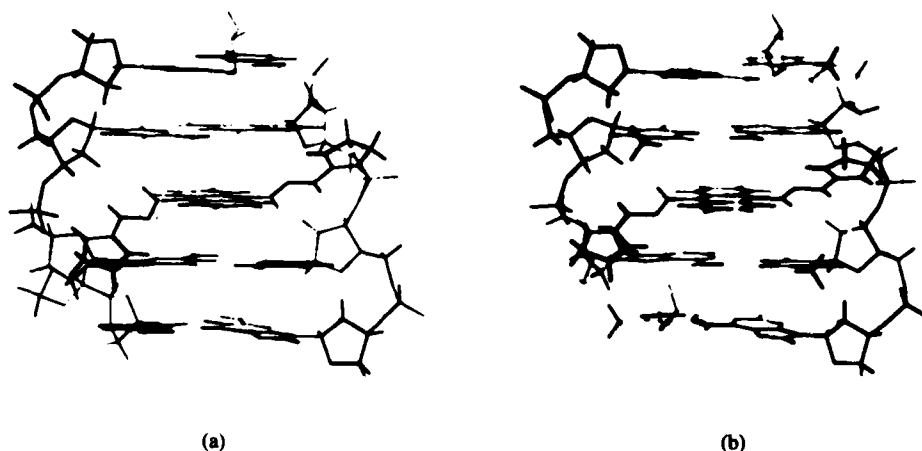
Figure 3. Computer drawn optimized structures of bisantrene with d(GCGC)$_2$ and d(ATAT)$_2$.

drugs are located in the same groove of the oligonucleotides, stretching symmetrically up and down, toward the two backbone strands of DNA. We have investigated also conformations of the complexes with the two side chains of the drugs in different grooves, and with a different orientation of the chromophore. All these conformations correspond to much weaker intermolecular interaction energies.

The results listed in Table I indicate that:

1. For both drugs the intermolecular interaction energy, ΔE_{inter}, and the overall energy balance of the interaction, δE, show a significant sequence specificity favoring the tetramer d(GCGC)$_2$ over d(ATAT)$_2$. The preference is, in terms of δE, 11.5 kcal/mol and 14.2 kcal/mol for I and II, respectively.

2. For both drugs and both tetramers, major groove binding is obviously preferred over minor groove binding.

3. For both d(GCGC)$_2$ and d(ATAT)$_2$ sequences and both major and minor groove bindings, the active bisantrene shows a greater affinity for intercalative binding than does the inactive LC 230487. The preferences for major groove binding, favoring I over II, are 43.9 and 46.6 kcal/mol, in the cases of binding to d(GCGC)$_2$ and to d(ATAT)$_2$ sequences, respectively. For minor groove binding, these preferences are 16.8 and 14.8 kcal/mol, still favoring I.

4. Altogether, the order of the overall energy balance, δE, for the intercalative binding is, for both compounds I and II:

<div align="center">

d(GCGC)$_2$ >   d(ATAT)$_2$ >   d(ATAT)$_2$ >   (GCGC)$_2$

major groove   major groove   minor groove   minor groove.

</div>
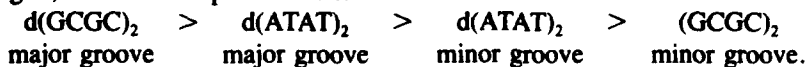
the strongest association occuring with the d(GCGC)$_2$ sequence and with two side chains of the drugs in the major groove of the tetranucleotide.

Up to this point of the analysis of the results of the intercalative binding the difference between the antitumorally active bisantrene and the inactive derivative LC 230487, expressed in energy terms, appears essentially quantitative, the most ef-

ficient interaction characterizing the active compound. It is obvious, however, that this situation must correspond also to structural differences in the detailed mode of binding. In order to elucidate this aspect of the situation we may look into the details of the intermolecular interactions established between the drugs and the receptor oligonucleotides (Table II).

The two compounds, I and II, are diprotonated on their side chains. The results of *ab initio* computations show that the positive charge is not localized exclusively on the nitrogen of the ammonium group, but is partially delocalized on the adjacent atoms. For compound I, these partially charged H atoms are those of $N_2(N_{2'})$, $N_3(N_{3'})$, and $N_4(N_{4'})$. For II, in addition to the strongly positively charged H atom of $N_2(N_{2'})$, H atoms of the methyl groups linked to N_3 and N_4 are also partially charged, although significantly weaker. These charged hydrogen atoms of the side chains play an essential role in interactions with the oligonucleotides, which stabilize the complexes.

We may consider from that point of view the *differences* in these interactions which produce a preference of both compounds for (1) major groove intercalation, (2) interaction to GC sequences, and (3) which are responsible for better intercalative interaction of bisantrene over LC 230487.

Major Groove Versus Minor Groove Binding

The two side chains of each drug behaving in a symmetrical fashion, we may limit our attention to the behavior of one of them in each case.

Major Groove Binding

Side chain A of I forms two hydrogen bonds: one between H of N_3 and the ionic oxygen O_1 of the central phosphate group $P2(d_{H-O} = 1.77 - 1.80$ Å), and another between H of N_2 and the same O_1 of $P2(d_{H-O} = 2.50 - 2.53$ Å). Some interactions between compound I and base pairs are also observed: between H of N_2 and N_7 of the purine G3 or A3 of the intercalation site, between H of N_4 and N_7 of G_3, and between N_1 and H of the amino group of A3. Due to longer distances between the corresponding atoms (d_{H-N7} or $d_{N-H} = 3.08 - 3.48$ Å), these interactions are weaker.

Compound II interacts with the tetramers in a less efficient way than I. In particular, only one hydrogen bond involving the phosphate group is found for each side chain: it occurs between an H of C_3 and the ionic oxygen O_1 of the phosphate group $P2(d_{H-O} = 1.68 \sim 1.72$ Å). The distances between H of N2, a strongly charged atom, and the ionic oxygen of P2 are too long ($d_{H-O} = 3.2 - 3.4$ Å) to form a normal hydrogen bond. For $d(GCGC)_2$, two additional bonds are established: the first involves an H of C_4 and N_7 of G3 of the intercalation site ($d_{H-N7} = 2.80$ Å): the second an H of C4 and O_6 of G3 ($d_{H-O6} = 2.64$ Å). For $d(ATAT)_2$, due to the repulsion of the amino group of adenine, H atoms of C_4 of the side chain cannot come close enough to N_7 of the base for an effective bond to be formed.

Minor Groove Binding

For both I and II, each side chain forms two hydrogen bonds with electron-rich sites in this groove. For I, they occur between the H of N_2 and $O_{1'}$ of the deoxyribose

linked to bases G3 or A3 (d_{H-O} = 2.16 Å), and in an elongated fashion between H of N_3 and $O_{3'}$, of deoxyribose linked to bases C2 or T2 (d_{H-O} = 2.96 Å). For II, the distances between the corresponding atoms H(C3) and H(N2) with $O_{3'}$, (S—C2 or S—T2) and $O_{1'}$, (S—G3 or S—G3) are 2.17 Å and 2.41 Å, respectively.

For both I and II, the binding energies in the minor groove are considerably disfavored with respect to those in the major groove.

Binding to GC Versus AT Sequences

Considering the most efficient bindings of both drugs with these sequences which occur through the major groove, we observe that the preference for binding to the GC oligomer springs both from a better interaction of the side chains with the phosphates of that sequence and their parallel better interaction with the nucleic acid bases. This last result is due essentially to the greater attraction exerted by N7 of guanine over N7 of adenine for electrophilic reactants due to the greater value of the molecular electrostatic potential at the former position [29].

Binding of I Versus Binding of II

The less efficient intercalative binding of II as compared to I, considered with the best receptor, $d(GCGC)_2$, is accountable for by the less intimate interaction of the former both with the ribose phosphate backbone and with the bases of the oligonucleotide. It may be worth stressing that while the principal hydrogen bonds of I involve NH donating groups, those of II involve only CH donating groups, which although activated by the delocalization of the positive charge upon them do not produce as effective hydrogen bonds. Altogether the network of H bonds established by II is significantly restricted with respect to that of I. In particular, no strong H bond could be found for II, which would involve its N_2 (or $N_{2'}$) hydrogen.

Conclusion

The present results indicate that although both compounds bisantrene and LC 230487 are capable of intercalative binding to double-stranded oligonucleotides, with the same sequence specificity, they differ in the detailed pattern of the complexation to the point of producing, for this type of binding, a significantly greater energy of complexation for bisantrene.

On the basis of this result it could be tempting to assume that the existence of antitumor activity in I and its absence in II, may be related to this difference in the pattern and strength of intercalative binding. That the story may, however, be more complex or even substantially different is evident if we recall the experimental result, quoted in the introduction to this paper, which indicates nearly identical measured affinity constants for the interaction with DNA of LC 230487 and of bisantrene. In view of our theoretical result, predicting a substantial difference in the affinity of these two compounds for the intercalative binding, a plausible suggestion seems to be that this situation may signify a possible involvement of LC 230487 in a mode of binding other than intercalation, for example an exterior association with the DNA backbone which does not confer antitumor activity. In these circumstances the presence of this activity in bisantrene, and its absence in LC 230487 could be attributed

to differences in binding *mode,* the presence of the activity being dependent on the presence or dominance of the intercalative interaction.

The present case would not constitute the unique example of closely related compound, showing a similar affinity constant for DNA but a very different behavior as antitumor agents. An outstanding example of a similar situation is offered by *m*-AMSA, a potent chemotherapeutic drug and *o*-AMSA, inactive in this respect [30], as well as by other compounds of the acridine drugs series [15, 31]. We have undertaken in our laboratory an exploration of the possible competitive exterior binding of a number of essentially intercalative drugs.

Acknowledgments

Bibliography

[1] J. W. Lown, *Molecular Aspects of Anti-Cancer Drug Action,* S. Neidle and M. J. Waring, Eds. (Verlag, Chemie, Weiheim, 1983), pp. 283–314.

[2] J. W. Lown, Adv. Free Rad. Biol. Med. **1,** 225–264 (1985).

[3] J. W. Lown, R. Reszka, P. Kolodziejczyk, and W. D. Wilson, *Molecular Mechanisms of Carcinogenic and Antitumor Activity,* B. Pullman and C. Chagas, Eds. (The Vatican Press, in press).

[4] J. W. Lown, S. N. Sonatic, S. F. Yen, J. A. Plambeck, H. H. Peters, E. M. Acton, and G. A. Gordon, Drugs Exptl. Clin. Res. **11,** 735–744 (1984).

[5] H. D. H. Showalter, D. W. Fry, W. R. Leopold, J. W. Lown, J. A. Plambeck, and K. Reszka, Anti-Cancer Drug Design **1,** 73–85 (1986).

[6] H. D. H. Showalter, J. L. Johnson, J. M. Hoftiezer, W. R. Turner, L. M. Werbel, W. R. Leopold, J. L. Shillic, R. C. Jackson, and E. L. Elslager, J. Med. Chem. **30,** 121–131 (1987).

[7] K. X. Chen, N. Gresh, and B. Pullman, Nucleic Acid. Res. **14,** 3799–3812 (1986).

[8] K. X. Chen, N. Gresh, and B. Pullman, *Anti-Cancer Drug Design,* **2,** 79–84 (1987).

[9] B. Pullman, in *Anthracycline and Anthracedione-Based Anticancer Agents,* J. W. Lown Ed. (Elsevier Science Publishers, Amsterdam, in press).

[10] K. C. Murdick, R. G. Child, Y. Lin, J. D. Warren, P. F. Fabio, V. J. Lee, P. T. Izzo, S. A. Lang, R. B. Angier, R. V. Citarella, R. E. Walker, and F. E. Durr, J. Med. Chem. **25,** 505–508 (1982).

[11] G. T. Bowden, D. Garcia, Y-M Pengard, and D. A. Alberts, Cancer Res. **42,** 2660–2665 (1982).

[12] J. W. Lown, C. C. Hanstock, R. D. Bradley, and D. G. Scraba, Molec. Pharmacol. **25,** 178–184 (1984).

[13] W. O. Foye, P. S. Karnik, and S. K. Sengupta, Anti-Cancer Drug Design **1,** 65–71 (1986).

[14] R. V. Citarella, R. E. Wallau, K. C. Murdock, R. B. Angier, F. E. Durr, and M. Forbes, Cancer Res. **42,** 440–444 (1982).

[15] J. Feigon, W. A. Denny, W. Leupin, and D. R. Kearns, J. Med. Chem. **27,** 450–465 (1984).

[16] K. Miller, *Proceedings of the Second SUNYA Conversation in the Discipline Biomolecular Stereodynamics,* Vol. 2, R. H. Sarma, Ed. (Adenine Press, New York, 1981), pp. 469–486.

[17] K. Miller and J. Pycior, Biopolymers **18,** 2683–2719 (1979).

[18] K. X. Chen, N. Gresh, and B. Pullman, J. Biomol. Struct. Dynamics **3,** 445–466 (1985).

[19] K. X. Chen, N. Gresh, and B. Pullman, Nucl. Acid Res. **14,** 2251–2267 (1986).

[20] K. X. Chen, N. Gresh, and B. Pullman, Molec. Pharmacol. **30**, 279–286 (1986).
[21] K. X. Chen, N. Gresh, and B. Pullman, Nucl. Acid Res. **14**, 9103–9115 (1986).
[22] K. X. Chen, N. Gresh, and B. Pullman, Nucl. Acid Res. **15**, 2175–2189 (1987).
[23] N. Gresh, P. Claverie, and A. Pullman, Theoret. Chim. Acta **66** 1–20 (1984).
[24] N. Gresh, P. Claverie, and A. Pullman, Int. J. Quantum Chem. **29**, 101–118 (1986).
[25] A. Pullman, K. Zakrzewska, and D. Perahia, Int. J. Quantum Chem. **16**, 395–403 (1979).
[26] S. Arnott and D. Hukins, J. Molec. Biol. **81**, 93–105 (1973).
[27] B. Pullman, N. Gresh, H. Berthod, and A. Pullman, Theoret. Chim. Acta **44**, 151–163 (1977).
[28] M. Powell, Report CSS, 15 AERE Harwell Library, Routine VA13A (1975).
[29] A Pullman and B. Pullman, Quart Rev. Biophys. **14**, 289–380 (1981).
[30] W. R. Wilson, G. F. Whitmore, and R. P. Hall, Cancer Res. **41**, 2817–2821 (1981).
[31] See, e.g., W. A. Denny, B. C. Baguley, B. F. Cain, and M. J. Waring, in *Molecular Aspects of Anti-Cancer Drug Action*, S. Neidle and M. J. Waring, Eds. (Verlag Chemie, Weinheim, 1983), pp. 1–34.

Basis Set and Correlation Effects on Computed Negative Ion Hydrogen Bond Energies of the Complexes $AH_n \cdot AH_{n-1}^{-1}$: $AH_n = NH_3$, OH_2, and FH

JANET E. DEL BENE

Department of Chemistry, Youngstown State University, Youngstown, Ohio 44555, U.S.A.

Abstract

Basis set and correlation effects on computed hydrogen bond energies of the negative ion complexes $AH_n \cdot AH_{n-1}^{-1}$, for $AH_n = NH_3$, OH_2, and FH, have been evaluated. The addition of diffuse functions on nonhydrogen atoms to valence double- and triple-split plus polarization basis sets [6-31G(d, p) and 6-311G(d, p)] significantly decreases binding energies by 9–19 kcal/mol, depending on the particular complex and the level of theory. Adding diffuse functions to hydrogens has a negligible effect, while replacing the single set of polarization functions on each atom by two sets alters energies by 1 kcal/mol or less. Electron correlation increases the hydrogen bond energies of these complexes and has a greater effect for basis sets without diffuse functions. Since the Hartee-Fock energies computed with these basis sets are already too large, correlation energy calculations should not be performed in these cases. For basis sets including diffuse functions, the correlation energy contribution to the binding energies of these complexes is significant, with the Møller-Plesset second-order term being the largest term and having a stabilizing effect of from 3–6 kcal/mol. The third and fourth order terms are smaller, and may be of opposite sign. As a result, the MP2 and MP4 energies differ by no more than 1 kcal/mol, with the MP2 stabilization energy being greater except for $N_2H_5^{-1}$. The computed standard solvation enthalpy of OH^{-1} by H_2O based on either $MP4/6-311+G(2d, 2p)$ or $MP2/6-31+G(d, p)$ electronic energies is -26.8 kcal/mol, in excellent agreement with a recent gas-phase experimental measurement.

Introduction

Many studies have demonstrated that calculated *ab initio* association energies are dependent on the basis set and the theoretical method used for the calculations [1-28]. The specific details of these dependencies vary with the particular type of interaction. In an attempt to characterize these dependencies, systematic studies have been undertaken in this laboratory of the effects on binding energies of augmenting split-valence plus polarization basis sets, and of electron correlation at various levels of Møller-Plesset perturbation theory. The first studies were concerned with the effects of basis set and electron correlation on computed protonation and lithium ion association energies for a series of oxygen and nitrogen bases [14]. Subsequent studies examined basis set and correlation effects on the computed binding energies of neutral and positive ion hydrogen-bonded complexes $(AH_n)_2$ and $AH_n \cdot AH_{n+1}^{+1}$, for $AH_n = NH_3$, OH_2, and FH [27, 28]. The present study is an analogous investigation of basis set and correlation effects on the corresponding negative ion complexes

$AH_n \cdot AH_{n-1}^{-1}$ (alternatively, $A_2H_{2n-1}^{-1}$) for the same hydrides. Previous studies of the structures and energies of such negative ion complexes at various levels of theoretical treatment have been reported [29–40]. The performance of a given level of theory will be judged by comparison with higher levels of theory, and for the case of $O_2H_3^{-1}$, by comparison with a recent experimental measurement.

Method

The standard enthalpy at 298 K for the reaction involving the solvation of the anion by the corresponding solvent is defined as ΔH^{298} for the gas-phase reaction

$$AH_n + AH_{n-1}^{-1} \rightarrow AH_n \cdot AH_{n-1}^{-1} \tag{1}$$

The electronic contribution to this enthalpy is defined as ΔE_e, for the same reaction. In this study, electronic energies have been calculated using Hartree-Fock (HF) theory, with correlation included using Møller-Plesset perturbation theory at second (MP2), third (MP3), and fourth order (MP4) [41,42]. The correlation energy calculations exclude the 1s electrons on nonhydrogen atoms.

The electronic energies have been calculated using basis sets derived from a split-valence plus polarization basis [6-31$G(d,p)$] [43,44], and a valence triple-split plus polarization basis [6-311$G(d,p)$] [45]. These basis sets were augmented with diffuse functions on nonhydrogen atoms, giving 6-31+$G(d,p)$ and 6-311+$G(d,p)$ [24,25], and diffuse functions on all atoms [6-31++$G(d,p)$ and 6-311++$G(d,p)$]. Another enhancement was the replacement of the standard set of d functions on nonhydrogen atoms by two sets, with exponents a factor of 2 larger and smaller than the standard value [6-31$G(2d,p)$ and 6-311$G(2d,p)$], and a similar change in the first polarization space of hydrogen atoms [6-31$G(2d,2p)$ and 6-311$G(2d,2p)$]. Basis sets derived from combinations of these enhancements were also employed. No counterpoise estimates of basis set deficiencies have been made in this work. Recent studies have suggested that the counterpoise correction does not improve reliability, and probably should not be done [19,22]. It was noted that in addition to those basis set superposition effects for which the counterpoise correction is intended to compensate (or overcompensate), there are more fundamental deficiences in double-polarized basis sets which have important effects on the predicted energetics of hydrogen-bonded systems. The intent in the present study is to evaluate the performance of "uncorrected" basis sets. It is expected that superposition errors should be small for the larger basis sets used. The good agreement between the computed and the experimental enthalpy for reaction (1) for $O_2H_3^{-1}$ supports this expectation.

All calculations reported in this work were carried out on the optimized Hartree-Fock 6-31$G(d)$ geometries of the monomers and complexes. For F_2H^{-1} and $N_2H_5^{-1}$, calculations were performed on the equilibrium structures of $D_{\infty h}$ and C_1 symmetry, respectively. For $O_2H_3^{-1}$, both the Hartree-Fock equilibrium C_1 structure and a structure of C_2 symmetry have been investigated. The latter structure has a symmetrically bound proton, and is energetically similar to the C_1 structure at correlated levels of theory. Since the structures of hydrogen bonded complexes also show some dependence on the theoretical method [22], Hartree-Fock structures for all complexes $AH_n \cdot AH_{n-1}^{-1}$ have also been obtained with the 6-31+$G(d)$ basis set. Although some

structural differences are found, these do not appear to have large effects on the computed stabilization energies. Thus, for reaction (1), corresponding MP2/6-31+$G(d,p)$ energies computed at HF/6-31$G(d)$ and HF/6-31+$G(d)$ geometries differ by 0.1, 0.3, and 1.1 kcal/mol for F_2H^{-1}, $O_2H_3^{-1}$, and $N_2H_5^{-1}$, respectively.

Results and Discussion

Computed electronic negative ion hydrogen bond energies at various levels of theory are reported in Table I. Relationships derived from the data of Table I are summarized below. The first seven relationships refer to basis set effects, the last two to correlation effects.

TABLE I.   Computed electronic negative ion hydrogen bond energies (kcal/mol)[a]

Basis set	HF	MP2	MP3	MP4SDQ[b]	MP4
		F_2H^{-1}			
6-31$G(d,p)$	−57.2	−64.1	−62.4	−62.6	−63.4
6-31+$G(d,p)$	−42.5	−45.2	−45.5	−44.4	−44.4
6-31++$G(d,p)$	−42.5	−45.2	−45.5	−44.4	−44.5
6-31$G(2d,p)$	−56.4	−63.4	−61.4	−61.6	−62.7
6-31$G(2d,2p)$	−56.4	−63.8	−62.0	−62.2	−63.3
6-31+$G(2d,2p)$	−41.3	−44.0	−44.4	−43.2	−43.3
6-311$G(d,p)$	−51.6	−58.1	−56.6	−56.8	−57.7
6-311+$G(d,p)$	−41.6	−44.3	−44.8	−43.6	−43.6
6-311++$G(d,p)$	−41.6	−44.3	−44.8	−43.6	−43.6
6-311$G(2d,p)$	−52.5	−59.5	−57.9	−58.0	−59.1
6-311$G(2d,2p)$	−52.1	−59.0	−57.2	−57.4	−58.6
6-311+$G(2d,2p)$	−41.5	−44.8	−45.1	−44.0	−44.3
		$O_2H_3^{-1} - C_1$			
6-31$G(d,p)$	−35.1	−42.2	−40.1	−40.4	−41.3
6-31+$G(d,p)$	−24.1	−27.7	−27.6	−26.8	−27.2
6-31++$G(d,p)$	−24.2	−27.9	−27.7	−27.0	−27.5
6-31$G(2d,p)$	−34.5	−42.0	−39.8	−40.0	−41.2
6-31$G(2d,2p)$	−34.6	−42.8	−40.6	−40.9	−42.1
6-31+$G(2d,2p)$	−23.4	−27.5	−27.4	−26.6	−27.3
6-311$G(d,p)$	−34.6	−42.8	−40.6	−41.0	−42.2
6-311+$G(d,p)$	−23.6	−27.7	−27.8	−26.9	−27.4
6-311++$G(d,p)$	−23.7	−27.9	−27.9	−27.1	−27.6
6-311$G(2d,p)$	−34.7	−43.0	−40.6	−40.8	−42.3
6-311$G(2d,2p)$	−34.3	−42.8	−40.4	−40.7	−42.3
6-311+$G(2d,2p)$	−23.3	−27.8	−27.8	−27.0	−27.7
		$O_2H_3^{-1} - C_2$			
6-31$G(d,p)$	−34.5	−43.3	−40.8	−41.1	−42.1
6-31+$G(d,p)$	−22.6	−27.9	−27.4	−26.6	−27.1
6-31++$G(d,p)$	−22.6	−28.1	−27.6	−26.7	−27.3
6-31$G(2d,p)$	−33.7	−43.0	−40.2	−40.4	−41.8
6-31$G(2d,2p)$	−33.7	−43.8	−41.1	−41.3	−42.8
6-31+$G(2d,2p)$	−21.8	−27.6	−27.1	−26.2	−27.1

TABLE I.   *(Continued)*

Basis set	HF	MP2	MP3	MP4SDQ[b]	MP4
$6\text{-}311G(d,p)$	−33.5	−43.7	−41.1	−41.5	−43.0
$6\text{-}311+G(d,p)$	−22.0	−28.0	−27.8	−26.8	−27.5
$6\text{-}311++G(d,p)$	−22.0	−28.2	−27.9	−26.9	−27.6
$6\text{-}311G(2d,p)$	−33.8	−44.1	−41.3	−41.5	−43.2
$6\text{-}311G(2d,2p)$	−33.2	−43.4	−40.5	−40.7	−42.6
$6\text{-}311+G(2d,2p)$	−21.7	−27.9	−27.5	−26.5	−27.5
		$N_2H_5^{-1c}$			
$6\text{-}31G(d,p)$	−19.4	−24.5	−23.2	−23.3	−24.0
$6\text{-}31+G(d,p)$	−10.8	−13.7	−13.5	−13.1	−13.6
$6\text{-}31G(2d,p)$	−18.8	−24.4	−23.0	−22.9	−23.8
$6\text{-}31G(2d,2p)$	−19.2	−25.4	−23.9	−23.9	−24.9
$6\text{-}31+G(2d,2p)$	−9.9	−13.4	−13.2	−12.9	−13.5
$6\text{-}311G(d,p)$	−19.6	−25.6	−24.1	−24.2	−25.2
$6\text{-}311+G(d,p)$	−10.3	−13.7	−13.5	−13.1	−13.7
$6\text{-}311G(2d,p)$	−19.4	−25.8	−24.1	−24.1	−25.2
$6\text{-}311G(2d,2p)$	−19.6	−26.7	−24.9	−25.0	−26.3
$6\text{-}311+G(2d,2p)$	−9.7	−13.5	−13.4	−13.0	−13.7

[a]Total energies are available from the author on request.

[b]Fourth order Møller-Plesset energy including all single, double, and quadruple substitutions, but omitting triple substitutions.

[c]Calculations employing basis sets with diffuse functions on hydrogens were not carried out for this complex.

Basis Set Effects

Changing from the Split-Valence $6\text{-}31G(d,p)$ Basis Set to the Valence Triple-Split $6\text{-}311G(d,p)$ Basis. Expanding from a valence double-split to a valence triple-split basis set has a relatively small effect on the stabilization energies of the complexes $O_2H_3^{-1}$ and $N_2H_5^{-1}$, which change by only 1 kcal/mol or less at Hartree-Fock and correlated levels of theory. In contrast, this same enhancement decreases the hydrogen bond energy of F_2H^{-1} by 5–6 kcal/mol. The large effect in this case may be attributed to the limitations of the double-split valence basis set for describing the F atom.

Adding Diffuse Functions on Nonhydrogen Atoms. The addition of diffuse functions on nonhydrogen (heavy) atoms is once again found to be the single most important enhancement of valence double- and triple-split basis sets for computing hydrogen bond energies. This enhancement dramatically decreases the binding energies of all complexes at all levels of theory, with a larger effect observed with correlation than at Hartree-Fock. With the double-split basis set, the energy lowering ranges from 8.6 kcal/mol for $N_2H_5^{-1}$ at Hartree-Fock to 19.0 kcal/mol for F_2H^{-1} at MP4. With the triple-split basis set, the lowering ranges from 9.3 kcal/mol at Hartree-Fock for $N_2H_5^{-1}$ to 15.7 kcal/mol at MP2 for the C_2 form of $O_2H_3^{-1}$. At all levels of theory the energy lowering is greater for the more symmetric C_2 form of $O_2H_3^{-1}$ than for the C_1 form. The decrease in the stabilization energies of these negative ion hydrogen-

bonded complexes is greater than that observed in the corresponding neutral and positive ion hydrogen-bonded complexes. The importance of diffuse functions for describing interactions involving negative ions has been noted previously by others, and is attributed to the ability of diffuse functions to better describe the tails of functions associated with the lone pairs of electrons on atoms which are the basic sites for bond formation [13, 22, 24–26].

Adding Diffuse Functions on Hydrogens. Addition of diffuse functions to hydrogen atoms has little effect on the stabilization energies of these complexes. The presence of such functions increases stabilities but by no more than 0.3 kcal/mol at any level of theory.

Since the $6-31+G(d,p)$ and $6-311G(d,p)$ basis sets are the same size, it is appropriate to compare the stabilization energies computed with each. Corresponding Hartree-Fock and Møller-Plesset hydrogen bond energies obtained with the $6-31+G(d,p)$ basis set are lower by 9–16 kcal/mol than those computed with the $6-311G(d,p)$ basis, and are in better agreement with the energies computed with the largest basis set employed in this study. Since the presence of diffuse functions on hydrogen atoms has essentially no effect on stabilization energies, these data demonstrate that it is important to enhance the $6-31G(d,p)$ basis set by adding diffuse functions on nonhydrogen atoms rather than triple-split the valence shell when computing the stabilization energies of negative ion hydrogen-bonded complexes. The large decrease in stabilization energies due to the presence of diffuse functions argues for their absolute necessity for describing interactions involving negatively charged ions.

Splitting d Functions on Nonhydrogen Atoms into Two Sets. Enhancing the $6-31G(d,p)$ basis set by replacing the set of d polarization functions on nonhydrogen atoms by two sets decreases stabilization energies by 1 kcal/mol or less at all levels of theory. In contrast, this same enhancement of the $6-311G(d,p)$ basis set increases the stabilization energies of F_2H^{-1} by about 1 kcal/mol, but has little effect on the hydrogen bond energies of $N_2H_5^{-1}$ and $O_2H_3^{-1}$.

Splitting p Functions on Hydrogens. The effect of replacing the standard set of p functions on hydrogens by two sets depends on the starting basis set and on the particular complex. This enhancement of the $6-31G(d,p)$ basis has little effect on the stabilization energies of these negative ion hydrogen bonded complexes, increasing them by 1 kcal/mol or less at all levels of theory. With the $6-311G(d,p)$ basis the splitting of the p functions decreases the hydrogen bond energies of F_2H^{-1} and $O_2H_3^{-1}$ but increases the hydrogen bond energies of $N_2H_5^{-1}$. These changes do not exceed 1 kcal/mol.

$6-31+G(2d,2p)$ versus $6-311+G(2d,2p)$. The enhancement of both the valence double- and triple-split basis sets by the addition of diffuse functions on heavy atoms and the splitting of the polarization functions on all atoms into two sets results in computed $6-31+G(2d,2p)$ and $6-311+G(2d,2p)$ hydrogen bond energies for each complex which are similar at corresponding levels of theory. The $6-311+G(2d,2p)$ energy of F_2H^{-1} is always greater than the corresponding $6-311+G(2d,2p)$ energy, but the difference does not exceed 1 kcal/mol. Corresponding differences for the other complexes are less than 0.5 kcal/mol. Moreover, the differences between corresponding $6-31+G(d,p)$ and $6-311+G(2d,2p)$ energies are about 1 kcal/mol at

Hartree-Fock, and no more than 0.5 kcal/mol at correlated levels of theory. Given the computational expense of the larger basis set and the similarity of results, the $6\text{-}31+G(d,p)$ basis would appear to be the basis set of choice for studies of negative ion hydrogen bond energies in larger systems.

Additivity. The specific changes in the computed $6\text{-}31G(d,p)$ and $6\text{-}311G(d,p)$ hydrogen bond energies will be equal to those computed directly with the $6\text{-}31+$ $G(2d,2p)$ and $6\text{-}311+G(2d,2p)$ basis sets. For these negative ion hydrogen-bonded may be added to the hydrogen bond energies computed with each of these two starting basis sets, respectively. If basis set enhancement effects are additive, the resulting hydrogen bond energies will be equal to those computed directly with the $6\text{-}31+$ $G(2d,2p)$ and $6\text{-}311+G(2d,2p)$ basis sets. For these negative ion hydrogen-bonded complexes, the sum of the $6\text{-}31G(d,p)$ hydrogen bond energy and the specific changes in the stabilization energy brought about by each enhancement approximates the computed $6\text{-}31+G(2d,2p)$ hydrogen bond energy to within 1.2 kcal/mol at all levels of theory. With the $6\text{-}311G(d,p)$ basis as the starting basis set, the approximate values are within 1.1 kcal/mol of the computed $6\text{-}311+G(2d,2p)$ hydrogen bond energies. It should be recognized that the changes in the hydrogen bond energies brought about by various enhancements to the basis sets are dominated by the addition of diffuse functions to nonhydrogen atoms. All other enhancements have relatively minor effects, and may be in opposite directions. Nevertheless, it is interesting to note that the same enhancements of these basis sets were more nearly additive in the corresponding positive ion hydrogen bonded complexes, where differences of only 0.8 and 0.2 kcal/mol were found between estimated and calculated stabilization energies with the $6\text{-}31+G(2d,2p)$ and $6\text{-}311+G(2d,2p)$ basis sets, respectively [28].

Correlation Effects

It is apparent from the data presented thus far that diffuse functions in the basis set are absolutely necessary for adequately describing the stabilization energies of these negative ion hydrogen-bonded complexes. The presence of such functions leads to significant decreases in binding energies. In contrast, the overall effect of electron correlation is to increase the binding energies of these complexes, with the increase being significantly larger when computed with basis sets without diffuse functions. Thus for these basis sets, correlation further increases the stabilization energies which are already too large at Hartree-Fock. Therefore, correlation calculations should not be done in these cases. In the following analyses of the effects of electron correlation on stabilization energies, comments will be limited to trends observed for basis sets which contain diffuse functions on nonhydrogen atoms.

Inclusion of Correlation. The correlation energy contribution to the hydrogen bond energy leads to an increase in the binding energies of all complexes. The correlation contribution varies with the particular complex, and is a significant part of the total stabilization energy. At MP4/$6\text{-}311+G(2d,2p)$ it ranges from 2.8 kcal/mol (7% of the Hartree-Fock energy) for F_2H^{-1}, to 5.8 kcal/mol (27%) for the C_2 form of $O_2H_3^{-1}$, and 4.0 kcal/mol (41%) for $N_2H_5^{-1}$.

Hydrogen Bond Energies at Various Levels of Correlation. The second-order contribution to the correlation energy increases the stability of all complexes, and is significantly larger than any other correlation term. The MP2 correlation contribution ranges from about 3 kcal/mol for F_2H^{-1} to 6 kcal/mol for the C_2 form of $O_2H_3^{-1}$. The MP3 and MP4 correlation energy contributions to the hydrogen bond energies are smaller than the MP2 contribution, and may be of the same or opposite sign, depending on the particular complex. The MP2 and MP4 stabilization energies of each complex differ by no more than 1 kcal/mol when computed with basis sets derived from either the $6\text{-}31G(d,p)$ or $6\text{-}311G(d,p)$ basis sets. In general, the stabilization energy computed at MP2 is larger than at MP4, except for $N_2H_5^{-1}$, in which case the MP4 energies may be slightly larger. The triples contribution to the fourth-order term stabilizes these complexes by 1 kcal/mol or less.

Comparisons

An extensive study of F_2H^{-1} has been reported [22] recently in which the optimized geometry of this complex was computed with correlation at MP2 employing the $6\text{-}311++G(2d,2p)$ basis set. A single-point MP4 calculation with the $6\text{-}311++G(3df,3pd)$ basis set was then carried out. This basis set contains three sets of d functions and a set of f functions on nonhydrogen atoms, and three sets of p functions and a set of d functions on hydrogens. The computed stabilization energy is -45.6 kcal/mol. This compares with the $MP4/6\text{-}311+G(2d,2p)$ energy of -44.3 kcal/mol obtained in the present study.

The experimental value of the gas-phase reaction enthalpy for the solvation of OH^{-1} by H_2O has been measured recently, and found to be -26.8 kcal/mol [46]. Combining the $MP4/6\text{-}311+G(2d,2p)$ electronic association energy for the Hartree-Fock equilibrium C_1 structure of $O_2H_3^{-1}$, the zero-point and thermal vibrational energy corrections at $HF/6\text{-}31G(d)$ for this structure, and the remaining thermal terms [47], leads to the same computed enthalpy of -26.8 kcal/mol. The enthalpy computed from the $MP2/6\text{-}31+G(d,p)$ energy is also -26.8 kcal/mol.

Conclusions

In this study, basis set and correlation effects on the computed hydrogen bond energies of the negative ion hydrogen-bonded complexes $AH_n \cdot AH_{n-1}^{-1}$ for $AH_n = NH_3$, OH_2, and FH, have been evaluated. The addition of diffuse functions on nonhydrogen atoms to valence double- and triple-split plus polarization basis sets is absolutely necessary for adequately describing binding energies, which are decreased by 9–19 kcal/mol, depending on the particular complex and the level of theory. The effect is greater with correlation than at Hartree-Fock. Addition of diffuse functions to hydrogen atoms has a negligible effect on stabilization energies, while replacing the single set of polarization functions on each atom by two sets alters energies by 1 kcal/mol or less. In contrast to the energy-lowering effects of augmenting the $6\text{-}31G(d,p)$ and $6\text{-}311G(d,p)$ basis sets, the overall effect of electron correlation is to increase hydrogen bond energies. Thus, for basis sets without diffuse

functions, correlation further increases binding energies which are already too large at Hartree-Fock. Therefore, correlation calculations should not be done in these cases. For basis sets including diffuse functions, the correlation energy contribution to the hydrogen bond energies of these complexes is significant, with the Møller-Plesset second-order term being the largest term and having a stabilizing effect of from 3–6 kcal/mol. The third- and fourth-order terms are smaller and may be of the opposite sign. As a result, the MP2 and MP4 energies differ by no more than 1 kcal/mol, with the MP2 energy being greater except for $N_2H_5^{-1}$. The computed standard solvation enthalpy of OH^{-1} by H_2O based on either $MP2/6\text{-}31+G(d,p)$ or $MP4/6\text{-}311+G(2d,2p)$ electronic energies is -26.8 kcal/mol, in excellent agreement with a recent gas-phase measurement.

Acknowledgments

References

[1] G. H. F. Diercksen, W. P. Kraemer, and B. O. Roos, Theor. Chim. Acta. 36, 249 (1975).

[2] R. L. Woodin, F. A. Houle, and W. A. Goddard, Chem. Phys. 14, 461 (1976).

[3] R. Kari and I. G. Csizmadia, Int. J. Quan. Chem. 11, 441 (1977).

[4] P. Kollman and S. Rothenberg, J. Am. Chem. Soc. 99, 1333 (1977).

[5] H. Berthod and A. Pullman, Israel J. Chem. 19, 299 (1980).

[6] R. A. Eades, R. Scanlon, M. R. Ellenberger, D. A. Dixon, and D. S. Marynick, J. Phys. Chem. 84, 2840 (1980).

[7] M. J. Frisch, J. E. Del Bene, K. Raghavachari, and J. A. Pople, Chem. Phys. Lett. 83, 240 (1981).

[8] R. A. Eades, D. A. Weil, D. A. Dixon, and C. H. Douglass, J. Phys. Chem. 85, 981 (1981).

[9] S. F. Smith, J. Chandrasekhar, and W. L. Jorgensen, J. Phys. Chem. 86, 3308 (1982).

[10] J. E. Del Bene, M. J. Frisch, K. Raghavachari, and J. A. Pople, J. Phys. Chem. 86, 1529 (1982).

[11] J. E. Del Bene, Chem. Phys. Lett. 94, 213 (1983); J. Comp. Chem. 5, 381 (1984).

[12] J. E. Del Bene, M. J. Frisch, K. Raghavachari, J. A. Pople, and P. v. R. Schleyer, J. Phys. Chem. 87, 73 (1983).

[13] Z. Latajka and S. Scheiner, Chem. Phys. Lett. 105, 435 (1984); J. Chem. Phys. 81, 407 (1984).

[14] J. E. Del Bene, J. Comp. Chem. 6, 296 (1985); 7, 259 (1986).

[15] M. J. Frisch, J. A. Pople, and J. E. Del Bene, J. Phys. Chem. 89, 3664 (1985).

[16] J. E. Del Bene, M. J. Frisch, and J. A. Pople, J. Phys. Chem. 89, 3669 (1985).

[17] M. D. Newton and N. R. Kestner, Chem. Phys. Lett. 94, 198 (1983).

[18] N. R. Kestner, M. D. Newton, and T. L. Mathers, Int. J. Quantum Chem. Symp. 17, 431 (1983).

[19] D. W. Schwenke and D. G. Truhlar, J. Chem. Phys. 82, 2418 (1985).

[20] J. F. Gaw, Y. Yamaguchi, M. A. Vincent, and H. F. Schaefer III, J. Am. Chem. Soc. 106, 3133 (1984).

[21] D. W. Michael, C. E. Dykstra, and J. M. Lisy, J. Chem. Phys. 81, 5998 (1984).

[22] M. J. Frisch, J. E. Del Bene, J. S. Binkley, and H. F. Schaefer III, J. Chem. Phys. 84, 2279 (1986).

[23] D. J. DeFrees and A. D. McLean, J. Comp. Chem. 7, 321 (1986).

[24] T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. v. R. Schleyer, J. Comp. Chem. 4, 294 (1983).

[25] G. W. Spitznagel, T. Clark, J. Chandrasekhar, and P. v. R. Schleyer, J. Comp. Chem. 3, 363 (1982).

[26] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, Ab Initio Molecular Orbital Theory, (John Wiley and Sons, New York, 1986).

[27] J. E. Del Bene, J. Chem. Phys. 86, 2110 (1987).

[28] J. E. Del Bene, J. Comp. Chem. (in press).
[29] P. A. Kollman and L. C. Allen, J. Am. Chem. Soc. **92**, 6101 (1970).
[30] M. D. Newton and S. Ehrenson, J. Am. Chem. Soc. **93**, 4971 (1971).
[31] W. P. Kraemer and G. H. F. Diercksen, Theor. Chim. Acta. **23**, 398 (1972).
[32] A. D. Issacson and K. Morokuma, J. Am. Chem. Soc. **97**, 4453 (1975).
[33] A. Stoegard, A. Strich, J. Almlof, and B. Roos, Chem. Phys. **8**, 405 (1975).
[34] B. O. Roos, W. P. Kraemer, and G. H. F. Diercksen, Theor. Chim. Acta. **42**, 77 (1976).
[35] G. Alagona, E. Scrocco, and J. Tomasi, Theor. Chim. Acta. **47**, 133 (1978).
[36] W. J. Bouma and L. Radom, Chem. Phys. Lett. **64**, 216 (1979).
[37] W. L. Jorgensen and M. Ibrahim, J. Comp. Chem. **2**, 7 (1981).
[38] C. M. Rohlfing, L. C. Allen, C. M. Cook, and H. B. Schlegel, J. Chem. Phys. **78**, 2498 (1983).
[39] S. Ikuta, J. Comp. Chem. **5**, 374 (1984).
[40] J. Gao, D. S. Garner, and W. L. Jorgensen, J. Am. Chem. Soc. **108**, 4784 (1986).
[41] J. A. Pople, J. S. Binkley, and R. Seeger, Int. J. Quantum Chem. Symp. **10**, 1 (1976).
[42] R. Krishnan and J. A. Pople, Int. J. Quantum Chem. **14**, 91 (1976).
[43] P. C. Hariharan and J. A. Pople, Theor. Chim. Acta. **28**, 213 (1973).
[44] J. D. Dill and J. A. Pople, J. Chem. Phys. **62**, 2921 (1975).
[45] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).
[46] M. Meot-Ner (Mautner), J. Am. Chem. Soc. **108**, 6189 (1986).
[47] J. E. Del Bene, in *Molecular Structure and Energetics*, Vol. 1, J. F. Liebman and A. Greenberg, Eds. (VCH Publishers, Inc. Deerfield Beach, FL, 1986).

Ab Initio MRD-CI Calculations on Protonated Cyclic Ethers

I. Protonation Pathways Involve Multipotential Surfaces (Protonation of Oxetane)
II. Differences from SCF in Dominant Configurations Upon Opening Non-Protonated Oxirane Rings (Epoxides)

JOYCE J. KAUFMAN, P. C. HARIHARAN, SZCZEPAN ROSZAK,* AND
P. B. KEEGSTRA

Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland 21218, U.S.A.

I. Introduction

Cyclic ethers are significant in both nonbiomedical and biomedical areas. Cyclic ethers undergo cationic polymerization. Oxetanes (the four-member rings) substituted with exotic energetic substituents form energetic polymers. The initiation step is governed by the basicity of the cyclic ether.

We have previously carried out *ab initio* MODPOT/VRDDO SCF calculations and generated electrostatic molecular potential contour (EMPC) maps around these energetic-substituted ethers which predicted their propensities to polymerize even prior to the synthesis of the monomers themselves [1]. The subsequent propagation step in cationic polymerization involves the attack of protonated oxetane on oxetane with concomitant opening of the oxetane ring. For opening the oxetane and protonated oxetane rings and for the reactions between them we carried out *ab initio* MRD-CI (multireference double excitation–configuration interaction) calculations (by the method of Buenker and Peyerimhoff, Ref. 2) based on localized molecular orbitals, including explicitly in the CI the localized occupied and virtual molecular orbitals in the reaction region and folding the remainder of the occupied localized molecular orbitals into an effective CI Hamiltonian [3–5].

In the biomedical area, oxetanes have been shown to induce the enzyme, aryl hydrocarbon hydroxylase [6] which is also involved in metabolic activation of polycyclic aromatic hydrocarbon (PAH) carcinogens.

The metabolic activation of polycyclic aromatic hydrocarbons (PAHs) from precarcinogens to proximate carcinogens to ultimate carcinogens involves epoxidation of the PAHs [7] and opening of the epoxide rings either protonated or nonprotonated.

We have previously published detailed *ab initio* MODPOT/VRDDO/MERGE SCF calculations on the potent carcinogens benzo(a)pyrene [BP] and 3-methylcholanthrene and

---

their metabolites (proximate carcinogens: epoxides and dihydrodiols; ultimate car-
cinogens: dihydrodiol epoxides, protonated dihydrodiol epoxides, and their ring-
opened species) [8,9] and on the EMPC maps generated around these molecules
[10-12]. In our original article on SCF calculations on the BP metabolites, we men-
tioned that as the epoxide ring is opened there appears to be a mixing of configurations.

In connection with protonation of PAH epoxides and dihydrodiol epoxides, we also
previously pointed [13] out that even though epoxide + H$^+$ is the reaction between
two closed-shell ground-state species, a single-determinant SCF calculation (even an
ab initio single-determinant SCF calculation) is insufficient to describe properly the
protonation process of any molecule. The reason is that closed-shell molecule + H$^+$
is an ion-molecule reaction. There are two types of ion-molecule reactions:

$$A^+ + B \qquad IP(A) < IP(B) \tag{1}$$
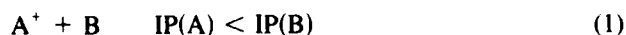
where the ionization potential (IP) of A is less than that of B. In this case there is the
possibility (but not the guarantee) that a single-determinant *ab initio* SCF calculation
could yield the correct potential energy surface behavior.

$$A^+ + B \qquad IP(A) > IP(B) \tag{2}$$
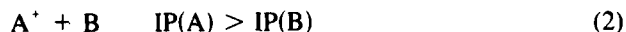
where the ionization potential of A is greater than that of B. (This is the case for all
protonation reactions.) In this case there has to be at least a pair of potential energy
surfaces (singlet and triplet) which arise from the lower energy asymptotic pair
A$^+$ + B. In our original article we sketched schematically these lower surfaces as
being repulsive. The physical argument about protonation being an ion–molecule
reaction of type 2 involving multipotential surfaces is general and applicable to proto-
nation reactions of all closed shell molecules whose ionization potentials are lower
than that of the hydrogen atom.

For the present presentation, in the panel on Carcinogenesis, we carried out *ab ini-
tio* MODPOT MRD-CI calculations to investigate two points: (a) the multipotential
surface character of protonation reactions; (b) the character of the dominant configu-
rations as an epoxide ring is opened.

II. Methodology

The *ab initio* MODPOT/VRDDO/MERGE procedures are detailed in Ref. 14:

MODPOT: *ab initio* effective core model potentials which enable one to treat the va-
lence electrons not only explicitly, yet accurately;

VRDDO: a charge-conserving integral prescreening evaluation procedure which de-
cides whether the magnitude of the integrals in a particular block is suffi-
ciently large to warrant calculating explicitly (which we named
VRDDO — variable retention of diatomic differential overlap) (especially ef-
ficient for spatially extended systems);

MERGE: to save and reuse common skeletal integrals efficiently.

The *ab initio* MRD-CI calculations were carried out by the procedure of Buenker and
Peyerimhoff [2]. First a small CI is run (several hundred to a thousand configura-
tions), from this the most important configurations are chosen as reference configura-
tions (up to 80 configurations can be included as reference configurations), all single

and double excitations are generated relative to all the reference configurations, the energy contribution of each configuration is estimated by a perturbation procedure, all configurations contributing more energy than a certain threshold are included explicitly in the CI calculation (E_{CI}), the energy contribution of all of the configurations not included explicitly in the CI calculation is added back in by a perturbative procedure and extrapolated (E_{EXT}) and finally a Davidson-type correction can be added to account for size consistency (E_{FULL}).

For a typical MRD-CI run for the protonation pathway of oxetane 231,721 CSFs (configuration state functions) were generated from 28 reference configurations and ~4500 CSFs were selected for explicit inclusion in the CI calculation (using a threshold for 10 microhartrees). The MRD-CI calculations were carried out from an R(O-H$^+$) distance of 1.5 bohrs to a distance of 10.0 bohrs. We carried out the MRD-CI calculations for both the linear in-plane (C_{2v}) attack and the out-of-plane C_s attack without imposing any symmetry for the CI wave functions for consistency since even a slight further distortion in the direction of the proton attack will destroy any symmetry. Thus we took 5 roots of the CI matrix. However, while we did not impose symmetry in the CI calculation the results still contain the symmetry information. To facilitate a clear transformation from C_{2v} to C_s symmetry rotations we describe the symmetry operations for both. Commensurate with Herzberg's conventions [15]: (a) For C_{2v} attack the oxetane molecule is in the xz plane with the oxygen along the z axis, and the two reflection planes are σ_{xz} and σ_{yz}. The in-plane orbitals are a_1 and b_1 and the out-of-plane orbitals are a_2 and b_2; (b) For C_s attack the oxetane molecule is in the xz plane this time with the oxygen along the x axis and the reflection plane is ρ_{xy}. The a_1 and b_2 orbitals in C_{2v} symmetry correspond to a' orbitals in C_s symmetry and the b_1 orbital in C_{2v} symmetry corresponds to an a'' orbital in C_s symmetry.

For opening the nonprotonated oxirane or protonated oxirane ring we carried out the MRD-CI calculations using canonical molecular orbitals including the entire valence space. From the 23389 CSFs generated, 1675 CSFs were included explicitly in the CI calculation (a threshold of 10 microhartrees), the energies of remaining configurations were added back in by a perturbation procedure, and a Davidson-type correction was added for size consistency.

III. Results and Discussion

A. Protonation Paths of Oxetanes Involve Multipotential Surfaces

As mentioned above, all protonation reactions are ion–molecule reactions $A^+ + B$ of type 2 where the ionization potential of A (the H atom) is higher than that of B (the molecule). Thus there should be potential energy surfaces arising at the asymptotes from the lower energy asymptotic pair of species $A + B^+$.

We carried out very detailed *ab initio* MRD-CI calculations for proton attack on oxetane. The primary focus of these MRD-CI studies was to demonstrate that the lowest asymptote for dissociation of a protonated species (BH$^+$) was to B$^+$ + H and not to B + H$^+$ (which is the asymptote pair to which a single-determinant SCF wave function would dissociate). There are calculations reported even now in which the path for protonation and deprotonation is depicted incorrectly as the single-determinant SCF potential energy surface. While applications of correlation to protonation and

proton transfer reactions have been reported in the past literature [16–22], none of these were MRD-CI calculations aimed at demonstrating the multideterminant and the multipotential potential energy surface character of the protonation or dissociation of the protonated space. Also, it is not possible to correct the deficiency in a single-determinant SCF potential energy curve for the multideterminant and multipotential surface character by subsequently using MBPT (even to any order on top of a single-determinant SCF wave function). MBPT only corrects the calculated potential curve by perturbation theory for a correlation energy correction to that curve. To use MBPT properly for the dissociation of BH^+, it would first be necessary to carry out MRD-CI calculations and then to apply MBPT *separately* to each of the multireference CI surfaces (with proper correction for intruder states).

The MRD-CI calculations were carried out using localized orbitals (obtained by a Boys localization subsequent to the SCF). This localization leads to much more compact desirable molecular orbitals for the CI calculations. For this series of protonation calculations we included the entire valence orbital space (occupied and virtual). Thus, this is equivalent by a unitary transformation to using the canonical orbitals in the MRD-CI. Our previous *ab initio* MRD-CI [4, 5] (and additional *ab initio* MC-SCF/CI derivative geometry optimization calculations [23] using the GAMESS program [24]) calculations on protonated oxetanes had indicated that the equilibrium geometry for a protonated oxetane was C_{2v} with the oxetane ring planar and the $O\text{-}H^+$ bond in the plane of the molecule along the C_2 axis. We carried out the MRD-CI calculations for a linear attack on oxygen by the proton in the plane and also for an out-of-plane attack by the proton along an oxygen lone pair bond direction. The potential energy curves are shown in Figures 1 and 2. The curves are labelled on the asymptotes to the species to which they dissociate smoothly. The results shown for the singlet states (Fig. 1) validate our original hypothesis of a lower singlet potential energy surface arising from the lower energy pair of species oxetane$^+$ + H.

The results in Figure 1 also indicate an even greater complexity for the potential energy surfaces than we had originally hypothesized. The lowest energy root from 1.5 bohrs to 4.5 bohrs (including the equilibrium internuclear $O\text{-}H^+$ distances of protonated oxetane at 2.0 bohrs) is $^1A_1$. At 4.5 bohrs the lowest root becomes $^1B_1$ to an asymptote of oxetane$^+$ (ground state) + H. At 4.5 bohrs the $^1A_1$ curve continues as the 2nd root of the CI matrix smoothly to oxetane$^+$ (first excited singlet state) + H. The third root of the CI matrix, also a $^1A_1$ state, has a minimum at 2.0 bohrs, a hump at 2.75 bohrs and then continues down to the separated products oxetane (ground state) + H^+. The behavior of the next two higher roots of the CI matrix is also complicated.

Similar results are obtained for the MRD-CI calculations where the proton is attacking oxetane from an out-of-plane lone pair direction.

The MRD-CI results for the protonation pathway also confirm our earlier MRD-CI results that the equilibrium geometry for protonated oxetane has the proton on the oxygen linear and in the plane of the oxetane ring.

In Tables I and II are shown the SCF and MRD-CI energies as a function of H-O distance in the protonation and deprotonation of oxetane.

PROTONATION OF OXETANE

along the O1-H$^+$ bond (head on, in plane)

Ab-Initio MODPOT MRD-CI Extrapolated CI Energies [a.u]
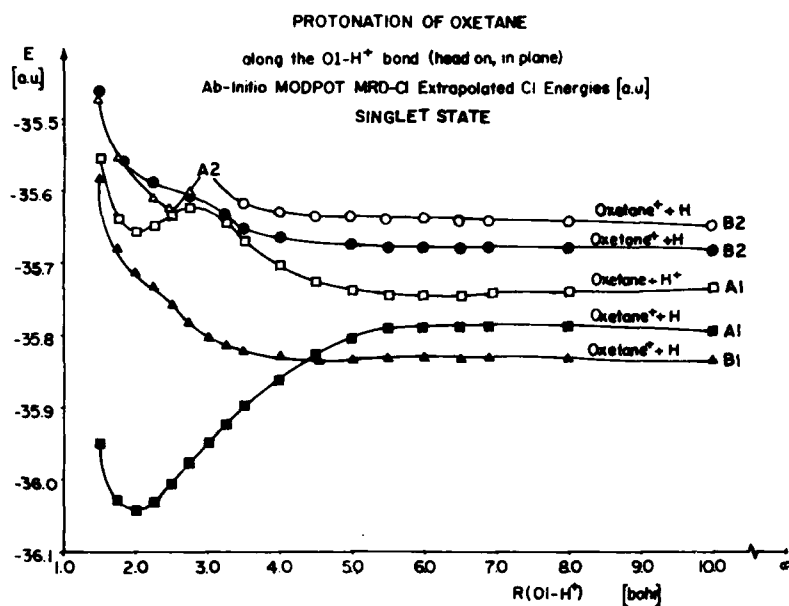
SINGLET STATE

Figure 1.   Protonation of oxetane. Along the O1-H$^+$ bond (head on, in plane). *Ab initio* MODPOT MRD-CI extrapolated CI energies (a.u.) singlet state.

PROTONATION OF OXETANE

H$^+$ approaching oxetane along the lone pair direction

Ab-Initio MODPOT MRD-CI Extrapolated CI Energies [a.u]
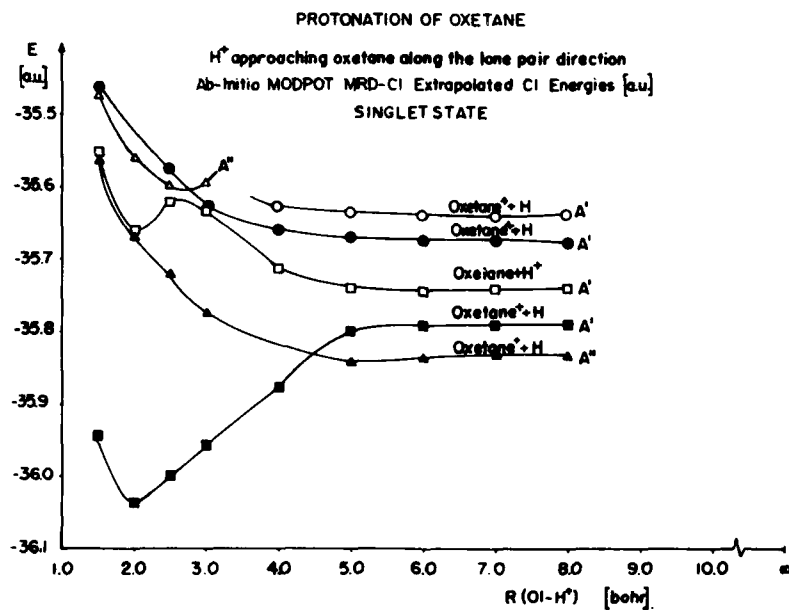
SINGLET STATE

Figure 2.   Protonation of oxetane. H$^+$ approaching oxetane along the lone pair direction. *Ab initio* MODPOT MRD-CI extrapolated CI energies (a.u.) singlet state.

TABLE I. Protonation and deprotonation of oxetane (in plane attack): *Ab initio* MODPOT SCF and MRD-CI (extrapolated CI) energies (a.u.) as a function of O1-H⁺ distances (bohrs).

R(OH⁺)	1.5	2.0	2.5	3.0	4.0	5.0	8.0
E_{SCF}	-35.727424	-35.810593	-35.764750	-35.700286	-35.601313	-35.550850	-35.515064
Protonation E_{MRD-CI}	-35.553923	-35.658251	-35.631026	-35.618364	-35.703081	-35.736611	-35.739917
Deprotonation E_{MRD-CI}	-35.931307	-36.042947	-36.005860	-35.947602	-35.856397	-35.806400	-35.788240

TABLE II. Protonation and deprotonation of oxetane (out of plane attack): *Ab initio* MODPOT SCF and MRD-CI (extrapolated CI) energies (a.u.) as a function of O1-H⁺ distances (bohrs).

R(O1-H⁺)	1.5	2.0	2.5	3.0	4.0	5.0	8.0
E_{SCF}	-35.716536	-35.802237	-35.759405	-35.695669	-35.596166	-35.545405	-35.511880
Protonation							
E_{MRD-CI}	-35.532170	-35.626394	-35.617291	-35.633666	-35.712720	-35.740980	-35.738087
Deprotonation							
E_{MRD-CI}	-35.943823	-36.037697	-36.009876	-35.956916	-35.876544	-35.788875	-35.791311

The MRD-CI results show that the ground state equilibrium geometry of $^1A_1$ protonated oxetane goes to the separated ($^1A_1$) products oxetane$^+$ ($^2A_1$) + H. To what then do the single-determinant SCF calculations for protonation which one sees carried out and published correspond? We laid the $^1A_1$ potential surfaces for the single-determinant SCF calculation over the MRD-CI curves (using the same energy value for the equilibrium geometry minimum of protonated oxetane). The SCF curve connects smoothly to the asymptote oxetane + H$^+$, however, the SCF curve starts to veer higher than the correct MRD-CI lowest $^1A_1$ curve at only 2.75 bohrs and continues to veer higher until it reaches the asymptote of the second highest MRD-CI $^1A_1$ curve oxetane + H$^+$. The single-determinant SCF wave function begins to deviate in fundamental character from the correct MRD-CI wave function. The form of the single-determinant SCF Hamiltonian forces this incorrect behavior. As an additional check we calculated the gross atomic population on the hydrogen from the SCF wave functions where it goes down smoothly from 0.51 at the equilibrium geometry of 2.0 bohrs to 0.00 at 10.0 bohrs and from the lowest MRD-CI wave functions where it goes up from 0.51 at the equilibrium geometry of 2.00 bohrs to 1.00 at 10.0 bohrs.

The implications of these research results are general. All protonation reactions H$^+$ + B even of closed-shell ground-state molecules (where the ionization potential of the B molecule is less than that of the hydrogen atom) can only be described correctly by MRD-CI (or possibly CI) calculations including all the potential energy surfaces for all the states which arise from separated species H + B$^+$ lower in energy at the asymptotes than H$^+$ + B and all other states which cross those potential energy surfaces. All other ion-molecule reactions A$^+$ + B [where the ionization potential of B is less than that of A (including those where A$^+$ is an ultimate carcinogen which attacks DNA, RNA, etc.)] can also only be properly described by MRD-CI (or possibly CI) calculations of the multipotential surfaces.

faces arising from the asymptotic species AH$^+$ + B to transfer a proton to form A + BH$^+$ will be different from those arising from BH$^+$ + A to transfer a proton to form AH$^+$ + B. To date, even the *ab initio* calculations reported for proton transfer in biological/biomedical systems appear to have been only *ab initio* SCF calculations (perhaps with the addition of MBPT corrections). As we pointed out above MBPT cannot correct for the lack of multideterminant character in a single-determinant SCF calculation. Our findings that the fundamental character of single-determinant SCF potential surfaces begin to deviate in from the MRD-CI surfaces as early as 0.75 bohr from the equilibrium position of the proton on a species implies that MRD-CI (or possibly CI calculations) [for all the states arising from the asymptotes AH$^+$ + B or A + BH$^+$ and curves which cross these] will be necessary to describe these processes correctly. We are extending these MRD-CI studies to the proton transfer reactions.

B. Differences in Dominant Configurations in Opening Nonprotonated or Protonated Oxirane Rings (Epoxides)

Oxiranes (the 3-member ring cyclic ethers) are the simplest of the epoxides. We already knew from our previous MRD-CI calculations on opening oxetane or protonated oxetane rings (the 4-member ring cyclic ethers) that the single-determinant SCF wave function for the open oxetane ring had a c^2 contribution of only 0.69 to the MRD-CI

wave function while the single-determinant SCF wave function for the open proto-
nated oxetane ring had a c^2 contribution of 0.93 to the MRD-CI wave function. On
opening the protonated oxetane ring both electrons remain on the oxygen. This situa-
tion is even more exaggerated in opening nonprotonated oxirane rings. For an oxirane
ring the single-determinant SCF wave function contributes a c^2 of only 0.40 to the
MRD-CI wave function. The single-determinant SCF wave function for an open proto-
nated oxirane ring contributes a c^2 of 0.92 to the MRD-CI wave function (just as for
the open protonated oxetane ring). Thus a single-determinant SCF wave function for
opening a protonated cyclic ether ring is a reasonable description of the system while
a single-determinant SCF wave function is a completely incorrect description for open-
ing a neutral cyclic ether ring.

Thus to calculate properly the pathway for opening an epoxide ring in a metaboli-
cally activated PAH carcinogen (or an epoxide ring in any other carcinogen or
metabolically activated carcinogen) will necessitate MRD-CI or CI calculations.

Acknowledgments

Bibliography

[1] F. L. Tobin, P. C. Hariharan, J. J. Kaufman, and R. S. Miller, Int. J. Quantum Chem. S15, 203
(1981).

[2] (a) R. J. Buenker, S. D. Peyerimhoff, and W. Butscher, Mol. Phys. 35, 771 (1978); (b) R. J.
Buenker, in *Studies in Physical and Theoretical Chemistry*, Vol. 21, Current Aspects of Quantum
Chemistry 1981, R. Carbo, Ed. (Elsevier Scientific, Amsterdam, 1982), pp. 17–34; (c) R. J.
Buenker, in *Proceedings of Workshop on Quantum Chemistry and Molecular Physics in Wollongong*,
Australia, February 1980; (d) R. J. Buenker and R. A. Phillips, J. Mol. Struct. Theochem. 123, 291–
300 (1985).

[3] J. J. Kaufman, P. C. Hariharan, and S. Roszak. "Ab-Initio Multireference Double Excitation Configu-
ration Interaction Calculations Based on Localized Orbitals for Molecular Decompositions and Reac-
tions for Large Systems," presented at the Sanibel International Symposium on Quantum Chemistry,
Solid State Theory, Many-Body Phenomena and Computational Quantum Chemistry, Marineland,
Florida, March 1986.

[4] J. J. Kaufman, P. C. Hariharan, S. Roszak, and P. B. Keegstra, "Ab-Initio Electrostatic Molecular
Potential Contour Maps for Initia-tion Step and Ab-Initio MRD-CI Calculations for Propagation Step of
Cationic Polymerization of Oxetanes," an invited lecture presented at the IUPAC 5th International
Symposium on Ring-Opening Polymerization, Blois, France, June 1986. Makromolecular Chemie,
special Symposium Issue. 6, 315–330 (1986).

[5] J. J. Kaufman, "Ab-Initio MRD-CI Calculations for Cationic Polymerization of Oxetanes Based on Localized Orbitals," presented at Sanibel International Symposium on Atomic, Molecular and Solid State Theory, Scattering Problems, Many-Body Phenomena, and Statistical and Computational Methods, Marineland, Florida, March 1987. In press, Int. J. Quantum Chem., Symposium Issue.

[6] Y. N. Chas, "Metabolism of Foreign Compounds," presented at the meeting Principles of Toxicology, The Johns Hopkins University School of Public Health and Hygiene, Baltimore, Maryland, May 1982.

[7] S. K. Yang, D. W. McCourt, P. P. Roller, and H. V. Gelboin, Proc. Natl. Acad. Sci., U.S.A. 73, 2594 (1976).

[8] J. J. Kaufman, H. E. Popkie, S. Palalikit, and P. C. Hariharan, Int. J. Quantum Chem. 14, 793 (1978).

[9] P. C. Hariharan, H. E. Popkie, and J. J. Kaufman, Int. J. Quantum Chem. 20, 645 (1981).

[10] P. C. Hariharan, J. J. Kaufman, and C. Petrongolo, Int. J. Quantum Chem. QBS6, 223 (1979).

[11] P. C. Hariharan, J. J. Kaufman, and C. Petrongolo, Int. J. Quantum Chem. 20, 1083 (1981).

[12] J. J. Kaufman, P. C. Hariharan, W. S. Koski, and K. Balasubramanian, "Quantum Chemical and Other Theoretical Studies of Carcinogens, Their Metabolic Activation and Attack on DNA Constituents," an invited plenary lecture presented at the meeting The Molecular Basis of Cancer, Buffalo, New York, May 1984, in, *The Molecular Basis of Cancer*, Part A: *Macromolecular Structures, Carcinogens, and Oncogenes*, R. Rein, Ed. (Alan R. Liss, Inc., New York, 1985), pp. 263–275.

[13] J. J. Kaufman. Int. J. Quantum Chem. QBS6, 503 (1979).

[14] J. J. Kaufman, H. E. Popkie, and P. C. Hariharan, "New Optimal Strategies for Ab-Initio Quantum Chemical Calculations on Large Drugs, Carcinogens, Teratogens and Biomolecules," an invited lecture presented at the Symposium on Computer Assisted Drug Design. Division of Computers in Chemistry at the American Chemical Society National Meeting, Honolulu, Hawaii, April 1979, in *Computer Assisted Drug Design*, E. C. Olson and R. E. Christoffersen, Eds. (ACS Symposium Series 112, Am. Chem. Soc., Washington, D.C., 1979), pp. 415–435.

[15] G. Herzberg, *Electronic Spectra of Polyatomic Molecules*. (Van Nostrand Company, Inc., Princeton, 1966), pp. 563, 564, 577.

[16] W. Meyer, W. Jakubetz, and P. Schuster, Chem. Phys. Lett. 21, 97 (1973).

[17] H. Lischka, Theoret. Chim. Acta 31, 39 (1973).

[18] W. Meyer, J. Chem. Phys. 58, 1017 (1973).

[19] A. Stogard, A. Strich, J. Almlof, and B. Roos, Chem. Phys. 8, 405 (1975).

[20] G. Karlstrom, B. Jonsson, B. Roos, and H. Wennerstrom, J. Am. Chem. Soc. 98, 6851 (1976).

[21] M. M. Szczesniak and S. Scheiner, J. Chem. Phys. 77, 4586 (1982).

[22] F. Keil and R. Ahlrichs, J. Am. Chem. Soc. 98, 4787 (1976).

[23] P. C. Hariharan and J. J. Kaufman, The Johns Hopkins University, Summer 1986.

[24] "GAMESS;" (a) M. Duppuis, NRC^2, Berkeley, California, Summer 1981; (b) V. R. Saunders and M. Guest, Atomic and Molecular Theory Group, Daresbury Laboratory, Warrington, England.

Ab Initio Investigation of the Structure of Hydrogen Halide-Amine Complexes in the Gas Phase and in a Polarizable Medium

INGRID J. KURNIG* AND STEVE SCHEINER[†,‡]

*Department of Chemistry & Biochemistry, Southern Illinois University,
Carbondale, Illinois 62901, U.S.A.*

Abstract

The geometries of all 12 complexes in which HF, HCl, or HBr is paired with NH_3, $NMeH_2$, NMe_2H, or NMe_3 are optimized with the MINI-1 basis set. As the basicity of the amine is increased via progressive methylation, or as the proton affinity of the halide is diminished, the proton equilibrium position shifts toward the nitrogen, but in no case is this shift far enough to classify the complex as an ion pair. When the effects of a polarizable medium are included via the SCRF formalism, the shift of the proton toward the nitrogen is enhanced by increases in the solute-solvent interaction such that relatively modest coupling leads to complexes of ion-pair type. In all cases, complexes containing HBr are the most sensitive to either the basicity of the amine or the influence of the medium whereas the HF analogs are affected very little.

Introduction

Over the last several decades, *ab initio* quantum mechanical calculations have served as a rich source of information covering a wide range of chemical reactions. The conceptual simplicity of proton transfers, coupled with their widespread occurrence in chemical and biological processes, has endowed this type of reaction with a particular significance and fostered a large number of *ab initio* studies [1–5]. Recent calculations in this laboratory have led to the enunciation of a set of rules which govern proton transfer reactions and which may be used to predict the important energetic and kinetic parameters in any given case from first principles [6–9].

Whereas our previous calculations have applied rigorously only to isolated systems *in vacuo*, it is our belief that the principles apply in condensed phases as well, albeit with some modification to account for the influence of the surrounding medium. Our strategy for quantitating the environmental effects involves a partitioning into a number of conceptual stages, of diminishing importance. The largest effect would undoubtedly arise from the electrostatic fields emanating from any neighboring ions or dipoles. Hence, we first considered the effects upon the proton transfer potential of

---

* Fulbright scholar.
[†] Recipient of NIH Research Career Development Award.
[‡] Address correspondence to this author.

an array of such entities in various locations about a H-bonded system [10]. Consistent with prior work, our results supported the contention that the effects of ions and/ or dipoles can be extremely important but that they can be simply understood on the basis of interaction of their electric fields with the partial positive charge of the proton being transferred.

Even in the complete absence of ions or permanent dipoles, one would expect the proton transfer potential to be affected by the immersion of the H-bonded system in a polarizable medium. Indeed, there is ample evidence from prior calculations to support this notion [11–18]. The present communication reports the results of a systematic study of a number of closely related systems, immersed in a medium whose dielectric constant is smoothly varied over a wide range. For this purpose, we choose the H-bonded complexes composed of a hydrogen halide and an amine. HF, HCl, and HBr are chosen for the former, so as to cover a range of proton-donating ability; NH_3, $NMeH_2$, NMe_2H, and NMe_3 are proton acceptors of increasingly greater power.

Previous work of both a theoretical and experimental nature has demonstrated that hydrogen halide-amine complexes are characterized by a single-well proton transfer potential [19–25]. The position of this minimum, specifically, the proximity of the proton to the amine or halide, is quite sensitive to both the nature of the subunits involved and the characteristics of the medium in which it is immersed. For example, H_3N—HF is almost certainly a neutral pair in the gas phase while the complex between trimethylamine and HBr is better described as an ion-pair, Me_3NH^+—Br^-. We have placed the various systems inside a spherical cavity, surrounded by a dielectric, and monitored the equilibrium proton position as the dielectric constant is slowly increased. In this manner, we hope to gain insights into the effects of a polarizable medium upon the nature of the proton transfer process.

Details of Calculations

Ab initio calculations were carried out using the MONSTERGAUSS package of computer codes [26]. The geometries of all monomers were first completely optimized in order to calculate the deprotonation and dissociation energies reported below. A linear H bond was assumed in all complexes. The internal geometries of the amines were held fixed in their isolated monomer structures while the distances of the central proton to the halide and nitrogen atoms were optimized.

The effects of the homogeneous reaction field were included via the version of Tapia's SCRF program which is incorporated into MONSTERGAUSS. The reaction field susceptibility, that is, the solute-solvent coupling factor, is computed as

$$g = (2/a^3)(\varepsilon - 1)/(2\varepsilon + 1) \tag{1}$$

where ε is the static dielectric constant of the medium and a the radius of the spherical cavity in which the HX-amine system was placed [15–18, 27].

From previous calculations, it appears that the nature of the calculated proton transfer potential in hydrogen halide-amine systems is especially sensitive to the level of theory. For example, whereas SCF calculations suggest double-minimum potentials for complexes of HCl with methylated amines, inclusion of correlation leads to a potential with a single broad minimum in all cases [19]. Similarly with

BrH-NMeH$_2$, the two wells found in the SCF potential [22] coalesce into a single broad minimum in the correlated potential [20]. When viewed in tandem with a wealth of inert matrix spectroscopic data, the picture that emerges is one in which most, if not all, hydrogen halide-amine complexes are characterized by a single-well potential. The equilibrium position of the central proton shifts smoothly between the halide and N atoms as the relative basicities of the two component subunits are altered.

Since SCRF formalism has not been incorporated into correlated treatments, it was not possible for us to directly include the effects of electron correlation. Further considerations, such as the size of some of our systems and the large number of geometrical configurations required for our systematic study, effectively rule out the use of an extended polarized basis set. Sets of the double-ζ type would not be appropriate here since they lead to the erroneous conclusion of a double-minimum proton transfer potential for these systems [19].

Fortunately for our purposes, the economical MINI-1 basis set, developed by Tatewaki and Huzinaga [28] correctly reproduces the single-well nature of the potentials of all of the hydrogen-halide systems. Moreover, as may be seen in Table I, this basis set does a surprisingly good job with the deprotonation energies of the hydrogen halides and protonated amines. All experimental trends are reproduced correctly; viz. HF > HCl > HBr and the increasing basicity of the amines arising from progressive methylation. Quantitatively, comparison of the last two columns of Table I reveals that all energies are within 1 or 2 kcal/mol of the experimental quantities with the exception of HCl and HBr which are predicted to be too basic by about 30 kcal/mol. This discrepancy is not at all surprising in light of the large basis set superposition

TABLE I. Deprotonation Energies and Equilibrium Bond Lengths in Protonated Species.

		Deprotonation energy, kcal/mol		
		$\Delta E(0\ K)$	$\Delta H(298\ K)$	
	r(XH), Å	calc	calc[a]	expt[b]
HF	0.980	377.3	372.3	371.3
HCl	1.367	367.9	364.5	333.3
HBr	1.493	360.2	357.3	323.6
	r(NH), Å	calc	calc[c]	expt[d]
NH$_4^+$	1.085	211.5	202.8	205.0
H$_3$MeN$^+$	1.079	220.7	212.0	214.1
H$_2$Me$_2$N$^+$	1.075	227.7	219.0	220.5
HMe$_3$N$^+$	1.071	232.5	223.8	224.3

[a] Corrected by adding 3/2 RT and zero-point vibrational energy corrections from Ref. 29.

[b] Ref. 30.

[c] Corrected by adding 5/2 RT and zero-point vibrational energy corrections from Ref. 31 (10.2 assumed for ΔZPE of all amines)

[d] Ref. 32.

error expected for the single-center halide anions with a minimal basis set. Rather, what is surprising is the good agreement for HF, due presumably to cancellation of the superposition error with other deficiencies of the basis set.

In sum, due to the good reproduction of proton transfer potentials and deprotonation energies, we expect the SCF calculations with the MINI-1 basis set to provide a reasonable model of the systems of interest. The largest errors are expected for the complexes containing HCl and HBr for which the high deprotonation energies will make it more difficult to extract a proton, biasing the potentials away from the ion-pair. In a broader context, although the theoretical model may have certain unavoidable failings with respect to the specific systems, it will certainly provide insights into the manner in which the structure of a general H bond is affected by the difference in proton affinity between the two subunits and by its immersion in a dielectric bath.

Complexes in Vacuo

The optimized geometries of all 12 complexes, as well as their complexation energies *in vacuo*, are reported in Table II. It is clear from the first grouping that in complexes containing HF, neither the bond lengths to the central hydrogen nor the dissociation energies E^D are much affected by the nature of the base. In complexes containing either HCl or HBr, on the other hand, methylation of the amine leads to (i) substantial shortening of $r(NH)$, (ii) noticeable lengthening of $r(XH)$, and (iii) a larger E^D. The combined result of (i) and (ii) can be described as a shift of the proton from the halide to the nitrogen atom, consistent with the greater basicity of the amine. This enhanced proton acceptor ability makes for a stronger interaction with HX, hence the increase in E^D.

It would be useful to have a quantitative measure of the degree of sharing of the central proton between the halide and the amine, one that could be used in all of our systems. Since the ionic character of the complex clearly increases as the proton moves away from the halogen atom, one might draw a connection between $r(XH)$

TABLE II.   Calculated Geometries and Dissociation Energies of Complexes *In Vacuo*.

system	$r(NH)$, Å	$r(XH)$, Å	E^D, kcal/mol
FH-NH$_3$	1.744	0.996	14.2
FH-NMeH$_2$	1.733	0.997	13.8
FH-NMe$_2$H	1.727	0.997	14.1
FH-NMe$_3$	1.721	0.997	13.8
ClH-NH$_3$	1.579	1.443	14.5
ClH-NMeH$_2$	1.535	1.456	15.1
ClH-NMe$_2$H	1.497	1.468	15.8
ClH-NMe$_3$	1.468	1.479	16.2
BrH-NH$_3$	1.538	1.603	12.3
BrH-NMeH$_2$	1.456	1.637	13.3
BrH-NMe$_2$H	1.386	1.674	14.6
BrH-NMe$_3$	1.340	1.703	15.5

and the degree of proton transfer. However, the different size of each halogen atom would make such an interpretation rather misleading. For example, since Br is much larger than F, r(BrH) would naturally be much longer than r(FH), even when there is no proton transfer at all. Nor would r(NH) be suitable since there is some variation in the equilibrium bond length in the various protonated amines (see Table I).

Rather than use the bond lengths themselves, we have devised a parameter which incorporates the *stretches* of the bonds. Specifically, Δr(XH) is defined as the difference between r(XH) in the complex and r(XH) in the isolated hydrogen halide molecule, the values of which are listed in Table I. Similarly, Δr(NH) represents the stretch of the proton in the complex away from the nitrogen atom, relative to the appropriate protonated amine, the bond lengths of which are also contained in Table I. We define a "proton transfer parameter" as the difference between the two latter stretches.

$$\rho = \Delta r(XH) - \Delta r(NH) \tag{2}$$

When the proton within a given complex has been pulled an equal distance away from both the halogen and the nitrogen atoms, the two stretches are equal and $\rho = 0$, appropriate for an equally shared proton. Negative values of ρ correspond to a lesser stretch away from the halogen than from the amine, which we interpret as a neutral pair. Similarly, as ρ becomes progressively more positive, the proton is pulled more toward the amine and the complex takes on more ionic character.

The proton transfer parameter ρ is presented for all 12 of our complexes in Figure 1. The horizontal axis corresponds to the relative attracting power of the two subunits of each complex for the central proton. Specifically, the normalized proton affinity difference, NPAD is defined as

$$NPAD = [PA(Am) - PA(X^-)] / [PA(Am) + PA(X^-)] \tag{3}$$

where PA(Am) and PA(X$^-$) refer to the proton affinities of the amine and halide anion, respectively. We take these proton affinities to be the calculated deprotonation energies $\Delta E(0\,K)$ of the corresponding protonated species in Table I. This normalized quantity has found widespread use by spectroscopists over the years [23].

The data in Figure 1 are grouped together by hydrogen halide. The numbers labeling each point indicate the number of methyl groups on the amine. Hence, the far left point of the HF curve, labeled by 0, refers to the FH-NH$_3$ complex and the far right point to FH-NMe$_3$. The flatness of the HF curve is a manifestation of the insensitivity of the equilibrium proton position in FH-Am complexes to the basicity of the amine Am. Indeed, r(FH) undergoes essentially no change as the proton affinity of the amine increases over a range of 21 kcal/mol. The large negative values of ρ confirm the strong neutral-pair character of the complexes containing HF.

When HF is replaced by HCl, ρ remains in the negative domain but shows a great deal more sensitivity to NPAD. The 21 kcal/mol rise in proton affinity from NH$_3$ to NMe$_3$ increases ρ from -0.42 Å to -0.29 Å. The sensitivity is further enhanced in the complexes containing HBr where ρ varies from -0.34 Å to -0.06. Indeed, the small magnitude of the latter value of ρ would lead us to describe the proton in the BrH-NMe$_3$ complex as being almost equally shared between the Br and the base. A
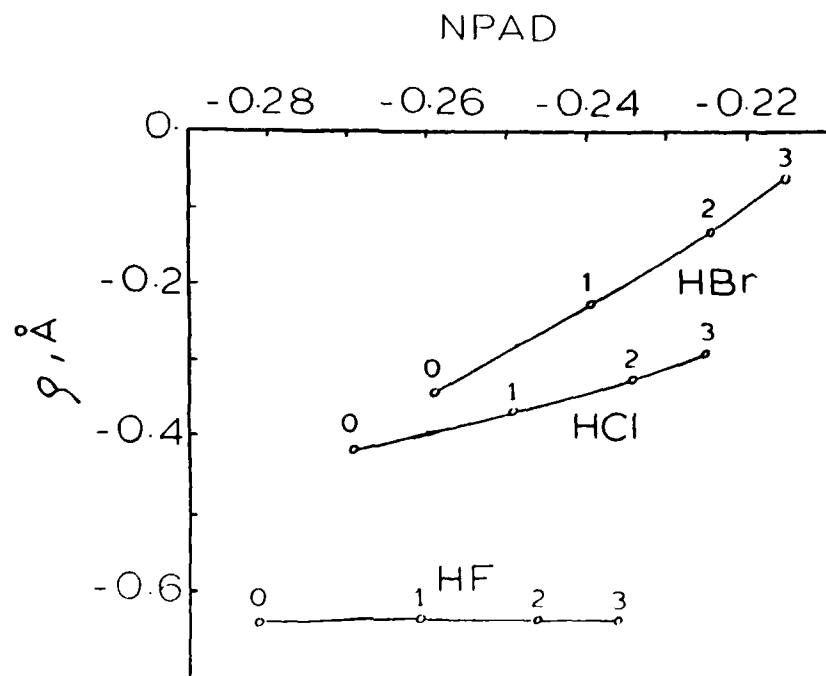
Figure 1.   Proton transfer parameter p for complexes in the gas phase plotted as a function of the normalized proton affinity difference, NPAD, between the amine and halide. The numerical label on each point refers to the number of methyl groups within the amine. The negative values of p indicate a neutral pair.

major finding is thus that the increased basicity of the amine caused by progressive methylation leads to an increased degree of proton transfer.

One may look at the results from the alternate perspective of holding the amine constant and changing the acid. Comparison of all three points marked by any number n of methyl groups on the amine immediately reveals that p increases in the sequence HF < HCl < HBr. That is, a more acidic hydrogen halide leads to a greater degree of proton transfer to the amine within the complex.

In summary, all of the complexes examined here would be best described *in vacuo* as neutral pairs in which the proton is more closely associated with the halogen than with the amine. The least acidic hydrogen halide, HF, shows little tendency to release its proton, while our strongest base, NMe_3, is nearly successful in achieving equal sharing of the proton with our strongest acid, HBr. It was mentioned above that the calculated proton affinities of Br^- and Cl^- are inflated when compared to experiment, exaggerating the ability of these halogens to hold on to a proton. It would, therefore, not be unreasonable to presume that the values of p would be appreciably larger in the real systems, probably surpassing zero for a number of complexes which would fall into the category of ion-pairs *in vacuo*.

One last facet meriting discussion concerns the "compactness" of each complex. Addition of $r(NH)$ and $r(XH)$ in Table II yields the distance between the N and X

atoms which we refer to as the H-bond length, R. In any series associated with a particular halide atom, the decrease in $r(NH)$ arising from progressive methylation of the amine outweighs the smaller increase in $r(XH)$ and R is therefore diminished. For example, R is equal to 2.740 Å for FH-NH$_3$ and decreases by 0.02 Å to 2.718 Å in FH-NMe$_3$. The corresponding reductions in R caused by trimethylation are 0.05 and 0.10 Å for the ClH and BrH series, respectively. Making use of our prior finding of greater ion-pair character induced by methylation, we may further connect the degree of proton transfer with a shorter H bond.

As a point of interest, let us consider the HBr curve in Figure 1 which would intercept the horizontal axis at about −0.21. Using Eq. (3), we get 235 kcal/mol as the proton affinity of an amine which could equally share a proton with Br$^-$, particularly notable since this value is some 125 kcal/mol smaller than the proton affinity of Br$^-$. The reason why an amine with a lower proton affinity can successfully compete with the anion rests in the fact that transfer of the proton to the amine produces an ion pair with an extremely strong electrostatic attraction. In contrast, the force holding the complex together when the proton resides on the halogen is much smaller since both subsystems are formally uncharged. This same principle is responsible for the aforementioned contraction of the H bond as the proton is transferred toward the amine.

Immersion in Dielectric Continuum

Let us now place each complex in a spherical cavity within a dielectric medium. As described earlier, the g parameter contains the strength of the electrostatic interaction of the complex with the solvent, by way of the cavity radius and the dielectric constant. The degree of proton transfer p for each complex is presented in graphical fashion as a function of g in Figure 2. The solid curves correspond to the HF complexes, dashed curves to HCl, and dotted to HBr. Each curve is labeled with the number of methyl groups on the amine.

It should be first noted that all curves rise as g is increased. Thus, the greater interaction with the solvent preferentially stabilizes the ion-pair, shifting the proton toward the amine and thus making p more positive. The solid curves which represent the HF-containing complexes rise most gradually, indicating that even a relatively strong interaction with solvent can only produce a slight shift of the proton toward the amine in these intrinsically very nonionic complexes.

The increases in p are much sharper for the other curves in Figure 2, indicating that the character of the complexes containing HCl and HBr is much more sensitive to interaction with the solvent. As shown by the intercepts of the dashed curves with the horizontal axis, equal sharing of the proton between HCl and the amine occurs for g in the range 0.001–0.002 au. In order to place this range in perspective, if we take the cavity radius to be 4 Å, the approximate maximum length of any of our HCl-containing complexes, a value of 0.002 au for g corresponds to a dielectric constant ε of 10.

The sensitivity of p to the influence of solvent is even greater for the complexes containing HBr, as illustrated by the steepness of the dotted curves. This sensitivity, in conjunction with the more nearly equal sharing of the proton even in the gas phase, leads to Br$^-$ . . $^+$HAm ion-pairs (i.e., $p > 0$) for relatively small values of g.
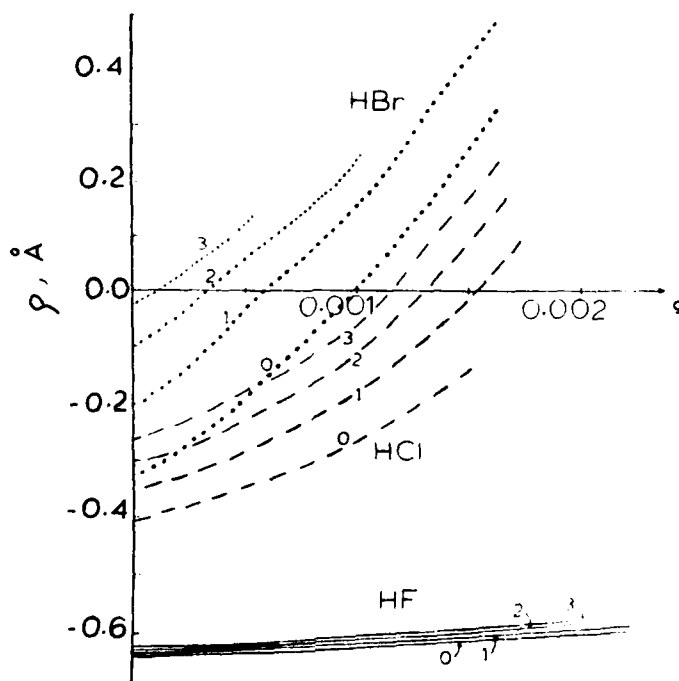
Figure 2. Degree of proton transfer shown in terms of the strength of the interaction of the complex with its surrounding dielectric continuum. Solid curves refer to HF complexes, broken to HCl, and dotted to HBr. Each curve is labeled with the number of methyl groups contained within the amine. g is in units of au.

For example, taking 4.7 Å as the radius of complexes containing HBr, the g value of 0.0005 au, for which a number of these complexes are in the ionic range of ρ, corresponds to a value of only 2 for ε. Keeping in mind the inflated proton affinities of the Cl^- and Br^- anions with the MINI-1 basis set, we may anticipate that a more realistic treatment would displace all the dashed and dotted curves upward to some extent and would result in a crossing of the horizontal axis at reduced values of g.

It may be noted that a number of the curves in Figure 2 do not extend all the way to the right end of the scale. These limits relate to the finite size of each complex which provides a minimum for the cavity radius a. Since the expression $(\varepsilon - 1)/(2\varepsilon + 1)$ has a maximum value of 0.5 for infinite ε, the entire expression for g in Eq. (1) is bounded above by the inverse cube of the molecular dimension. Attempts to reduce a below the van der Waals radius of any complex led to unrealistic expansion of the system.

Concomitant with the increasing ion-pair character of the various complexes as manifested by more positive values of ρ, the enhanced interaction with the medium also induces a contraction of each complex. In keeping with the earlier patterns, the FH-amine complexes are least sensitive, with R diminishing by up to 0.035 Å when g is increased from zero to 0.002 au. The H-bond contraction in complexes containing

HCl or HBr is substantially greater, up to 0.1 Å. This trend would undoubtedly continue for larger values of g except for the aforementioned problem that when the cavity radius a is reduced below the van der Waals length of the complex, an unrealistic expansion is observed.

Conclusions

Raising the basicity of the amine Am by progressive methylation shifts the equilibrium position of the proton in XH—Am toward the nitrogen atom. These shifts are most pronounced when X = Br, Cl while the proton position is almost completely stationary with respect to changes in the amine when X = F. Nevertheless, all 12 of the complexes studied here are predicted to be nonionic in the gas phase; in other words, the proton is stretched a smaller distance from the halide than from the nitrogen of the amine.

Immersion of the system within a dielectric continuum leads in all cases to a more ionic complex, as one might expect on the basis of the medium's ability to stabilize a solute with appreciable charge separation. Due to their nearly equal sharing of the proton even in the gas phase, and to their sensitivity to the medium, BrH—Am complexes require only a small solute-solvent coupling to be classified as ion-pairs. The Cl analogs require somewhat larger coupling, on the order of 0.001–0.002 au, to reach the same state. On the other hand, the proton position in FH—Am complexes changes very little, even when the dielectric constant is quite large. In all cases, greater ionic character of the complex leads to a contraction of the distance separating the halide from the amine.

Acknowledgments

Bibliography

[1] M. L. McKee, J. Am. Chem. Soc. 109, 559 (1987).

[2] H. Z. Cao, M. Allavena, O. Tapia, and E. M. Evleth, J. Phys. Chem. 89, 1581 (1985).

[3] J. L. Andres, M. Duran, A. Lledos, and J. Bertran, Chem. Phys. Lett. 124, 177 (1986).

[4] W. H. Jones, P. G. Mezey, and I. G. Csizmadia, J. Mol. Struct. (Theochem) 121, 85 (1985).

[5] H. Basch, M. Krauss, and W. J. Stevens, J. Am. Chem. Soc. 107, 7267 (1985).

[6] S. Scheiner, Acc. Chem. Res. 18, 174 (1985).

[7] E. A. Hillenbrand and S. Scheiner, J. Am. Chem. Soc. 108, 7178 (1986); ibid., 107, 7690 (1985).

[8] S. Scheiner and P. Redfern, J. Phys. Chem. 90, 2969 (1986).

[9] S. M. Cybulski and S. Scheiner, J. Am. Chem. Soc. 109, 4199 (1987).

[10] S. Scheiner, P. Redfern, and M. M. Szczesniak, J. Phys. Chem. 89, 262 (1985); I. J. Kurnig and S. Scheiner, Int. J. Quantum Chem: QBS 13, 71 (1986).

[11] J. Angyan and G. Naray-Szabo, Theor. Chim. Acta 64, 27 (1983).

[12] P. Th. van Duijnen and B. T. Thole, in Quantum Theory of Chemical Reactions, R. Daudel, A. Pullman, L. Salem, and A. Veillard, Eds. (Reidel, Dordrecht, 1982), Vol. 3, pp. 85–95; P. Th. van Duijnen and B. T. Thole, Chem. Phys. Lett. 83, 129 (1981); B. T. Thole and P. Th. van Duijnen, Biophys. Chem. 18, 53 (1983).

[13] H. Pfeiffer, G. Zundel, and E. G. Weidemann, J. Phys. Chem. **83**, 2544 (1979); R. Janoschek, E. G. Weidemann, H. Pfeiffer, and G. Zundel, J. Am. Chem. Soc. **94**, 2387 (1972).

[14] F. Zuccarello, A. Raudino, and G. Buemi, J. Mol. Struct. (Theochem) **107**, 215 (1984).

[15] O. Tapia, C. -I. Bränden and A. -M. Armbruster, in *Quantum Theory of Chemical Reactions*, R. Daudel, A. Pullman, L. Salem, and A. Veillard, Eds. (Reidel, Dordrecht, 1982), Vol. 3, pp. 97–123; O. Tapia, *ibid.*, 1981, Vol. 2, pp. 25–72.

[16] J. E. Sanhueza and O. Tapia, J. Mol. Struct. (Theochem) **89**, 131 (1982).

[17] O. Tapia, F. M. L. G. Stamato, and Y. G. Smeyers, J. Mol. Struct. (Theochem) **123**, 67 (1985).

[18] O. Tapia, E. Poulain, and F. Sussman, Chem. Phys. Lett. **33**, 65 (1975).

[19] Z. Latajka, S. Sakai, K. Morokuma, and H. Ratajczak, Chem. Phys. Lett. **72**, 115 (1980).

[20] Z. Latajka, S. Scheiner, and H. Ratajczak, Chem. Phys. Lett. **135**, 367 (1987).

[21] P. G. Jasien and W. J. Stevens, Chem. Phys. Lett. **130**, 127 (1986).

[22] A. Brciz, A. Karpfen, H. Lischka, and P. Schuster, Chem. Phys. **89**, 337 (1984).

[23] A. J. Barnes and M. P. Wright, J. Chem. Soc., Faraday Trans. 2, **82**, 153 (1986).

[24] E. J. Goodwin, N. W. Howard, and A. C. Legon, Chem. Phys. Lett. **131**, 319 (1986).

[25] N. S. Golubyev and G. S. Denisov, Khim. Fiz. **5**, 563 (1982).

[26] M. R. Peterson and R. A. Poirier, Department of Chemistry, 1984, MONSTERGAUSS, University of Toronto, Ontario, M5S 1A1.

[27] O. Tapia and O. Goscinski, Mol. Phys. **29**, 1653 (1975); O. Tapia, F. Sussman, and E. Poulain, J. Theor. Biol. **71**, 49 (1978); R. Costanciel and O. Tapia, Theor. Chim. Acta, **48**, 75 (1978).

[28] H. Tatewaki and S. Huzinaga, J. Comput. Chem. **1**, 205 (1980); Y. Sakai, H. Tatewaki, and S. Huzinaga, *ibid.*, **2**, 100 (1981).

[29] P. W. Atkins, *Physical Chemistry*, W. H. Freeman, San Francisco, 1982, p. 589.

[30] J. E. Bartmess and R. T. McIver, Jr., in *Gas Phase Ion Chemistry*, M. T. Bowers, Ed. (Academic Press, New York, vol. 2, 1985), Ch. 11.

[31] J. E. Del Bene, M. J. Frisch, K. Raghavachari, and J. A. Pople, J. Phys. Chem. **86**, 1529 (1982).

[32] D. H. Aue and M. T. Bowers, in *Gas Phase Ion Chemistry*, M. T. Bowers, Ed. (Academic Press, New York, 1985) vol. 2, Ch. 9.

An MNDO Molecular Orbital Study of the Reactions of Protonated Oxirane with Guanine

GEORGE P. FORD AND CHRISTOPHER T. SMITH

Department of Chemistry, Southern Methodist University, Dallas, Texas 75275 U.S.A.

Abstract

The bimolecular hydroxyalkylation of the N^2, 3-, O^6, and 7-position of guanine by protonated oxirane has been studied using the MNDO molecular orbital procedure. The enthalpies of activation (relative to the isolated reactants) were calculated to be 12.5, 11.4, −4.8, and −4.9 kcal mol $^{-1}$, respectively. The transition state geometries were characteristically S_N1-like. The forming bonds reach *ca.* 0–2% of their final strengths while cleavage of the breaking bonds is *ca.* 77–87% complete. Their relative energies are dominated by electrostatic interactions between the reacting moieties with little to no charge transfer involved. The relevance of these results to the reactions of carcinogenic oxiranes and their derivatives with nucleic acids is discussed.

Introduction

It is widely believed that the carcinogenic effects of certain chemical compounds are related to their abilities to covalently modify the nucleic acids [1]. However, the kinds and amounts of adducts formed in such reactions depend markedly on the nature of the reactive electrophile [2]. An important class of carcinogens incorporate the oxirane ring system. Although the parent oxirane is a rather weak carcinogen, some of its derivatives are among the most potent known [3]. A number of biologically significant examples of this kind arise through the metabolic oxidation of substituted alkenes and arenes [4,5]. These include the aflatoxins [6], the polycyclic aromatic hydrocarbons [5], and various simple vinylic compounds [4].

For some years we have been using quantum mechanical calculations in an attempt to understand the physicochemical determinants of the complex regioselectivities observed in the reactions of various electrophiles with nucleic acids and their components [7–14]. As part of this effort we recently completed MNDO semiempirical molecular orbital calculations [15] on the reactions of protonated oxirane (1) [11] and several monosubstituted derivatives [13, 14] with a group of simple nucleophiles. The present paper extends this work to the reactions of 1 at four key nucleophilic sites of guanine (2) itself:

FORD AND SMITH

R = -CH$_2$CH$_2$OH
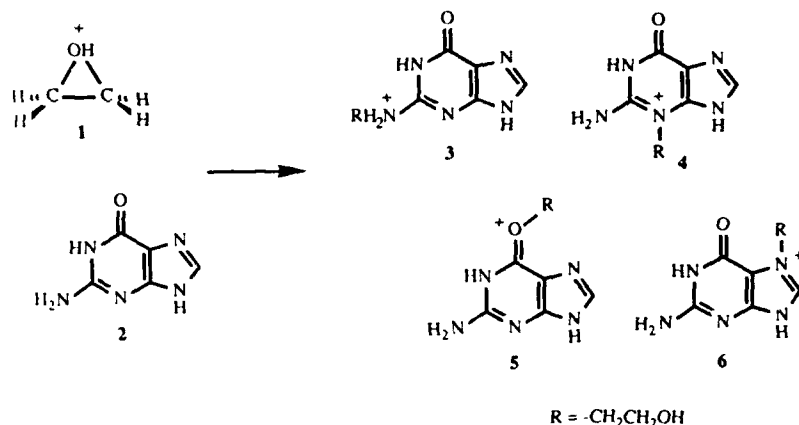
Calculations

Calculations were carried out using the standard MNDO procedure [15] with programs either adapted from versions originating in the laboratories of Professor Dewar for our Harris minicomputers, or an IBM version of the MOPAC package [16]. The geometries of all species were fully optimized with no geometrical constraints. Approximate transition state geometries were constructed on the basis of our previous calculations [11] and refined by minimizing the square of the gradient vector in the usual way. Examination of the eigenvector associated with the single negative eigenvalue of the Cartesian force constant matrix for each transition state showed it to be a true saddle point for the required process [17].

Results

The calculated energetic data for the reactions leading to 3–6 are collected in Table I. In contrast to our earlier studies [11, 14] we did not specifically optimize the geometries of the ion-dipole complexes separating the reactants and transition states which characterize reactions of this kind. Their existence is of course implied in the two cases for which the activation enthalpies, $\Delta H^\ddagger$ (calculated relative to the isolated reactants), were predicted to be negative. The heats of reaction are given for conversion to both the *gauche* and *anti* products. As expected [11, 12] the gauche conformations were predicted to be the more stable. The alternative conformations are illustrated for 6 in Figure 1. The *gauche* conformation in which the OH group lies closest to the lowest numbered position of guanine is arbitrarily designated *gauche* 1 in Table I. The calculated transition state geometries are shown in Figure 2 with key structural and electronic data summarized in Table II. The transition states for these formally bimolecular substitution reactions were calculated to be highly unsymmetrical with virtually no significant bond formation to the incoming nucleophile. Only for attack at the 7-nitrogen is any more than an infinitesimal degree of charge transfer between the reacting moieties predicted. According to our previously suggested index of bond formation, $\Delta n\%$ [12], even here the forming bond reaches only ~2% of its final strength (Table III). The transition state geometries and energies are evidently

TABLE I. Calculated energetics for the gas-phase bimolecular ring opening of Protonated oxirane (1) by guanine (2).[a]

Alkylation site (product)	ΔH_f^{b}	$\Delta H^{\ddagger c}$	ΔH^{d}		
			anti	gauche 1	gauche 2
N^2- (3)	193.2	12.5	−32.8	−38.1	−37.7
3- (4)	192.1	11.4	e	−53.1	−51.4
O^6- (5)	175.9	−4.8	−62.8	−65.4	—
7- (6)	175.8	−4.9	−66.5	−69.2	−70.9
None[f]	187.5	14.5			

[a]Energies in kcal mol^{-1}, distances in Angstroms.

[b]Heat of formation of transition state. ΔH_f (1) = 173.0 kcal mol^{-1}; ΔH_f (2) = 7.7 kcal mol^{-1}.

[c]Activation energy relative to isolated reactants.

[d]Heats of reaction to alternative product conformations. See text.

[e]No stable minimum corresponding to this conformation.

[f]Unimolecular ring opening.

gauche 1   anti   gauche 2
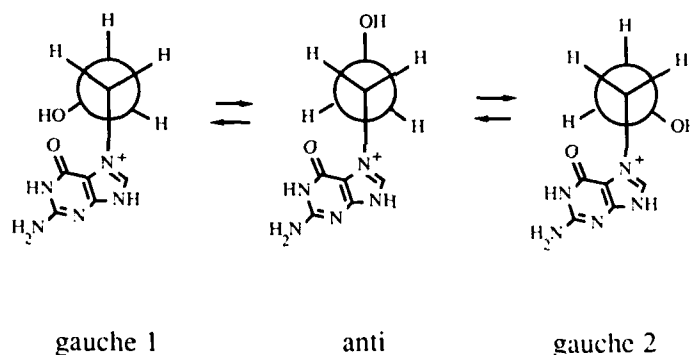
Figure 1. Alternative conformations of 2-(7-guanyl)-ethanol (6).

dominated by electrostatic effects. Although the orientations of the reacting moieties follow general expectations for bimolecular displacements, the angles between the entering and exiting atoms (θ in Table II) are significantly distorted from colinearity. This allows the most favorable placement of the oxirane in the regions of local electrostatic attraction. Thus, for attack at both the O^6- and 7-positions, the centroid of charge of the latter resides in the deep attractive well between these two positions, the individual geometries differing only in the orientation of the exiting group [17]. In the transition state for attack at N^2-, the oxirane moiety is significantly displaced toward the electrostatically attractive region above the adjacent 3-nitrogen. A simple estimate of the electrostatic interaction, E_{el}, between the approaching reactants was made using Eq. 1 where q_i and q_j are the MNDO charges in the guanine and oxirane moieties, respectively, and $r_{i,j}$ the corresponding internuclear separations. These calculations were performed in two ways. Both were based on the calculated
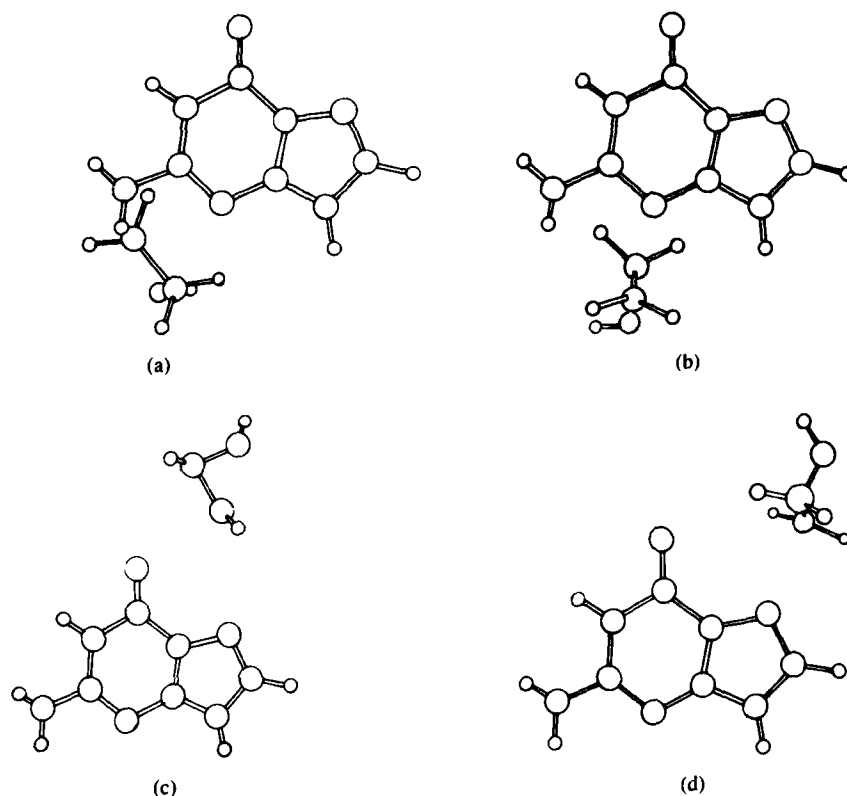
Figure 2.   Calculated transition state geometries for attack of protonated oxirane at the
guanine (a) N^2-; (b) 3-; (c) O^6-; and (d) 7-positions.

$$E_{el} = \frac{1}{2} \sum_{i}^{\text{guanine}} \sum_{j}^{\text{oxirane}} \frac{q_i q_j}{r_{i,j}} \tag{1}$$

transition state geometries for attack at the specified site. However, in the first the
charges were calculated for each moiety unperturbed by the other. In the second, des-
ignated E_{el}', the charge distribution was taken directly from the calculated transition
state itself. Both reveal trends in qualitative accord with the published electrostatic
potential maps of guanine [18] while the difference between them, δE_{el}, provides a
quantitative measure of the additional attractive interaction arising from their mutual
polarization. The latter is evidently (Table III) greatest for electrophiles approaching
in the vicinity of the O^6- and 7-positions and least for approach to the amino group.
From the calculated charge distribution the high polarizability at the former positions
is associated with increased weights of dipolar resonance forms akin to 7 and 8.

TABLE II. Calculated transition state properties for the gas-phase bimolecular hydroxy-alkylation of 2 by protonated oxirane.[a]

Alkylation site	r_1	r_2	θ	ϕ	iv^b	$\delta q(CH_2)^c$ $\times 10^3$	Σq^d $\times 10^3$
N^2-	3.34	1.96	154	84	409	308	1
3-	2.97	1.93	148	83	439	303	3
O^6-	2.87	1.90	139	81	447	311	4
7-	2.48	1.84	152	78	538	257	38
None[e]	—	1.99	—	86	386	307	—

[a]Bond lengths in Angstroms and bond angles in degrees. Definition of r_1, r_2, θ, and ϕ:
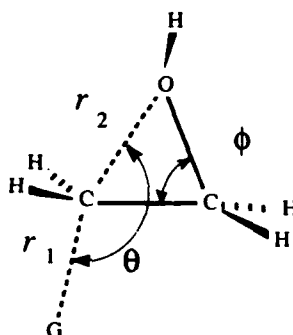
[b]Frequency of imaginary vibration interconverting reactants and products (cm^{-1}).
[c]Charge at the reacting CH_2 group in the transition state relative to 1 (0.409).
[d]Total charge associated with guanine moiety.
[e]Unimolecular ring opening.

TABLE III. Extents of bond formation and cleavage,[a] electrostatic interactions,[b] and deformation energies[c] in the calculated transition states.

Alkylation site	$\Delta n_1\%$	$\Delta n_2\%$	E_{el}	E'_{el}	δE_{el}	$\delta \Delta H_f(1)$	$\delta \Delta H_f(2)$
N^2-	0.1	86.5	-1.2	-5.1	-3.9	14.5	1.7
3-	0.4	84.3	-3.2	-9.3	-6.1	14.3	3.0
O^6-	0.4	82.7	~16.1	~25.4	-9.3	14.1	0.8
7-	2.3	76.7	~15.3	~23.4	-8.1	13.2	0.8
None[d]	—	89.2	—	—	—	—	—

[a]Orders of the forming (r_1) and breaking (r_2) bonds as a percentage of the net change for the reaction (c.f. Ref. 12): $\Delta n\% = 100[\exp(-r^t/0.26) \sim \exp(-r^R/0.26)]/[\exp(-r^P/0.26) - \exp(-r^R/0.26]$
[b]E_{el} electrostatic interaction between guanine and oxirane moieties in the transition states. C.f. Eq. (1) and text. $\delta E_{el} = E'_{el} - E_{el}$
[c]$\delta \Delta H_f(1)$, $\delta \Delta H_f(2)$: heats of formation of the protonated oxirane and guanine moieties in the corresponding transition states, relative to those of the equilibrium geometries of the isolated molecules.
[d]Unimolecular ring opening.

Also included in Table III are the amounts by which the heats of formation of the individual reacting moieties in their transition state geometries differ from those of the isolated molecules. These again are indicative of a situation in which the oxirane ring, although residing in the electrostatically attractive regions of the guanine, opens with little nucleophilic assistance. In reaching the transition state geometry, the guanine moiety undergoes distortions equivalent to only 1–3 kcal mol^{-1} while the energetic requirements of the oxirane moieties are very similar to the activation energy for unimolecular ring opening.

Discussion

The very early, or S_N1-like character in the transition states for the gas-phase bimolecular ring opening of protonated oxirane (1) was evident in our earlier calculations for simple nucleophiles [11]. A parallel situation for related reactions in aqueous solution had been suggested some time earlier by Parker and Isaacs who designated them "borderline" S_N2 processes [19]. While calculations for the gas-phase reactions should be of significant value in interpreting those in aqueous solution, a quantitative extrapolation between the two phases naturally involves uncertainties. We can be reasonably confident that the unimolecular ring opening of 2 will be less facile in aqueous solution than in the gas phase. Thus, the specific intermolecular hydrogen bond which contributes ca. 18 kcal mol^{-1} to the hydration-free energies of protonated ethers [9] will be significantly weakened as cleavage of the CO bond begins to dissipate the positive charge formally associated with the onium group. Using ideas deduced from our earlier work on the alkanediazonium ions [8] raising the endothermicity of the unimolecular reaction implies an increased role of covalent interactions in the transition state for the corresponding bimolecular process. This in turn is expected to facilitate attack at the guanine 7- relative to the O^6-position. The activation enthalpies for attack of protonated oxirane at these positions in the gas phase are predicted to be almost identical. While if the foregoing arguments are correct, hydration effects should tip this balance in favor of 7-alkylation. Indeed only 7-alkylated products have been reported for the reactions of oxirane and its simple alkyl derivatives with deoxyguanosine in aqueous solution [20, 21]. Thus, the apparent absence of O^6-alkylation may well be explicable in these terms. Guanine 3-alkylation has not been reported for simple epoxides although minor adducts of other alkylating agents at this site have been detected [1]. The least favorable alkylation site is predicted to be the exocyclic amino group, N^2. Again, alkylation by simple epoxides has not been reported at this site and, only with very rare exceptions is it a target for related alkylating agents [1]. Interestingly, however, styrene oxide reacts with deoxyguanosine in aqueous solution to give ca. 12% and 28% of the corresponding O^6 and N^2 adducts 9 and 10, in addition to products resulting from attack at the 7-position [22]. Moreover, for a number of aryl epoxides derived from polycyclic aromatic hydrocarbons, the corresponding N^2 adducts are the major reaction products [5]. We have previously suggested that in aqueous solution, adduct formation at the N^2-position should be more favorable than implied by the gas phase calculations due to specific hydrogen bonding to the polar N^+-H bonds of the nitrogen undergoing quaternization [9]. While not manifest in the reactions of simple alkylating agents, this effect could be-

come significant for the more reactive aryl epoxides, where the intrinsic differences between the activation energies at alternative sites are less. We have yet to scrutinize the reactions of aryl epoxides theoretically. However, on the basis of the present results we are tempted to suggest an additional factor which is expected to reduce the difference between the activation barriers for guanine 7 and N^2 attack in the case of aryl epoxides. As noted in the previous section, the greater electrostatic attraction experienced by electrophiles approaching the former site is due, in part, to the greater polarizability of the guanine undergoing this mode of attack. This effect should, in turn, be most pronounced in transition states involving the aliphatic epoxides where the positive charge is highly localized on the reacting methylene group. For aryl epoxides, where the corresponding position is a benzylic carbon, the charge will be significantly delocalized and polarization effects therefore less important. An effect of this kind is probably responsible for the clear shift from 7- to N^2-benzylation of deoxyguanosine observed by Dipple and co-workers [23, 24] for a series of benzyl chlorides (11a–e) bearing increasingly electron-donating substituents.
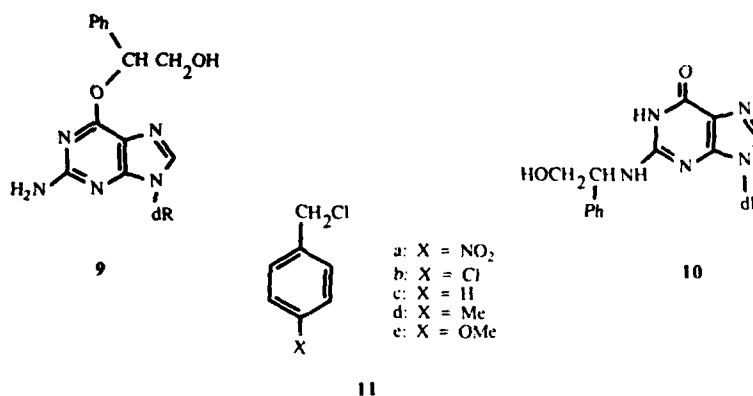
a: X = NO₂
b: X = Cl
c: X = H
d: X = Me
e: X = OMe

9   11   10

Acknowledgment

Bibliography

[1] B. Singer and D. Grunberger, Molecular Biology of Mutagens and Carcinogens. (Plenum, New York), 1983.
[2] K. Hemminki, Arch. Toxicol. 52, 249 (1983).
[3] IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. Allyl Compounds, Aldehydes, Epoxides, and Peroxides, vol. 36. (International Agency for Research on Cancer, Lyon, 1985), p. 189. H. Vaino, K. Hemminki, and J. Wilbourn, Carcinogenesis 6, 1653 (1985).
[4] L. Ehrenberg and S. Hussain, Mutat. Res. 86, 1 (1981).
[5] A. Dipple, R. C. Moschel, and A. H. Bigger, Chemical Carcinogens, ACS Monogr. 182, vol. 1, C. E. Searle, Ed. (American Chemical Society, Washington, D.C., 1984), pp. 41-174.

[6] W. F. Busby, Jr., and G. N. Wogan, *Chemical Carcinogens*, ACS Monogr. 182, vol. 2, C. E. Searle, Ed. (American Chemical Society, Washington, D.C., 1984), pp. 945–1136.

[7] G. P. Ford and J. D. Scribner, J. Am. Chem. Soc. **103**, 4281 (1981).

[8] G. P. Ford and J. D. Scribner, J. Am. Chem. Soc. **105**, 349 (1983).

[9] G. P. Ford and J. D. Scribner, J. Org. Chem. **48**, 2226 (1983).

[10] G. P. Ford, J. Am. Chem. Soc. **108**, 5104 (1986).

[11] G. P. Ford and C. T. Smith, Int. J. Quantum Chem. **13**, 107 (1986).

[12] G. P. Ford and C. T. Smith, J. Am. Chem. Soc. **109**, 1325 (1987).

[13] G. P. Ford and C. T. Smith, J.C.S. Chem. Commun. 44 (1987).

[14] G. P. Ford and C. T. Smith, J. Comput. Chem. (in preparation).

[15] M. J. S. Dewar and W. Thiel, J. Am. Chem. Soc. **99**, 4899, 4907 (1977).

[16] S. Olivella, QCPE Bull. **4**, 109 (1984).

[17] The transition state geometries shown in Figure 2 represent those of lowest energy for attack at each site. We also located a series of slightly higher energy (0–1 kcal mol^{-1}) transition structures with the alternative configuration about the oxirane oxygen. Further transition states, related to those shown in Figure 2 by rotations of *ca.* 90° about the forming bond, were also located. The corresponding activation energies via these were: 3-, 12.0 kcal mol^{-1}; O^6-, −4.4 and −4.7 kcal mol^{-1}; 7-, −3.5 kcal mol^{-1}. In the interest of brevity the individual structures are not discussed in detail.

[18] R. Bonaccorsi, E. Scrocco, J. Tomasi, and A. Pullman, Theoret. Chim. Acta **36**, 339 (1975).

[19] R. E. Parker and N. S. Isaacs, Chem. Rev. **59**, 737 (1959).

[20] P. Brooks and P. D. Lawley, J. Chem. Soc. 3923 (1961).

[21] P. D. Lawley and M. Jarman, Biochem. J. **126**, 893 (1972).

[22] K. Hemminki and A. Hesso, Carcinogenesis **5**, 601 (1984); K. Hemminki, 7th Annual Interdisciplinary Cancer Research Workshop, Univ. New Orleans, Feb. 24, 1984.

[23] A. Dipple, R. C. Moschel, and W. R. Huggins, Drug Metab. Rev. **13**, 249 (1982).

[24] For an alternative statement of somewhat similar ideas see: R. C. Moschel, W. R. Huggins, and A. Dipple, J. Org. Chem. **44**, 3324 (1979).

Application of the Quantum Mechanics and Free Energy Perturbation Methods to Study Molecular Processes

PIOTR CIEPLAK,* U. CHANDRA SINGH,† AND PETER A. KOLLMAN

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, California 94143, U.S.A.

Abstract

The molecular dynamics free energy perturbation method was applied to study the solvation effect on the tautomeric equilibria in water solution as well as association of the nucleic acid base pairs in water solution and *in vacuo*. Tautomerization energies *in vacuo* calculated by the *ab initio* SCF-HF method differed from experiment by 1–2 kcal/mol, even if geometry optimization was performed and MP2 correlation energy calculated at $6\text{-}31G*$ basis set was added.

Introduction

A new approach applying quantum mechanics (QM) together with molecular dynamic free energy perturbation (FEP/MD) [1, 2] methods has been used to investigate chemical processes such as tautomeric equilibria and the association of the nucleic acid bases in vacuo and in water solution.

The gas-phase energy differences, were calculated for 2-oxo-pyridine (I), 2-oxo-pyrmidine (II), and cytosine (III) tautomers (Fig. 1) by the *ab initio* Hartree–Fock method in the extended 6-31G* basis set for geometries optimized at 3-21G level. For the test case of 2-oxo-pyridine tautomers, geometry optimization and the calculation of the correlation energy by the MP2 [3] method with a $6\text{-}31G*$ basis set were also performed. The FEP/MD method was applied to study the hydration effect on tautomeric equilibria. Combined results of the QM and the FEP/MD method give results in good agreement with experiment [4–7].

Preliminary calculations [8] using the FEP/MD method for nucleic acid base association in water and *in vacuo* have been carried out. Five complexes were considered: adenine-thymine Watson-Crick H-bonded pair, adenine-thymine stacked, adenine-adenine stacked, guanine-cytosine Watson-Crick H-bonded, and guanine-cytosine stacked. The stacked complexes were calculated to be slightly more stable than the H-bonded in water, whereas *in vacuo* the H-bonded complexes are favored, which is consistent with experiment [9].
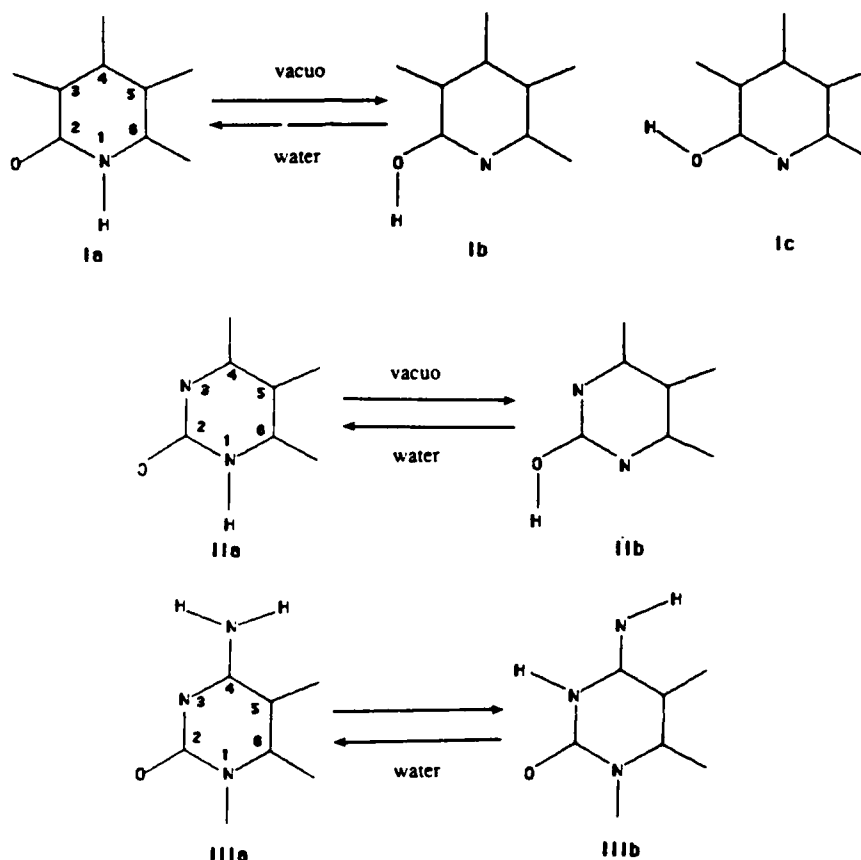
Figure 1.   Tautomeric structures, atom nomenclature, and numbering scheme for 2-oxo-
pyrydine (I), 2-oxo-pyrimidine (II), and cytosine (III).

This FEP/MD method seems to be general and is a powerful method to obtain equi-
librium constants and Gibbs free energy differences of solvation from simulations.

Free Energy Perturbation Method Formalism

Several statistical mechanical procedures have been developed to compute free en-
ergy differences for solutions [10–12]. The recent applications [1, 12–19] of the ther-
modynamic free energy perturbation method confirm its general applicability in
Monte Carlo (MC) as well as molecular dynamics (MD) simulations. It is also a robust
tool for calculation of the relative free energy differences in the general case, not
only when solvation processes are of interest. Here we summarize the theoretical ba-
sis of the original method given by Zwanzig [20].

In a canonical ensemble (e.g., generated by standard MC or MD with constant temperature) the Helmholtz free energy is:

$$F = -kT \ln Z \tag{1}$$

where Z is the partition function determined by the Hamiltonian $H(p, q)$

$$Z = \frac{1}{h^{3N}N!} \int \exp\left(-H\frac{(p,q)}{kT}\right) d\mathbf{p}\, d\mathbf{q}. \tag{2}$$

The aim is to calculate free energy difference ΔF between systems (or states) A and B whose Hamiltonians differ by perturbation ΔH:

$$\mathbf{H_B} = \mathbf{H_A} + \mathbf{\Delta H} \tag{3}$$

this yields:

$$\Delta F = F_B - F_A = -kT \ln \frac{Z_B}{Z_A} = -kT \ln \frac{\int \exp\left(\frac{(-H_A + \Delta H)}{kT}\right) d\mathbf{p}\, d\mathbf{q}}{\int \exp\left(\frac{-H_A}{kT}\right) d\mathbf{p}\, d\mathbf{q}} \tag{4}$$

and:

$$\Delta F = -kT \ln \frac{\int \exp\left(\frac{-H_A}{kT}\right) \exp\left(\frac{-\Delta H}{kT}\right) d\mathbf{p}\, d\mathbf{q}}{\int \exp\left(\frac{-H_A}{kT}\right) d\mathbf{p}\, d\mathbf{q}} \tag{5}$$

One should keep in mind that, in general, this last equation can be obtained from Eq. (4) if and only if H_A and ΔH commute, which is true in the case when the classical expression for the Hamiltonian of the system under consideration is used. In this case we can rewrite Eq. (5) as follows:

$$\mathbf{\Delta F} = -\mathbf{kT} \ln \left\langle \exp\left(\frac{-\mathbf{\Delta H}}{\mathbf{kT}}\right)\right\rangle_A, \tag{6}$$

where $\langle\ \rangle_A$ means average over the system A. In the P, T ensemble Gibbs free energy ΔG is obtained. Expression (6) can be used immediately in Monte Carlo and molecular dynamics simulations to calculate ΔF.

Direct calculation of the expectation value in equation (6) via MC or MD encounters convergence problems if ΔH is large. To overcome this problem the umbrella sampling method developed by Torrie and Valleau [21] can be used. Alternatively, one can combine many small perturbations coupled to a dimensionless parameter λ along the path between A and B and sum up free energy differences obtained for each of these small perturbation steps. This is the basis for the *windowing* procedure, that is:

$$\Delta G_i = -kT \ln \left\langle \exp\left(\frac{-\Delta H(\lambda \to \lambda')}{kT}\right)\right\rangle_\lambda \tag{7}$$

$$\Delta G = \sum_{\lambda=0}^{\lambda=1} G_i. \tag{8}$$

For each window separate MD (or MC) simulations with equilibration and data collection stages are performed. If infinitesimal steps in λ are chosen, the $\langle \rangle_\lambda$ does not fluctuate (i.e., the system can be regarded to be in equilibrium) this yields:

$$\Delta G_i = (H_{\lambda'} - H_\lambda)_\lambda.$$ (9)

This expression is the essence of the so called *slow growth* procedure in which one needs only to define the total time (in the case of MD) or total number of configuration to be generated (in the case of MC) during which conversion from state A to state B proceeds.

Methods

In our calculations, Gaussian-80-UCSF [22] and Gaussian-82 [23] programs were used for the *ab initio* calculations. Partial atomic charges used in MD simulation were calculated by fitting to the electrostatic potential obtained from *ab initio* 6-31G* [24] wave functions [25] for isolated molecules.

The molecular simulation program AMBER-UCSF (Version 3.0) [2] with its force field was used to calculate free energy differences by the FEP/MD method. Molecular dynamic simulations were carried out at T = 300 K. In the solution calculations the TIP3P [26] water model was used with 1 atm constant pressure and periodic boundary conditions. Complete computational details are given in Refs. 8 and 19.

The free energy perturbation method was incorporated into the molecular dynamics module of the AMBER program in the following way [1, 2]. The molecular mechanical energy is calculated according to the following formula:

$$E_{\text{Total}} = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)]$$
$$+ \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\varepsilon R_{ij}} \right] + \sum_{\text{H-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right]$$ (10)

where the first three terms represent the difference in energy between a geometry in which the bond lengths, bond angles, and dihedral angles have ideal values and the actual geometry. The remaining terms represent nonbonded-van der Waals and electrostatic interactions. The last term (10-12) is used for atoms involved in hydrogen bonding. To represent changes or "mutations" of a given group of atoms from state A to state B coupled to the dimensionless parameter λ the appropriate values of K_r, r_{eq}, K_θ, θ_{eq}, E_{dihedral}, q_i, A_{ij}, B_{ij}, C_{ij}, and D_{ij} for a given λ were calculated according to the linear interpolation prescription:

$$X(\lambda) = \lambda X^A + (1 - \lambda)X^B$$ (11)

Since the Helmholz free energy is a state function, the computed free energy changes are path independent (i.e., should not depend on the manner in which the "mutation" from state A to B is performed). We also assume that in our calculation the changes in ΔH are mainly due to the potential energy and contributions from kinetic energy change cancel out and/or are negligible.

The above procedure was tested extensively for several cases where accurate experimental data are available. For example the calculated [1, 13] $\Delta\Delta G$ hydration between CH_3OH and CH_3CH_3 is equal to 6.6–6.9 kcal/mol, whereas the experimental value is 6.9 kcal/mol. In the case of $\Delta\Delta G$ for the reaction $NH_4^+(H_3O)_3 \rightarrow H_3O^+(H_3O)_3$ 20.0 kcal/mol was obtained [1] versus 21.3 kcal/mol for the experimental value. If the above reaction was performed during MD simulation in water bath the value of 21.1 kcal/mol was calculated [1] which is also very close to the experimental value of $\sim$ 20 kcal/mol. The above scheme was also applied to study the solvation of amino acids and nucleic acid bases [28] as well as to calculate energetics of protein inhibitor binding [17] and site-specific mutagenesis phenomena [29]. In this paper we present some results obtained by the FEP/MD method applied to study tautomerism and association phenomena in water solutions.

In our specific examples, the transformation of one tautomer into another in water solution was carried out by changing the parameters in such a way to achieve the simultaneous vanishing tautomeric proton in one position (e.g., oxygen) and growing this proton in another position (e.g., nitrogen).

To model association (or rather dissociation) process for the system A-B *in vacuo*, the charges, and the van der Waals parameters for one of the molecules were decreased to zero during the MD simulation. To model the solvation of A and B or the A-B complex, a simulation was done on the system with the solute fully represented and then its electrostatic and van der Waals parameters decreased to zero. The results were evaluated with the thermodynamic cycle in Figure 2.

Tautomerism

In Table I the tautomerization energies are summarized. Zero-point energy differences $\Delta\epsilon_0$ between tautomers were estimated by the MINDO/3 method and since they are in the range 0.6–0.9 kcal/mol they cannot be neglected because correction of this magnitude significantly influences the tautomerization energy. We have neglected the differences which arise from temperature dependence of vibrational and rotational energies and entropies because at the MINDO/3 level these are less than 0.1 kcal/mol and 0.2 kcal/mol, respectively.

Our *ab initio* calculations are extensions of the work done by Scanlan and Hillior [29] for cytosine tautomerism and Schlegel [30] on the 2-oxopyridone/2-hydrox-
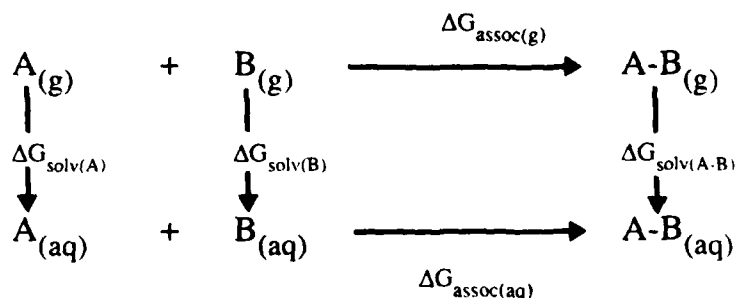
Figure 2. Thermodynamic cycle used to calculate ΔG_{assoc} of nucleic base pairs in water.

TABLE I. Tautomerization energies ΔE (kcal/mol) (defined as a difference between appropriate total energies obtained from *ab initio* calculations) for 2-oxo-pyridine (I), 2-oxopyrimidine (II) and cytosine (III) isomers in the gas phase calculated at different levels.

	$\Delta E =$ $E(\text{Ib}) - E(\text{Ia})$	$\Delta E \approx$ $E(\text{Ic}) - E(\text{Ia})$	$\Delta E =$ $E(\text{IIb}) - E(\text{IIa})$	$\Delta E =$ $E(\text{IIIa}) - E(\text{IIIb})$
a) HF/3-21G				
geometry optimization	1.67	17.19	1.14	−0.39
b) HF/6-31G* geometry				
optimized in 3-21G	0.55	6.69	−1.85	−0.61
c) HF/6-31G*				
geometry optimization	0.10	6.14		
d) MP2/6-31G* geometry				
optimized in HF/6-31G*	−1.61	4.40		
Zero point vibration				
energy differences				
(from MINDO/3)	−0.73	−0.85	−0.63	0.58
Estimates[a] of b)	−0.18	5.84	−2.48	−0.03
tautomerization c)	−0.63	5.29		
energy d)	−2.34	3.55		
Experimental $\Delta\Delta G$	−0.4 ± 0.7 (IR)[4]	—	−2.4[4]	
	−0.5 ± 0.8 (UV)[4]			
	−0.6 ± 0.1 (X − PES)[5]			

[a]Values given under labels b, c, d refer to the calculations done, b) with HF/6-31G* and geometry optimized in 3-21G basis set, c) HF/6-31G* geometry optimization, and d) MP2/6-31G* geometry optimized in 6-31G* basis set, respectively.

ypyridine equilibrium. It was also found that to obtain proper results for tautomerization energies in the gas phase, the proper choice for geometries of the tautomers, and method for the energy calculation are needed.

Table I shows that enhancing the 3-21G basis set to 6-31G* improves agreement with experiment for the 2-oxo-pyridine tautomerism. Further geometry optimization at the 6-31G* basis set moves the position of tautomerization energy in the right direction and closer to the observed value, but still the keto form is more stable. Addition of the correlation energy calculated within the 6-31G*/MP2 level stabilizes the hydroxy form by 1.7 kcal/mol. This overestimate is opposite in direction to that obtained by Schlegel et al. [30] at the 6-31G/MP2 level, where the keto form was stabilized by 0.8 kcal/mol.

The overall agreement of the calculated data for the gas phase tautomeric equilibria for 2-oxo-pyrimidine is satisfactory, since the estimated error for the experimental results is not known.

Experimental data for the gas-phase tautomerization energy of cytosine is not known and our calculations suggest that significant amounts of both forms should be observable in the gas phase.

In Table II we summarize the results for the free energies and equilibrium constants obtained here and in experiments. The $\Delta\Delta G_{tot}$ (column 5), which is the free

TABLE II. Comparison of the free energy differences (kcal/mol) and equilibrium constants for tautomerization in the H_2O solution at 300 K obtained from MD simulation and experiment.

n	Tautomer pair	NH_2O in MD simulation	$\Delta\Delta G_{solv}$	$\Delta\Delta G_{tot}$	K_{calc}	$\Delta\Delta G_{exp}$	K_{exp}
	2-oxopyridine		Ib → Ia	Ib → Ia	Ia/Ib		
			−5.4 ± 0.3	b) − 5.2 ± 0.4	6,390		Ia/Ib
1	Ib ⇌ Ia	600		c) − 4.7 ± 0.4	2,980	−4.1[4]	910
				d) − 3.0 ± 0.4	164		
			Ic → Ib	Ic → Ib	Ib/Ic		
			0.9 ± 0.1	b) − 5.0 ± 0.1	4,200		
2	Ic ⇌ Ib	584		c) − 5.0 ± 0.1	4,200		
				d) − 4.9 ± 0.1	3,500		
			Ic → Ia	Ic → Ia	Ia/Ic		
			−4.4 ± 0.1	b) − 10.2 ± 0.1			
3	Ic ⇌ Ia	584		c) − 9.7 ± 0.1			
				d) − 7.9 ± 0.1	711,000		
			Ic → dIa	Ic → Ia	Ia/Ic		
		based on	−4.4 ± 0.4	−10.2 ± 0.4			
3'	Ic ⇌ Ia	1 and 2		c) − 9.7 ± 0.4			
				d) − 7.9 ± 0.4	711,000		
	2-oxopyrimidine		IIb → IIa				
			−5.5 ± 0.4	IIb → IIa	IIa/IIb		IIa/IIb
4	IIb ⇌ IIa	578		−2.9 ± 0.2	154	<−1.6[4]	>15 (ethanol)
	cytosine						IIIa/IIIb
			IIIb → IIIa	IIIb → IIIa	IIIa/IIIb	−6.3[6]	39,810
5	IIIb ⇌ IIIa	588	−4.2 ± 0.2	−4.2 ± 0.2	1235	−4.1[7]	1.000

Different level of calculations − b, c, d as in Table I. $\Delta\Delta G_{tot}$ is defined by the equation 12. $\Delta\Delta G_{solv}$ is the value calculated by Eq. 7, K is experimental or calculated from $\Delta\Delta G$ equilibrium constant.

energy difference between tautomers in solution, is estimated according to the following equation:

$$\Delta\Delta G_{tot} = \Delta\Delta G_{solv} + \Delta\epsilon_0 + \Delta E_{QM} + \Delta S_{QM}. \tag{12}$$

where $\Delta\Delta G_{solv}$ is the solvation free energy difference between the tautomers calculated by the FEP/MD simulations, $\Delta\epsilon_0$ is the difference in zero-point vibrational energies between the two tautomers (from MINDO/3), ΔE_{QM} is the calculated gas phase tautomerization energy and ΔS_{QM} is the difference in the entropies which could be calculated by quantum mechanical methods, but which we assume to be negligible, given the MINDO/3 results noted above. In general, there is qualitative agreement in all cases except for the cytosine tautomers, where the experimental results are rather inconclusive and depend on the measurement method [6, 7]. One general conclusion that can be derived from our calculations is that the keto and amino forms of these molecules prevail in water solutions mainly because of the solvation effect.

The robustness of our method is supported by the fact that the sum of the free energies for the three simulations involving the cyclic transformations Ia → Ib, Ib → Ic, Ic → Ia, which should be exactly 0 kcal/mol is indeed 0.0 ± 0.4 kcal/mol. This accuracy is achieved despite a difference in the number of water molecules in the simulations.

The effect of solvation on the isomerization of the Ib to Ic isomer also shows a competition between intrinsic energies and solvation effects. Table II, simulation no. 2 shows that the solvation free energy difference and the zero-point vibration energy contribution favors Ic (angle N-C-O-H = 180°), but the internal electronic energy is lower by 6.0 kcal/mol for the Ib isomer, thus, in net, the Ib form is favored.

In the case of the equilibrium IIa ⇄ IIb, only the data measured in ethanol solution are available. However, changing solvents from ethanol to water the keto form, with its higher dipole moment, will be preferentially stabilized. This will make the difference in free energy more negative than -1.6 kcal/mol and, thus, closer to our calculated value.

After our calculations were completed, we have learned that Kwiatkowski et al. [31, 32] have carried out *ab initio* calculations of the relative gas phase energies for the nucleic acid bases at a similar level as our calculations for 2-oxo-pyridine. They found that the MBPT(2) correlation correction calculated at the 6-31G* basis set level stabilizes the amino-keto (IIIa) isomer over keto-imino (IIIb) cytosine by 0.7 kcal/ mol. If this correction were added to our estimate of the IIIa–IIIb tautomerization energy in solution, it would become -2.3 rather than -1.6 kcal/mol and yield a result closer to the experimental data.

Association of Base Pairs

In Table III the association free energies for base pairs obtained by the FEP/MD simulations are collected and compared with the available experimental data [9]. The $\Delta G_{assoc(aq)}$ was calculated from the thermodynamic cycle presented in Figure 2 by subtracting $\Delta G_{solv(A)}$ and $\Delta G_{solv(B)}$ (taken from Ref. 27) from the sum of the $\Delta G_{assoc(g)}$ and $\Delta G_{solv(A-B)}$. Each of those numbers were obtained from numerical FEP/MD simulations, thus they should be regarded as "first principle" results.

TABLE III.   Association free energy (kcal/mol) for nucleic acid base pairs in H-bonded and stacking configurations in the gas phase and water solution in T = 300 K.

	A − T H-bond	A/T stack.	G − C H-bond	G/C stack.	A/A stack.
Calculated $\Delta G_{assoc(g)}$	-1.51 ± 0.23	0.38 ± 0.15	-6.78 ± 0.16	-2.61 ± 0.05	1.29 ± 0.12
Calculated $\Delta G_{assoc(aq)}$	0.18	-1.86	-0.91	-2.22	-2.16
Experimental $\Delta G_{assoc(aq)}$	—	-1.15	—	-0.71	-1.80

Experimental data were taken from Ref. 9.

The numerical results for $\Delta G_{assoc(g)}$ presented in Table III were corrected for changes in vibrational entropies during complex formation in T = 300K and for expressing ΔG at different standard state (i.e., 1 atm versus 1 Molar state, to which experimental data are referred) to be compared with experimental values.

FEP/MD results show that the method is able to properly predict that H-bonded configurations are preferred in the gas phase for adenine-thymine and guanine-cytosine pairs as well as the preference in water solution for the stacked configurations. This is because in solution, water can form H bonds with the bases in the stacked configuration that are comparable in strength to base-base hydrogen bonds.

For the $\Delta G_{assoc(aq)}$ no errors bars are given in Table III, but we expect these to be of the order of 2–3 kcal/mol. Despite these large errors our simulations give reasonable values for the absolute ΔG of solution associations which are in the range of experiment.

Conclusions

The FEP/MD method seems to be general and is found to be the only acceptable method to obtain equilibrium constants and free energy differences from simulations. This method can be regarded also as a supportive and a predictive tool for the experimentalist, especially in the field of site-specific mutagenesis, drug-DNA specific interactions, and solvation studies. Our MD simulation results for the calculation of the solvation free energy differences between two isomers give a rather good agreement with experiment and suggests that they can be used to predict equilibrium constants in other cases where no experimental data is available.

Acknowledgments

References

[1] U. C. Singh, F. K. Brown, P. A. Bash, and P. A. Kollman, J. Am. Chem. Soc. in press, 1987.

[2] U. C. Singh, P. K. Weiner, J. W. Caldwell, and P. A. Kollman, AMBER (UCSF), version 3.0, Department of Pharmaceutical Chemistry, University of California, San Francisco, 1986.

[3] C. Moller and M. S. Plesset, Phys. Rev. 46, 618 (1934).

[4] P. Beak, F. S. Fry Jr., and F. Steele, J. Am. Chem. Soc. 98, 171 (1976).

[5] R. S. Brown, A. Tse, and J. C. Vederas, J. Am. Chem. Soc. 102, 1174 (1980).

[6] M. Dreyfus, O. Bensaude, G. Dodin, and J. E. Dubois, J. Am. Chem. Soc. 98, 6338 (1976).

[7] A. R. Katrizky and A. J. Waring, J. Chem. Soc. 3046 (1963).

[8] P. Cieplak and P. A. Kollman, "Calculation of the Free Energy of Association of Nucleic Acids in Water," to be submitted to J. Am. Chem. Soc. (1987).

[9] N. I. Nakano and S. J. Igarashi, Biochemistry 9, 577 (1980).

[10] J. P. Valleau and G. Torrie in Statistical Mechanics, Part A, B. J. Berne, Ed. (Plenum Press, New York, 1977), p. 169.

[11] A. Pohorille and L. R. Pratt, in Methods in Enzymology, 1986.

[12] J. P. M. Postma, H. J. C. Berendsen, and J. R. Haak, Faraday Symp. Chem. Soc. 17, 55 (1982).

[13] W. L. Jorgensen and C. Ravimohan, J. Chem. Phys. 83, 3050 (1985).

[14] B. L. Tembe and J. A. McCammon, Comput. Chem. 8, 281 (1984).

[15] T. P. Lybrand, I. Ghosh, and J. A. McCammon, J. Am. Chem. Soc. 107, 7793 (1985).

[16] T. P. Lybrand, J. A. McCammon, and G. Wipff, Proc. Natl. Acad. Sci. USA 83, 833 (1986).

[17] P. A. Bash, U. C. Singh, F. K. Brown, R. Langridge, and P. A. Kollman, Science 235, 574 (1987).

[18] P. Cieplak, T. P. Lybrand, and P. A. Kollman, "Calculation of Free Energy Changes in Ion-Water Clusters Using Non-additive Potentials and Monte Carlo Methods," J. Chem. Phys. (1987) (in press).

[19] P. Cieplak, P. Bash, U. C. Singh, and P. A. Kollman, "A Theoretical Study of Tautomerism in the Gas Phase and Aqueous Solution: A Combined Use of "State-of-the-Art" ab initio Quantum Mechanics and Free Energy Perturbation Methods," J. Am. Chem. Soc. (1987) (submitted).

[20] R. W. Zwanzig, J. Chem. Phys. 22, 1420 (1954).

[21] G. Torrie and J. P. Valleau, J. Comp. Phys. 23, 187 (1977).

[22] U. C. Singh and P. A. Kollman, G80-USCF, QCPE Program #446, 1982.

[23] J. S. Binkley, R. A. Whiteside, K. Raghavachari, R. Seeger, D. J. DeFrees, H. B. Schlegel, M. J. Frisch, J. A. Pople, and L. R. Kahn, Gaussian-82, Carnegie-Mellon University, 1982.

[24] P. C. Hariharan and J. A. Pople, Theor. Chim. Acta 28, 213 (1973).

[25] U. C. Singh and P. A. Kollman, J. Comput. Chem. 5, 129 (1984).

[26] W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, J. Chem. Phys. 79, 926 (1983).

[27] P. A. Bash, U. C. Singh, R. Langridge, and P. A. Kollman, Science, (1987) (in press).

[28] S. N. Rao, P. A. Bash, U. C. Singh, and P. A. Kollman, "Free Energy Perturbation Calculations on Binding and Catalysis: Mutating Asn 155 in Subtilisin," Nature (1987) (submitted).

[29] J. M. Scanlan and I. H. Hillier, J. Am. Chem. Soc. 106, 3737 (1984).

[30] H. B. Schlegel, P. Gund, and E. M. Fluder, J. Am. Chem. Soc. 104, 5347 (1982).

[31] J. S. Kwiatkowski, W. B. Person, K. Szczepaniak, and M. Szczesniak, Acta Biochimica Polonica (1986).

[32] J. S. Kwiatkowski, R. J. Bartlett, and W. B. Person, J. Am. Chem. Soc. (1986).

Spin Density Distribution in Oxygen-Liganded Model Heme Proteins: Predictions of ^{17}O Hyperfine Broadening of ESR Spectra of Metmyoglobin, Cytochrome C Peroxidase, Catalase, and Cytochrome P450

GILDA LOEW AND JACK COLLINS

SRI International, 333 Ravenswood Ave., Menlo Park, California 94025, U.S.A.

LEK CHANTRANUPONG AND AHMAD WALEH

Rockefeller University, Molecular Theory Lab, 701 Welch Road, Suite 213, Palo Alto, California, U.S.A.

Abstract

In heme proteins, the axial ligands bound directly to the iron are important modulators of biological function. A common spectroscopic technique used to detect the presence of heme units with oxygen-containing ligands, is the broadening of the electron spin resonance (ESR) spectrum by hyperfine interactions with unpaired spin density in ^{17}O-enriched systems. To be useful as a means of identifying such ligands, there must be a measurable level of unpaired spin density on the oxygen ligand. In this study, we have used the semiempirical INDO/ROHF quantum mechanical method to calculate and compare the spin density localized on the axial oxygen ligand in the active site of four model heme proteins, Metmyoglobin, cytochrome c peroxidase (CCP), P450cam, and catalase. In particular, we have attempted to determine for which systems the results of such an experiment would be a reliable indicator of the presence of water or other types of oxygen-containing ligands. Using the MetMb system, for which such broadening has been observed, to determine a threshold value of spin density on the oxygen atom needed to detect broadening of the ESR spectra, we have found one-hundredth less spin on the water ligand in P450cam, thus explaining the observed lack of broadening in the ESR spectra of the low spin resting state. In addition, we predict that the catalase system would, in principle, exhibit ^{17}O broadening of its ESR spectra but that CCP would not. Finally, given the similarity of CCP and HRP (horse radish peroxidase), our calculations suggest that the absence of broadening in the ESR spectra of HRP does not rule out the presence of water as a sixth ligand.

Introduction

All heme proteins share a common active site or prosthetic group consisting of an iron porphyrin (heme) unit complex which is a nearly planar entity embedded in the globular protein. The family of heme proteins can be divided into three classes according to their primary biological function: (1) oxygen transport proteins such as myoglobin and hemoglobin, (2) electron transfer agents such as the cytochromes, and (3) oxidative metabolizing enzymes such as the peroxidases, catalases, and cytochrome-P450s. In all classes, the biological function is centered on the heme unit and primarily on the iron itself [1-5]. Thus, the oxidation and spin state of the iron, the nature of the axial ligands, and the protein environment of the heme unit serve as subtle

modulators of biological behavior. The heme group is also the principal origin of spectroscopic features of these proteins. Both electronic spectra [6–9] and ground-state electromagnetic properties, such as quadrupole splitting in Mossbauer resonance spectra [10–14], anisotropic g values and anisotropic hyperfine splitting in electron and nuclear spin resonance spectra [15–23], and temperature-dependent magnetic moments [24–29], originate almost entirely on the heme unit. Thus, these spectroscopic properties can serve as sensitive probes of the stereoelectronic features of the heme unit central to biological function. These techniques have been very useful in probing such important properties as the oxidation state, the low lying spin states, and the nature of the axial ligands bound directly to the iron in the heme unit of different families of heme proteins. In addition, since the pioneering x-ray crystal structure determination of myoglobin [30], rapid advances in the field of protein crystallography have led to structural elucidation of a large number of heme proteins [30–33]. Paradoxically, while a number of long-standing problems are being resolved by these combined techniques, new ones are emerging. In the work reported here, the techniques of theoretical chemistry are used to help resolve apparently conflicting observations about the nature of the axial ligand bound to the iron which have been deduced from x-ray crystal structure and electron spin resonance (ESR) spectra studies of heme proteins.

Metmyoglobin (MetMb), the oxidized form of myoglobin, has long been known to exist in a high spin, ferric state in which the iron is axially bound to an imidazole group of a nearby histidine residue and to water [27, 31]. However, the nature of the axial ligands in the various classes of metabolizing heme proteins has, at least until very recently, been a much more unresolved question [1–5]. For the cytochrome P450s, a great deal of indirect experimental evidence has accumulated over the years which established a sulfur-containing moiety, most likely cysteine, as one of the axial ligands [5]. This deduction has recently been confirmed by the high resolution x-ray crystal structure determination of a soluble P450, camphor-bound P450-camphor (P450cam) [34]. As also suspected, in this substrate-bound, high spin form, there is no second axial ligand. The question still remained then of the nature of this sixth ligand thought to be responsible for the low spin ground state of the substrate-free, resting state of this enzyme. As recently reviewed [5], detailed spectroscopic measurements, including MCD and optical spectra, appeared to indicate that in the low spin ferric resting state either a nitrogen-containing ligand such as an imidazole, or an oxygen-containing ligand such as a serine or threonine was bound to the iron. Surprisingly, the x-ray structure of camphor-bound cytochrome P450 revealed no such amino acid close enough to be capable of binding as an axial ligand to the Fe [34]. A more recent x-ray structure determination [35] of the substrate-free P450cam supports this original assessment. Instead, in the absence of camphor, water appears to bind as the second axial ligand to the Fe with an Fe-O distance of 2.28 Å. There is also some evidence that the water ligand is part of a self-contained H-bonded network with 4 more water molecules in the substrate-binding site, which do not interact with the lipophilic residues that comprise this site. Few significant protein conformational changes were noted between the substrate free and bound state. Thus, it would appear that this new study has determined that the axial ligands in the resting state of cytochrome P450cam are a cysteine sulfur and a water oxygen moiety.

In an attempt to independently verify the nature of the second axial ligand, the ESR spectrum of P450cam in 65% ^{17}O-enriched H_2O was recently determined [36]. Magnetic nuclear-electron spin interactions, $(\mathbf{I} \cdot \mathbf{A} \cdot \mathbf{S})$, in paramagnetic molecules by electron nuclear dipolar coupling or isotropic contact interactions lead to a hyperfine splitting of their electron spin resonance absortion spectra. In a well developed formalism reviewed in detail elsewhere [37], it has been shown that the magnitude of the splitting is determined by the isotropic and anisotropic coupling constants which are proportional to the spin density at the nucleus of each atomic species. The isotropic contact terms which usually make a major contribution to the coupling constants and hence to observed splittings are non-zero only if the nuclei on which they are centered have non-zero nuclear spins. Thus, for paramagnetic ferric heme proteins, this method is useful as a probe of the identity of axial ligands provided that the ligand has a nuclear spin and also sufficient spin density centered on it to give a coupling constant large enough to result in observable splitting or at least broadening of ESR resonance lines. To use this method as a probe for water as an axial ligand, since ^{16}O does not have a nuclear spin, it is necessary to perform the ESR experiments in water highly enriched with ^{17}O. If broadening or splitting is obtained in such ESR experiments, it can be taken as strong evidence for the presence of an exchangeable O ligand such as water. For P450cam, however, no detectable broadening was observed in any of the three anisotropic ESR lines, not even in the sharp derivative line corresponding to $g = 2$. These negative results are in apparent conflict with the x-ray structure which indicates water oxygen as the most probable axial ligand. By contrast, in earlier work [38], such broadening was seen for MetMb known from x-ray structure [31] to have an H_2O molecule as a sixth ligand. The question addressed in this work is then: Can this apparent inconsistency in the ESR and x-ray data of P450cam be resolved and the origin of differences between P450cam and MetMb understood?

More generally, the question we are addressing is whether negative results in such ESR experiments need always be interpreted as the absence of water as an axial ligand or is there the possibility that in the low spin form of P450 and in other ferric heme systems, there could be an axial water ligand which might not make itself known through broadening of the $g = 2$ line in the presence of ^{17}O-enriched H_2O? In addition to P450cam, no such broadening was seen [38] for another type of metabolizing heme protein, horseradish peroxidase (HRP), a high spin ferric heme protein with an imidazole axial ligand resembling MetMb. It was hence concluded that H_2O is not a sixth ligand in that protein, though other indirect evidence points to its presence. There is as yet no x-ray structure for HRP. However, there is now a 1.7 Å resolution structure for a closely related heme protein, cytochrome c peroxidase (CCP) [33], which does have water as an axial ligand, as well as an imidazole, as in MetMb. Should one expect to see comparable broadening of the ESR signal of this high spin ferric heme protein in the presence of ^{17}O-enriched water? No such studies have as yet been reported.

In an interesting variation of the subject of axial ligands, a recent x-ray structure of another type of oxidative metabolizing heme protein, catalase, has been reported [32]. Peroxidases and catalases, share common first steps in their metabolic cycle [2–4],

both being twice oxidized by peroxide to their active Compound I form. However, while peroxidases have relatively nonspecific substrates, catalases have a high degree of substrate specificity and preference for unsubstituted hydrogen peroxide. This high spin ferric heme protein was found to have a single axial ligand, a tyrosine residue (Tyr357), which, at physiological pH, presumably binds primarily as a phenolate. If some method could be found to use ^{17}O-enriched tyrosine, would there be a detectable broadening of the ESR signal of this protein?
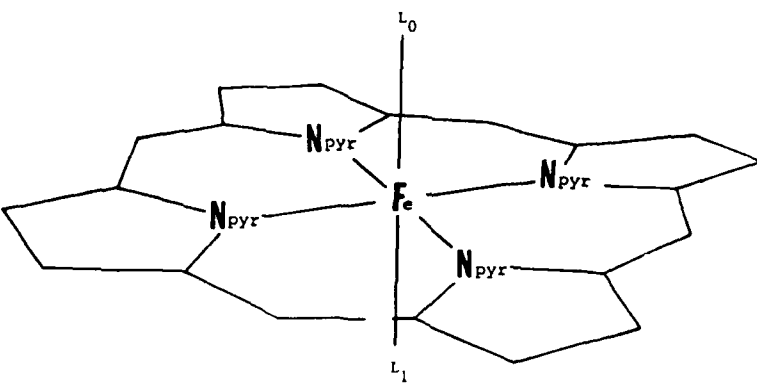
In this study then we have used the x-ray structure of the active site of these four classes of heme proteins, cytochrome P450cam [34] and CCP [33] kindly given to us by Dr. Thomas Poulos, of catalase [32], generously provided by Dr. Michael Rossmann, and of MetMb [31] obtained from the Brookhaven Data Bank [39] to calculate the electron and spin distribution in low-lying sextet, quartet, and doublet states and compare the extent of spin delocalization on the oxygen ligands with that calculated for MetMb for which ^{17}O broadening was detectable. The results have allowed us to predict the systems for which hyperfine broadening of the ESR spectra would be observable, and to what extent this property is a good measure of the presence of oxygen-containing axial ligands.

Methods and Procedures

All studies of the iron-porphyrin complexes have been made using an INDO/SCF/CI program [40] described in detail elsewhere [41–43]. It was developed primarily in the laboratory of Dr. Michael Zerner with the collaboration of this laboratory. This program includes parameterization for transition metals and has configuration interaction capabilities allowing calculation of electronic spectra. Recently, an open shell RHF formalism has been implemented which allows calculation of low-lying states of different multiplicity without spin contamination. This INDO/RHF procedure was used to calculate the electron and spin distribution and relative energies of the lowest lying sextet, quartet and doublet state for each model heme system, corresponding to 5, 3, and 1 unpaired electron states, respectively. A Mulliken population analysis [44] was then used with INDO deorthogonalized orbitals to calculate the electron density on each atomic center. Within the RHF formalism, the electron distribution on each atomic center summed over the open shell molecular orbitals yield, by definition, the spin density distribution on each atomic center.

The model heme active sites used in this calculation for MetMb, CCP, P450cam, and catalase are shown in Table I. They each consist of a porphine ring, the iron, and its axial ligands. For each protein, this active-site geometry was taken from the corresponding x-ray structure as indicated. These active sites represent the resting state of each protein. In the resting state of P450cam, while an oxygen atom has clearly been identified as an axial ligand, the exact nature of this ligand is not totally clear. It is possible that if it is a water molecule involved in an H-bonded network with 4 others, as suggested [35], it could have some degree of anionic, OH$^-$ character. It also appears that while the iron moves toward the heme plane in the camphor-free state, it remains nonplanar toward the cysteine. In previous calculations [43], we have shown that whether the axial ligand is water or hydroxide-like and the extent to which the Fe moves into the heme plane, can determine the high spin/low spin equilibrium in the

TABLE I. Geometry of 4-model heme proteins.

L_0

N_{pyr} N_{pyr} N_{pyr} Fe N_{pyr}

L_1

	[MetMb]$^{+1a}$	[CCP]$^{+1b}$	[P450]^{0c}	[Catalase]^{0d}
Lo	H$_2$O	H$_2$O	H$_2$O	Phenolate
L1	Imidazole	Imidazole	SCH3$^-$	—
Fe—O	1.90 Å	2.40 Å	2.24 Å	1.76 Å
Fe—L	2.02	1.93	2.32	—
Z^e_{Fe}	0.25	0.127	0.24	0.127

[a] X-ray Structure [Ref. 31].

[b] X-ray Structure; oxygen is slightly off axis [Ref. 33]. Long Fe—O bond length could be due to presence of polar HIS52 and TRP51 absent in MetMb.

[c] Coordinate from x-ray of P450 camphorbound P450cam [Ref. 34] with ligand distances from x-ray structure of camphor-free P450cam [Ref. 35].

[d] X-ray structure from Dr. Michael Rossman; oxygen is off axis, < FeOCϕ = 131° [Ref. 32].

[e] Distance the Fe is out of the plane of the porphyrin.

resting state. We have, therefore, calculated the electron and spin distributions for several models of the resting state of P450cam.

Results and Discussion

For MetMb, catalase, and CCP, the high spin state was calculated to be lower in energy than the doublet state, with a quartet state nearly degenerate with the sextet state, consistent with experimentally observed temperature-dependent magnetic moments [25–29]. The high spin ground state obtained for catalase is typical of a five-coordinated ferric–heme complex with a single anionic axial ligand. Such a ground state occurs, for example, in the camphor-bound P450cam in which the single axial ligand is a mercaptide. For camphor-free P450cam, with water and mercaptide as axial ligands, a low spin state is calculated to predominate if the Fe moves into the heme plane and the Fe-O distance shortens to 2.00 Å, or if there is some anionic character in the water ligand. Such charge transfer could be caused by postulated interaction of a water ligand with a network of H-bonded waters or other H-acceptor

groups. For example, if we assume an OH^- as an axial ligand with an Fe-O bond length of 1.75 Å, the low spin state is 16 kcal/mol lower in energy than the high spin state.

From the calculated spin distributions shown in Tables II and III, we see that for no model heme protein, in any spin state, is there a large amount of unpaired spin on the oxygen ligand. While all values are small, systematic variations observed among these very similar systems should be reliable. From the spin distribution on the oxygen ligands shown in Table II, we see that for all model heme systems, the spin density decreases with lower spin states, in the order $S = 5/2 > 3/2 > 1/2$. Also, calculated spin densities are appreciably larger on anionic than on neutral oxygen species.

The water oxygen of the high spin ferric MetMb was calculated to have 0.057e or about 1.1% of the total spin. For this protein, a detectable amount of broadening of the $g = 2$ signal was observed in the ESR spectra in the presence of ^{17}O-enriched H_2O [38]. Thus, the calculated value of oxygen spin densities for the other model heme proteins compared to MetMb can be used as a measure of the relative extent of broadening that would be observed for them.

By far the largest spin density found for an oxygen ligand is that on the phenolate oxygen of the high spin ferric catalase. This result is not surprising since this aromatic anion is more tightly bound to the iron than are the water ligands. Even in this system, however, only 5% of the total spin density is on the oxygen atom. If it is possible to exchange the tyrosine phenol oxygen for ^{17}O and incorporate it into the protein, a significant broadening of the ESR lines should be observed.* Even if the tyrosine is protonated, broadening should be detectable, since the results in Table II show the oxygen spin density on phenol is larger than that calculated for MetMb.

Using the crystal structure geometry, the low spin ferric P450 system, has virtually no spin density on the oxygen of the axial water ligand, with unpaired spin = 0.0005e, 100 times less than that of MetMb. Allowing the Fe—O distance of the water ligand to decrease or simulating its ionic character by OH^-, both lead to low spin ground states and somewhat increase the spin density on the oxygen but it remains at best about 1/6 that of MetMb. Since the broadening in MetMb was barely detectable, no measurable effect on ESR spectra would be observed for any low spin model of P450cam currently proposed. These results then account for the absence of a detectable broadening of the P450 ESR signal in the presence of high concentration of protein and $^{17}OH_2$ 65% in a manner consistent with the presence of water as an axial ligand in the resting state of P450cam.

Turning now to the fourth model heme protein considered, CCP, the greatly diminished spin densities on the water oxygens in the high spin state compared to MetMb, allow us to predict that no detectable broadening will be seen due to ^{17}O-enriched H_2O in the ESR spectra of CCP. Since CCP is similar to HRP, these results also imply

---

*A plausible procedure for preparation of ^{17}O-enriched tyrosine is to prepare it by use of phenylalanine hydroxylase in the presence of gaseous $^{17}O_2$ with phenylalanine as a substrate. The ^{17}O-enriched tyrosine could then be incorporated into bacterial catalase using a tyrosine-deficient mutant form (H. Beinert; private communication).

TABLE II. Spin Density[a] on oxygen ligand in different spin states of four model heme proteins

	[MetMb]$^{+1}$	[CCP]$^{+1}$	[P450]^0	[Catalase]^0
L	Imidazole	Imidazole[b]	SCH3$^-$	—
Lo	H$_2$O	H$_2$O	H$_2$O	Phenolate
S = 5/2	0.057	0.017	0.01(0.03)[c]	0.246
3/2	0.030	0.016	— (0.03)[c]	0.208
1/2	0.004	0.0004	0.0005(0.004)[c]	0.052
Lo		OH$^-$	OH$^-$	Phenol
Fe—O		2.40	1.75	1.76
S = 5/2		0.08(.07)[d]	0.16	0.088
3/2		0.07(.06)[d]	0.01	0.077
1/2		0.0004(.0003)[d]	0.01	0.03

[a] Spin densities given in units of e- and distances in Å.
[b] With imidazole anion as L no significant change in spin on oxygen.
[c] Calculated for Fe—OH$_2$ distance of 2.00 and Fe in plane.
[d] Values obtained with L=imidazole anion, L$_0$=OH$^-$.

TABLE III. Spin density[a] distribution in high and low spin states of four model heme proteins

Lo Ll	[MetMb] + 1 H$_2$O Imidazole		[CCP] + 1 H$_2$O Imidazole		[P450]0 H$_2$O SCH3$^-$		[Catalase]0 Phenolate	
	5/2	1/2	5/2	1/2	5/2	1/2	5/2	1/2
Fe	4.29	0.94	4.25	0.93	4.20	0.93	4.18	0.92
Ll								
(N) (S)	0.09	0.001	0.08	0.003	0.04	0.04	—	—
other	0.04	0.002	0.03	0.009	0.26	0.00	—	—
Lo								
(O)	0.057	0.004	0.017	0.0004	0.01	0.0005	0.25	0.05
other	0.017	0.000	0.004	0.0002	0.01	0.0000	0.05	0.003
Porphyrin	0.51	0.053	0.62	0.06	0.48	0.04	0.52	0.028

[a] Spin densities given in units of e-.

that water can be a sixth ligand in that system despite the absence of a detectable ^{17}O broadening. One of the reasons for the diminished spin density on the water oxygen on CCP, compared to MetMb, could be the much longer Fe—O bond length, 2.4 Å in CCP compared to 1.9 Å in MetMb. This bond lengthening has been attributed [1] to the presence of polar residues around the water-binding site which can H-bond to it in CCP; residues which are absent in MetMb. To simulate this effect, the CCP system was also calculated with an OH$^-$ as a sixth ligand. In this system, the spin density on the oxygen increased to a value larger than that for MetMb. If there is such appreciable charge donation to the water in CCP or if the experiments are done

at high pH, it is possible that broadening of the CCP and, by analogy, the HRP spectra would be observed, another possible indication of a water ligand in the latter system. However, under these conditions the peroxidases are likely to become low spin and the spin density on the oxygen would again be too low to detect (Table II).

There are also polar groups around the histidine residue in CCP which are absent in MetMb. As shown in Table II, simulating their effect by an anionic imidazolium ligand, does not appreciably change the spin distribution on the oxygen.

In summary, we have calculated and compared the spin density localized on the axial oxygen ligand in the active site of four model heme proteins, MetMb, CCP, P450cam, and catalase, using the x-ray structure geometries. Our results reveal no spin density on the water oxygen atom of the low spin state of P450cam. This resolves, in a consistent manner, the lack of observation of ^{17}O broadening in the ESR spectra of this system with the crystal structure determination of the presence of a water ligand in the resting state. Using the MetMb system to determine a threshold value of spin density on the oxygen to cause detectable broadening of the g = 2 line in the ESR spectra; we predict that the catalase system would, in principle, exhibit ^{17}O broadening of its ESR spectra but that CCP would not. Given the similarity of HRP to CCP, we also conclude that the absence of broadening of the ESR spectra of that enzyme does not rule out the presence of water as a sixth ligand, as it does in CCP. Further experiments should test these inferences made from small but systematic variations found in the spin distribution of these heme systems.

Acknowledgments

Bibliography

[1] T. L. Poulos and B. C. Finzel, in *Peptide and Protein Reviews*, M. T. W. Hearn, Ed. (Marcel Dekker, New York, 1984) p. 115.

[2] H. B. Dunford and J. S. Stillman, Coord. Chem. Rev. **19**, 187 (1976).

[3] G. R. Schonbaum and B. Chance in *The Enzymes*, 3rd ed., P. D. Boyer, Ed. (Academic Press, New York, 1976) pp. 363–408.

[4] W. D. Hewson and L. P. Hager in *The Porphyrins*, Vol. 7, D. Dolphin, Ed. (Academic Press, New York, 1979) pp. 295–332.

[5] J. H. Dawson and K. S. Eble, Advances Inorg. Bioinorg. Mech. **4**, 1-64 (1986).

[6] M. Gouterman in *The Porphyrins*, Vol. 3, D. Dolphin, Ed. (Academic Press, New York, 1978), p. 1.

[7] W. A. Eaton and J. Hofrichter, Methods Enzymol. **76**, 175 (1981).

[8] M. W. Makinen and W. A. Eaton, Nature (London) **247**, 62 (1974).

[9] M. W. Makinen, A. K. Churg, Y. Shen, and S. C. Hill in *Proceedings of the Symposium on Interaction between Iron and Proteins in Oxygen and Electron Transport*, C. Ho et al., Eds. (Elsevier North Holland, New York, 0000).

[10] T. H. Moss, A. Ehrenberg, and A. Bearden, Biochem. **8**, 4159 (1969).

[11] G. Lang, T. Asakura, and T. Yonetani, J. Phys. C **2**, 2246 (1969).

[12] Y. Maeda, A. Trautwein, U. Gonser, K. Yoshida, K. Kikuchi, T. Homma, and Y. Ogura, Biochim. Biophys. Acta **303**, 230 (1973).

[13] Y. Maeda, Y. Morita, and K. Yoshida, J. Biochem. **70**, 509 (1971).

[14] M. Sharrock, P. G. Debrunner, C. Schulz, J. D. Lipscomb, V. Marshall, and I. C. Gunsalus, Biochim. Biophys. Acta **420**, 8-26 (1976).

[15] J. P. Collman, T. N. Sorrell, and B. M. Hoffman, J. Am. Chem. Soc. **97**, 913 (1975).

[16] N. Harada and T. Omra, J. Biochem. (Tokyo) **87**, 1539 (1980).

[17] G. Lang, Atomic Energy Research Establishment Report No. 6171 (1969).

[18] G. Lang, T. Asakura, and T. Yonetani, Atomic Energy Research Establishment Report No. 6170 (1969).

[19] B. A. Wittenberg, L. Kampa, J. B. Wittenberg, W. E. Blumberg, and J. Peisach, J. Biol. Chem. **243**, 1863 (1968).

[20] M. Tamura and H. Hori, Biochim. Biophys. Acta **284**, 20 (1972).

[21] K. Torii and Y. Ogura, J. Biochem **64**, 171 (1968).

[22] J. E. Bennett, J. F. Gibson, and D. J. E. Ingram, Proc. Roy. Soc. London. Ser. **A240**, 67 (1957).

[23] R. L. Tsai, C. -A. Yu, I. C. Gunsalus, J. Peisach, W. E. Blumberg, W. H. Orme-Johnson, and H. Beinert, Proc. Natl. Acad. Sci. U.S.A. **66**, 1157 (1970).

[24] M. Tamura, Biochim. Biophys. Acta **243**, 239 (1971).

[25] M. Tamura, Biochim. Biophys. Acta **243**, 249 (1971).

[26] K. Yoshida, T. Iizukas, and Y. Ogura, J. Biochem. **68**, 849 (1970).

[27] A. Tasaki, J. Otsuka, and M. Kotani, Biochim. Biophys. Acta **140**, 284 (1967).

[28] K. Torii, Y. Ogura, J. Otsuka and A. Tasaki, J. Biochem. **66**, 791 (1969).

[29] T. Iizuka, M. Kotani and T. Yonetani, J. Biol. Chem. **246**, 4731 (1971).

[30] J. Kendrew, Nature, London **185**, 422 (1960).

[31] T. Tanaka, J. Mol. Biol. **110**, 537 (1977).

[32] Complete set of 2.0 Å refined coordinates were kindly provided by Dr. Michael Rossmann. T. J. Ried, III, M. R. N. Murthy, A. Sicignano, N. Tanaka, W. D. L. Smusick and M. G. Rossmann, Proc. Natl. Acad. Sci. (U.S.A.) **78**, 4767 (1981), M. R. N. Murthy, T. J. Reid, III, A. Sicignano, N. Tanaka and M. G. Rossmann, J. Mol. Biol. **152**, 465 (1981).

[33] Complete set of 1.7 Å refined coordinates were kindly provided by Dr. Thomas L. Poulos, B. D. Finzel, T. L. Poulos, and J. Kraut, J. Biol. Chem. **259**, 13027 (1984).

[34] Complete set of 1.7 Å refined coordinates were kindly provided by Dr. Thomas L. Poulus. T. L. Poulus, B. C. Finzel, I. C. Gunsalus, G. C. Wagner, and J. Kraut, J. Biol. Chem. **260**, 16122 (1985).

[35] T. L. Poulus, B. C. Finzel, and A. J. Howard, Biochem. **25**, 5314 (1986).

[36] V. Ullrich, P. M. H. Kroneck, and H. Beinert, private communication.

[37] J. Fajer and M. S. Davis, *The Porphyrins*, Vol. IV, *Physical Chemistry*, D. Dolphin, Ed. (Academic Press, New York, 1979) p. 198.

[38] S. Vuk-Pavlovic' and Y. Siderer, Biochem. Biophys. Res. Comm. **79**, 885 (1977).

[39] F. C. Bernstein, T. F. Koetzle, J. B. Graheme, F. M. Edgar, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, J. Mol. Biol. **112**, 535 (1977).

[40] A. D. Bacon and M. C. Zerner, Theor. Chim. Acta **53**, 21 (1979).

[41] M. C. Zerner, G. H. Loew, R. F. Kirchner, and U. T. Miller-Westerhoff, J. Am. Chem. Soc. **102**, 589 (1980).

[42] W. D. Edwards, B. Weiner, and M. C. Zerner, J. Am. Chem. Soc. **108**, 2196 (1986).

[43] G. H. Loew, J. R. Collins, B. T. Luke, A. Waleh, and A. Pudzianowski, Enzyme **36**, 54 (1986).

[44] R. S. Mulliken, J. Chem. Phys. **23**, 1833, 2338, 2343 (1955).

Equilibrium Geometry and Electrical Polarizability of Formic Acid, Formamide and Their Cyclic Hydrogen-Bonded Pairs

MAGDALENA DORY, JOSEPH DELHALLE, JOSEPH G. FRIPIAT, and
JEAN-MARIE ANDRE

Laboratoire de Chimie Théorique Appliquée, Facultés Universitaires Notre-Dame de la Paix, 61, Rue de Bruxelles, B-5000 Namur, Belgium

Abstract

The equilibrium geometry, stabilization energy, and electric polarizability of formic acid, formamide, and the three possible cyclic hydrogen-bonded pairs are obtained by *ab initio* calculations using the STO-3G, 4-31G, and 6-31G** bases. These three properties are found to be very much dependent on the basis set extension. The polarizability of the dimers is found to be basically additive in contribution from the monomeric moieties.

Introduction

Increasing attention is being given to studies of the dielectric and electronic properties of biological materials and to the ways in which they interact with electromagnetic energy [1,2]. It has also been realized that the nonlinear optical properties of some biological organic molecules were measurable [3,4], and potentially useful in quantum electronic devices [5]. One characteristic feature of many biological materials is to adopt a specific order due to the presence of hydrogen bonds. In the design of new synthetic materials with interesting electroactive properties the control of the organization at the molecular level is very important [6]. Thus it can be beneficial to learn from nature ways to achieve molecular organizations of specific interest.

Various approaches aimed at molecular organization exist and are actively studied: Langmuir-Blodgett film deposition, topochemical solid-state polymerization, liquid crystal formation, and so on. In the case of liquid crystalline systems, chain ordering in the nematic phase has recently been demonstrated for an organic system containing a carboxylic group which dimerizes by hydrogen bond formation [7]. Likewise, the property of hydrogen bonding to impose specific patterns to the molecular organization has been used elegantly in attempts to engineer molecular ferromagnets [8]. These two examples suggest potential applications of the hydrogen bond to control the structure of conjugated organic chains promising for their electrooptic properties. For example, phenomenological models predict an L^3 dependence of the polarizability with re vect to the length L of the conjugation path [9–12]. In long chains, however, conformational freedom can result in defects and twists which tend to destroy the extent of the conjugation and thus spoil the expected benefit of the L^3 behavior.

One possibility for preventing and/or minimizing the occurrence of these undesirable effects could be to prepare "long" conjugated chains with carboxylic and/or amide end groups capable of hydrogen bond pairing between neighboring chains as schematically indicated in Figure 1. Small chain systems of that type do exist and have been extensively studied from the point of view of their molecular packing [13, 14]. The synthesis of longer conjugated chains with end groups capable of forming hydrogen bonds is conceivable, but whether the overall chain conjugation will be interrupted, reduced, or enhanced by the cyclic hydrogen-bonded pairs remains an important factor to know beforehand. Indications can be obtained from the polarizability of such pairs compared to that of their monomers.

The components of the electrical polarizability tensor can in principle be estimated from quantum mechanical calculations, but serious conceptual and computational problems related to its practical evaluation are still unsolved [15]. Moreover, as indicated by numerous studies on simple compounds, the properties of hydrogen-bonded systems are rather difficult to obtain compared to those of more strongly bonded systems. The presently suggested computational strategies to approach such problems tend also to differ: to some authors large basis sets cannot be avoided, while others suggest that more limited basis sets can be used in a sensible way provided polarization functions are included in the basis describing the hydrogen atoms. Finally, a question that has received comparatively less attention is the effect of the geometry reorganizations taking place in the paired monomers and how to relate their properties to those of the isolated constituents when applying methods to correct for the basis set superposition error.

All these unsolved questions have relevance to the present work, and in spite of this largely unsettled situation, we have calculated the equilibrium geometry and electrical polarizability of formic acid, formamide, and their three possible cyclic hydrogen-bonded pairs. The results reported hereafter are meant to serve as reference for further studies on related but more realistic systems which, for practical reasons, will imply the use of basis sets of limited extent. Our expectations are that the qualitative trends obtained on the energies, most stable geometries, and electrical polarizabilities are reliable. The average polarizability has been calculated for the equilibrium geometries corresponding to the three basis sets (STO-3G, 4-31G, and 6-31G**) considered. In all cases a correction to the basis set superposition error (BSSE) has been applied using the original form of the counterpoise (CP) method.
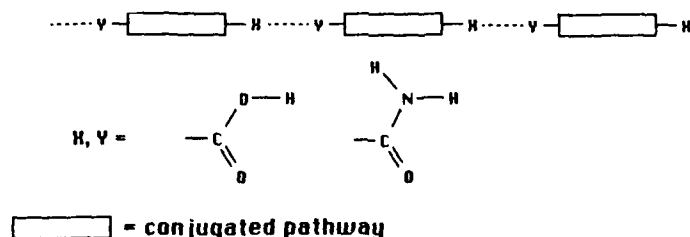
Figure 1.   Schematic representation of conjugated skeletons connected by cyclic hydrogen bonded-pairs involving carboxylic and amide functional groups.

Methodology

Calculations on the monomers and dimers, shown in Figure 2, have been performed with the Gaussian 82 series of programs [16], adapted for an FPS-164 processor attached to an IBM 4341 computer. A code for the finite-field steps has been added to the Gaussian 82 software. All the geometric parameters, denoted according to the general convention in Figure 3, for the molecules shown in Figure 2 and *constrained to be planar*, were optimized for each of the basis sets considered, namely, the minimal STO-3G basis set, the split valence 4-31G basis, and the 6-31G** split-valence basis including polarization functions [17]. The 6-31G** polarization basis set has been selected because it was designed for the description of weakly bonded systems such as those in which hydrogen is a bridging atom; it includes a single set of Gaussian p-type functions for each hydrogen atom. The Fletcher Powell procedure was used to minimize the forces on the nuclei. The standard threshold conditions of the Gaussian 82 program have been kept: 10^{-10} a.u. for the two-electron cutoff, 10^{-9} for the requested convergence on the density matrices, and 5×10^{-4} HartreeBohr^{-1} as the maximal residual force on the Cartesian components.

The Cartesian frame with respect to which the diagonal components α_{ii}, $i = x, y,$ and z of the electric polarizability tensor are calculated is given in Figure 3. Polarizability calculations have been performed using the finite-field (FF) approach [18, 19] where a term $-\mu.F$ describing the interaction between an external homogeneous field F and the molecular dipole moment μ is added to the unperturbed molecular Hamiltonian. In the presence of the external electric field, the molecular dipole moment is dependent on F, and the components α_{ij} of the electric polarizability tensor α can be expressed as:

$$\alpha_{ij} = [\partial\mu_i(\mathbf{F})/\partial F_j]_{\mathbf{F}=0}, \qquad i = x, y, \text{ and } z. \tag{1}$$

In practive they are evaluated utilizing an approximate differentiation operator,

$$\alpha_{ii} = [\mu_i(F_i) - \mu_i(-F_i)]/(2F_i) \qquad i = x, y, \text{ and } z, \tag{2}$$
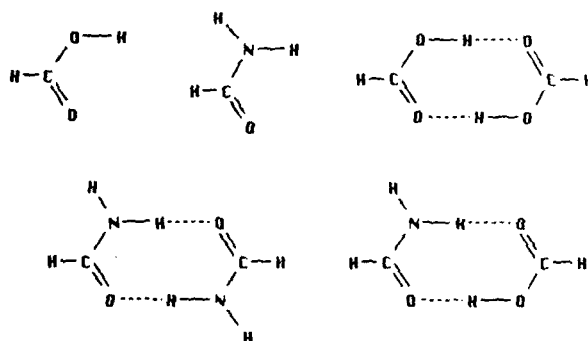
Figure 2. Schematic representation of the formic acid, formamide, and the three complexes considered in the present work
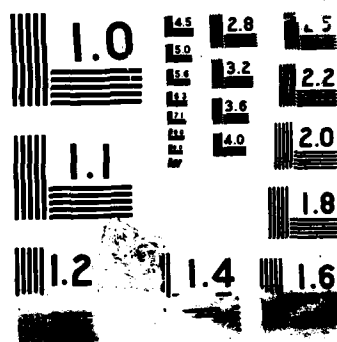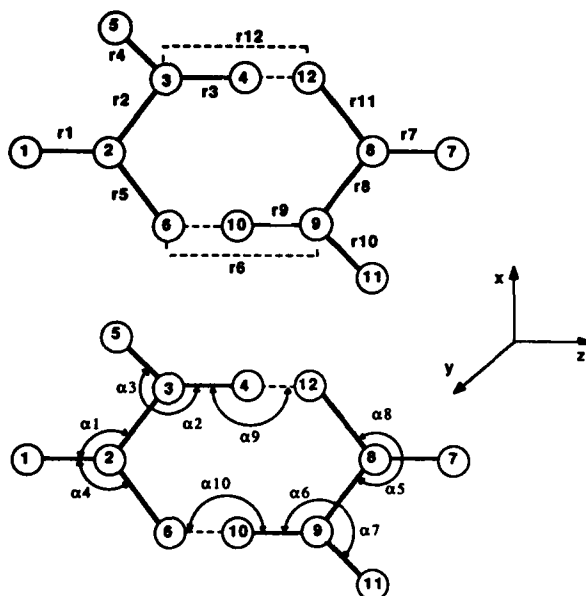
Figure 3. Notational convention for the geometrical parameters of the molecules and dimers optimized in this work.

in conjunction with the Romberg algorithm [20]. The values of the electric field components F_i actually used in the numerical procedure to compute μ_i are equal to ± 0.001 and ± 0.002 au (1 au of electric field $= 5.1423 \times 10^{11}$ Vm^{-1}). As found in other calculations [21, 23], we checked that these external electric field values yield consistent and numerically accurate polarizabilities. The values of the polarizability reported in this article are expressed in au (1 au of polarizability $= 1.6488 \times 10^{-41}$ C^{2}m^{2}J^{-1})

Due to the weakness of the hydrogen bond, incompleteness of the atomic basis set can cause properties such as the interaction energy, optimal geometries, dipole moment, etc. to be significantly in error (basis set superposition error or BSSE). The situation is expected to worsen in the case of the electric polarizability, for which it is known that quite diffuse functions are required. The trouble is that no completely reliable scheme seems to exist yet for eliminating or estimating superposition errors. Many papers [24–33] dealing with this problem have recently appeared in the literature. They basically investigate variants of the full counterpoise correction originally proposed by Boys and Bernardi [24], but no conclusive indication seems to emerge from which a given approach should definitely be preferred. Therefore in this work, we have adopted the conservative choice to use the original CP method [24].

Owing to the fact that important geometry relaxations take place upon dimerization, the counterpoise correction can be applied in many ways. For instance the "ghost" functions could be centered on positions characteristic of the optimized geometry of the isolated monomer or on those of the monomer in the relaxed geometry

of the dimer. Thus to study the influence of different possibilities on the calculated properties (Mulliken charge indices, dipole moments, stabilization energy, and average polarizability), several arrangements of either the molecular skeletons and/or the centers of the "ghost" basis functions, denoted by the capital letters **A** to **Q** in Figure 4, have been considered. In the figure the symbols (m) and (d), respectively, indicate the optimized structure of the isolated monomer and the monomeric moiety as it has relaxed in the cyclic hydrogen-bonded pair. In the hypothetical pairs where the monomeric moieties enter with the geometry of the isolated dimer, namely cases **C, H, M**, and **N**, the H_1-C_2 and C_8-H_7 directions of the partner molecules (or of the set of ghost functions) have arbitrarily been aligned, and the C_2----C_8 distance has been chosen to be that of the corresponding fully optimized dimer.

(a)
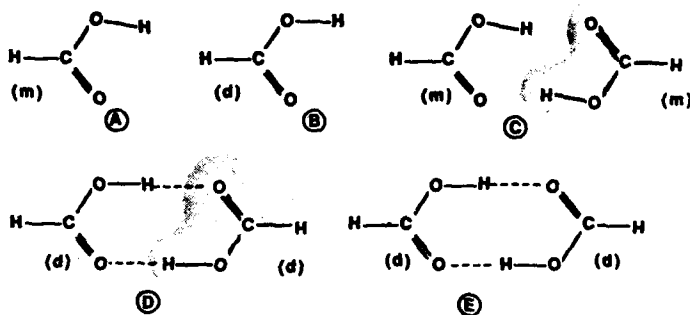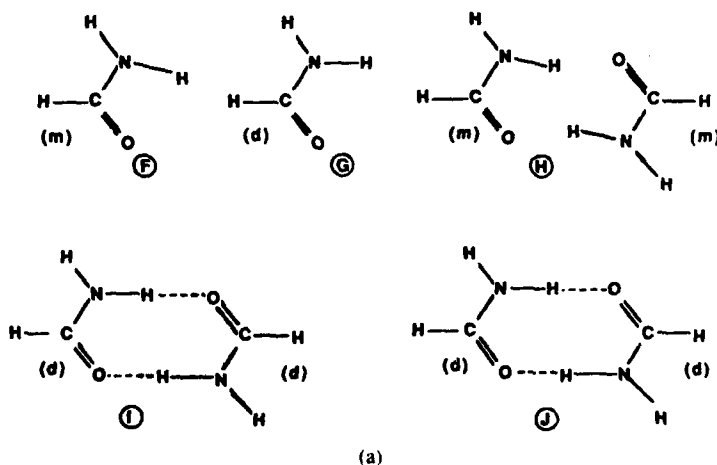
Figure 4. Graphical representation of the various study cases on isolated monomers, monomers in presence of the "ghost" functions (shaded areas in the figure), and dimers. (m) or (d) indicates which optimized geometry is used for monomeric moiety, either that of the isolated monomer (m) or the monomer (d) in the dimer. Capital letters label the different cases.

Constituents of the formic acid and formamide pair.

(b)
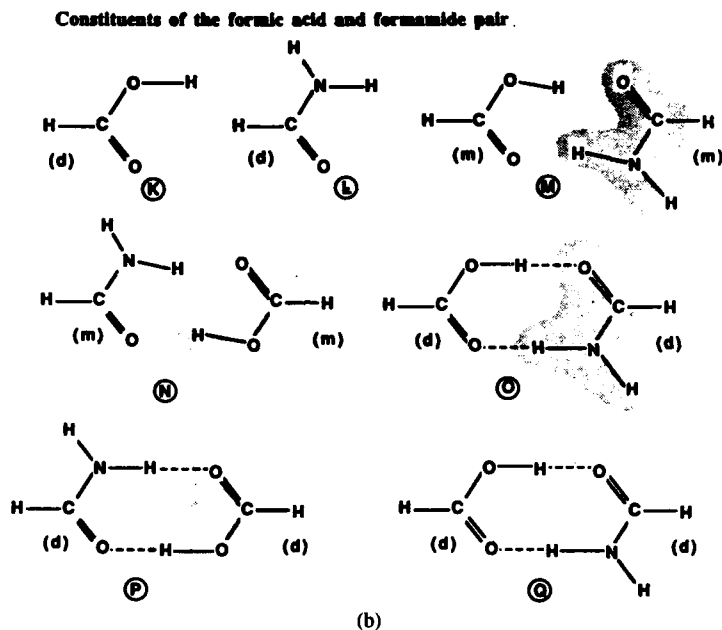
Figure 4.   (Continued.)

Results and Discussion

The molecular geometries predicted from geometry optimization, net atomic charges, overlap populations, dipole moments, and stabilization energies are reported below while the results on the average polarizability are discussed in a later section.

Basis Set Dependence of the Geometry, Charge Indices, and Dipole Moments.

Geometries. Formic acid and formamide are the simplest isoelectronic molecules capable of forming cyclic hydrogen-bonded pairs. As such they have been studied extensively both theoretically and experimentally. Since polarizability is a property sensitive to structure, it is important to use the optimum geometries corresponding to each basis set. To comply with these requirements, we have performed complete geometry optimizations at the STO-3G, 4-31G, and 6-31G** levels for all the molecules represented in Figure 2 where planarity was the only constraint imposed. Imposing planarity is justified by previous calculations, it is only questionable in the case of the formamide monomer for which the pyramidal form is calculated at the STO-3G level to be 2.28 kcal mol^{-1} more stable than the planar variety [34]. However, Sapse et al. [35], using the STO-6G, 6-31G, and 6-31G** bases, find the planar form to be the most stable structure.

Table I shows the optimum geometrical parameters defined in Figure 3 for the five molecules. The electron charge distribution indices — net atomic charges, overlap populations, and dipole moments — belonging to the investigated molecules are displayed in Table II where capital letters A to Q refer to the notation introduced in Fig-

TABLE I. Evolution of the STO-3G, 4-31G, and 6-31G** geometrical parameters, optimized bond lengths (in Å), and angles (in degrees) for formic acid, formamide, and their dimers.

	STO-3G	4-31G	6-31G**	Exper.
Formic acid: monomer and dimer				
r1	1.103 (1.107)	1.072 (1.072)	1.085 (1.084)	1.097[a] (1.082)[b]
r2	1.386 (1.348)	1.342 (1.313)	1.321 (1.298)	1.342 (1.323)
r3	0.990 (1.009)	0.956 (0.975)	0.949 (0.962)	0.972 (1.036)
r5	1.214 (1.231)	1.200 (1.217)	1.182 (1.196)	1.204 (1.220)
r6	(2.536)	(2.708)	(2.789)	(2.703)
α1	110.4 (112.1)	110.4 (111.8)	110.5 (111.5)	112.0 (118.4)
α2	104.8 (108.1)	114.9 (116.8)	108.9 (111.4)	106.3 (108.5)
α4	126.0 (122.2)	125.0 (122.9)	124.7 (122.5)	123.2 (115.4)
α9	(178.9)	(165.7)	(173.9)	
Formamide: monomer and dimer				
r1	1.105 (1.106)	1.081 (1.080)	1.093 (1.091)	1.098[c] (1.010, 1.090*)[d]
r2	1.403 (1.375)	1.346 (1.328)	1.348 (1.331)	1.352 (1.318, 1.326*)
r3	1.014 (1.039)	0.992 (1.006)	0.994 (1.004)	1.002 (0.890, 1.010*)
r4	1.013 (1.014)	0.989 (0.990)	0.991 (0.991)	1.002 (0.870, 1.010*)
r5	1.218 (1.234)	1.216 (1.230)	1.193 (1.205)	1.219 (1.241, 1.239*)
r6	(2.639)	(2.896)	(2.988)	(2.948)
α1	111.4 (116.6)	113.7 (115.1)	112.8 (114.6)	112.7 (114.5, 116.0*)
α2	120.5 (118.6)	119.5 (120.1)	119.1 (119.4)	118.5 (119.6, 119.0*)
α3	121.3 (120.1)	121.9 (121.0)	121.6 (120.6)	120.0 (118.9, 118.0*)
α4	124.3 (119.1)	121.5 (119.7)	122.3 (119.9)	122.4 (120.5, 119.1*)
α9	(171.2)	(168.4)	(168.1)	
Dimer of formic acid and formamide				
r1	(1.106)	(1.080)	(1.090)	
r2	(1.373)	(1.327)	(1.330)	
r3	(1.035)	(1.003)	(1.001)	
r4	(1.014)	(0.990)	(0.991)	
r5	(1.235)	(1.232)	(1.207)	
r6	(2.560)	(2.741)	(2.849)	
r7	(1.107)	(1.073)	(1.085)	
r8	(1.351)	(1.313)	(1.299)	
r9	(1.012)	(0.980)	(0.965)	
r11	(1.229)	(1.215)	(1.194)	
r12	(2.618)	(2.824)	(2.908)	
α1	(113.8)	(112.4)	(112.1)	
α2	(118.1)	(120.2)	(119.7)	
α3	(120.2)	(120.7)	(120.3)	
α4	(122.2)	(122.5)	(122.3)	
α5	(114.4)	(114.2)	(113.5)	
α6	(108.4)	(116.8)	(111.4)	
α8	(119.4)	(120.3)	(120.4)	
α9	(166.2)	(162.2)	(162.5)	
α10	(177.1)	(170.7)	(177.9)	

[a] Ref. 44. [b] Ref. 42. [c] Ref. 45. [d] Ref. 43.

The convention has been introduced in Fig. 3. The geometrical parameters for the dimers are given between parentheses. The experimental values for the formic acid and its dimer are denoted by a and b respectively while those of the formamide equivalents by c and d. The center of inversion in formic acid and formamide dimers yields redundancies such as r3 = r9 etc., which reduces the number of tabulated parameters.

Formic acid: monomer and dimer

Charge	A	B	C	D	E
2	+.255	+.254	+.266	+.266	+.281
	+.620	+.624	+.615	+.619	+.666
	+.591	+.589	+.587	+.587	+.618
3	−.285	−.281	−.281	−.277	−.328
	−.720	−.717	−.730	−.727	−.747
	−.567	−.555	−.580	−.568	−.582
4	+.207	+.212	+.211	+.217	+.273
	+.439	+.450	+.466	+.477	+.513
	+.362	+.371	+.386	+.394	+.418
6	−.251	−.257	−.229	−.234	−.305
	−.554	−.571	−.534	−.551	−.655
	−.521	−.538	−.502	−.518	−.595

Population	A	B	C	D	E
2–3	0.28	0.29	0.28	0.29	0.30
	0.14	0.12	0.14	0.12	0.19
	0.27	0.27	0.27	0.27	0.32
3–4	0.25	0.26	0.26	0.26	0.24
	0.25	0.25	0.25	0.24	0.23
	0.30	0.30	0.30	0.30	0.29
2–6	0.44	0.43	0.44	0.43	0.42
	0.47	0.46	0.45	0.45	0.40
	0.60	0.60	0.59	0.58	0.54
4–12	—	—	(0.04)	(0.04)	0.04
	—	—	(0.02)	(0.02)	0.05
	—	—	(0.02)	(0.02)	0.04
Electric dipole	0.625	0.913	0.887	1.119	0.0
	1.594	1.897	1.655	1.953	0.0
	1.627	1.920	1.678	1.965	0.0
Exp. value:	1.41[a]				

Formamide: monomer and dimer

Charge	F	G	H	I	J
2	+.254	+.242	+.262	+.252	+.262
	+.587	+.583	+.585	+.581	+.602
	+.561	+.558	+.557	+.554	+.568
3	−.438	−.429	−.436	−.426	−.462
	−.904	−.899	−.792	−.918	−.934
	−.730	−.721	−.743	−.734	−.744
4	+.200	+.205	+.203	+.209	+.265
	+.386	+.393	+.418	+.426	+.474
	+.322	+.327	+.343	+.349	+.385
6	−.266	−.274	−.253	−.255	−.312
	−.611	−.623	−.592	−.602	−.682
	−.562	−.576	−.543	−.556	−.617

Table II. (*Continued.*)

Formamide: monomer and dimer

	F	G	H	I	J
Population					
2–3	0.37	0.37	0.37	0.38	0.39
	0.11	0.10	0.12	0.10	0.17
	0.27	*0.26*	*0.27*	*0.26*	*0.31*
3–4	0.35	0.35	0.36	0.35	0.32
	0.31	0.31	0.30	0.30	0.26
	0.35	*0.34*	*0.34*	*0.34*	*0.32*
2–6	0.44	0.43	0.44	0.43	0.42
	0.51	0.50	0.49	0.49	0.46
	0.61	*0.61*	*0.60*	*0.60*	*0.56*
4–12	—	—	(0.02)	(0.02)	0.04
	—	—	(0.03)	(0.03)	0.05
	—	—	*(0.02)*	*(0.02)*	*0.03*
Electric Dipole	2.644	2.797	3.069	2.873	0.0
	4.470	4.645	4.742	4.564	0.0
	4.096	*4.169*	*4.261*	*4.337*	*0.0*
Exp. value:	3.71[b]				

Constituents of the formic acid and formamide pair

	K	L	M	N	O	P	Q
Charge							
2	+.245	—	~	+.263	—	+0.256	+.270
	+.587	—	~	+.585	—	+ .585	+.618
	+.560	—	~	*.557*	—	*+ .557*	*+.576*
3	−.430	—	~	−.435	—	− .427	−.461
	−.896	—	~	−.922	—	− .914	−.929
	−.718	—	~	*−.743*	—	*− .730*	*−.741*
4	+.205	—	—	+.204	—	+ .209	+.265
	+.391	—	—	+.419	—	+ .424	+.475
	+.326	—	—	*+.344*	—	*+ .347*	*+.387*
6	−.274	~	—	−.249	—	− .252	−.310
	−.626	—	—	−.593	—	− .604	−.703
	−.578	—	~	*−.543*	—	*− .558*	*−.630*
10	—	+.213	+.210	—	+.217	—	+.272
	—	+.452	+.464	—	+.479	—	+.513
	—	*+.373*	*+.381*	—	*+.606*	—	*.414*
9	—	−.280	−.282	—	−.276	—	−.331
	—	−.719	−.731	—	−.729	—	−.755
	—	*−.554*	*−.578*	—	*−.569*	—	*−.584*
8	—	+.252	+.264	—	+.263	—	+.272
	—	+.622	+.615	—	+.617	—	+.653
	—	*+.588*	*+.587*	—	*+.584*	—	*+.609*
12	—	−.257	−.235	—	−.238	—	−.305
	—	−.569	−.534	—	−.549	—	−.637
	—	*−.539*	*−.502*	—	*−.517*	—	*−.582*

Table II.   (*Continued.*)

Constituents of the formic acid and formamide pair

	K	L	M	N	O	P	Q
Population							
2–3	0.37	—	—	0.37	—	0.38	0.39
	0.10	—	—	0.12	—	0.10	0.19
	0.26	—	—	*0.27*	—	*0.26*	*0.32*
3–4	0.35	—	—	0.35	—	0.35	0.32
	0.31	—	—	0.30	—	0.30	0.27
	0.34	—	—	*0.34*	—	*0.34*	*0.32*
2–6	0.43	—	—	0.44	—	0.43	0.42
	0.50	—	—	0.49	—	0.49	0.44
	0.61	—	—	*0.60*	—	*0.59*	*0.55*
4–12	—	—	—	(0.03)	—	(0.03)	0.04
	—	—	—	(0.02)	—	(0.02	0.04
	—	—	—	*(0.02)*	—	*(0.02)*	*0.03*
6–10	—	—	(0.03)	—	(0.03)	—	0.04
	—	—	(0.02	—	(0.03)	—	0.06
	—	—	*(0.02)*	—	*(0.03)*	—	*0.04*
10–9	—	0.26	0.26	—	0.26	—	0.24
	—	0.25	0.24	—	0.24	—	0.22
	—	*0.30*	*0.30*	—	*0.29*	—	*0.28*
9–8	—	0.29	0.28	—	0.29	—	0.30
	—	0.12	0.14	—	0.12	—	0.18
	—	*0.27*	*0.27*	—	*0.27*	—	*0.31*
8–12	—	0.43	0.44	—	0.43	—	0.42
	—	0.46	0.45	—	0.45	—	0.41
	—	*0.59*	*0.59*	—	*0.59*	—	*0.56*
Electric dipole	0.917	2.799	0.852	2.911	1.088	3.099	2.293
	1.885	4.665	1.658	4.564	1.943	4.765	3.196
	1.901	*4.282*	*1.681*	*4.170*	*1.948*	*4.360*	*2.687*

[a]Ref. 55.
[b]Ref. 45.

ure 4. Table III lists the total energies E_T corresponding to the situations A to Q, and the stabilization (complexation) energies ΔE defined as the total energy of the dimer minus the energy of the monomers with or without corrections for the basis set superposition error.

Our results on optimum geometries are in most cases in very good agreement with previously published STO-3G [36, 37] and 4-31G [38–40] results on formic acid, formamide, and their respective dimers. We could not find corresponding results for the 6-31G** basis, but our results are reasonably consistent with those obtained by Mijoule et al. [41] on the formic acid dimer calculated with the 6-31G basis in spite of the linearity of the O—H .... O angle (α_9 in our notation) assumed by these authors, and by Sapse et al. on formamide obtained with the 6-31G and 6-31G* basis sets for the monomer and dimer. To the best of our knowledge, geometry optimiza-

TABLE III. STO-3G, 4-31G, and 6-31G** total energies E_T (au) and stabilization energies ΔE (kcal mol^{-1}) for the situations represented in Fig. 4.

Formic acid: monomer and dimer

E_T	A	B	C	D	E
sto-3G	-186.2179	-186.2149	-186.2306	-186.2290	-372.4599
4-31G	-188.4756	-188.4742	-188.4782	-188.4769	-376.9831
6-31G**	-188.7706	-188.7694	-188.7722	-188.7711	-377.5654
ΔE	E-2*A	E-2*B	E-2*C	E-2*D	
sto-3G	-15.13	-18.89	+0.82	-1.19	
4-31G	-20.02	-21.78	-16.76	-18.39	
6-31G**	-15.19	-16.70	-13.18	-14.56	

Exp. value: -16.1 kcal mol^{-1} [a]

Formamide: monomer and dimer

E_T	F	G	H	I	J
sto-3G	-166.6882	-166.6847	-166.6971	-166.6975	-333.3975
4-31G	-168.6816	-168.6807	-168.6840	-168.6833	-337.3906
6-31G**	-168.9405	-168.9397	-168.9421	-168.9414	-337.9020
ΔE	J-2*F	J-2*G	J-2*H	J-2*I	
sto-3G	-13.24	-17.64	-2.07	-1.57	
4-31G	-17.20	-18.33	-14.19	-15.06	
6-31G**	-13.18	-14.19	-11.17	-12.05	

Exp. value -14.0 kcal mol^{-1} [b]

Constituents of the formic acid and formamide pair

E_T	K	L	M	N	O	P	Q
sto-3G	-186.2138	-166.6858	-186.2279	-166.6996	-186.2261	-168.6998	-352.928
4-31G	-188.4734	-168.6806	-188.4781	-168.6840	-188.4761	-168.6832	-357.187
6-31G**	-188.7689	-168.9398	-188.7721	-168.9482	-188.7705	-168.9416	-357.733
ΔE	Q-(A + F)	Q-(B + G)	Q-(K + L)	Q-(M + N)	Q-(O + P)		
sto-3G	-14.25	-18.33	-18.33	-0.82	-1.82		
4-31G	-18.96	-20.40	-20.96	-15.88	-17.64		
6-31G**	-14.25	-15.50	-15.76	-12.24	-13.62		

[a] Ref. 56.
[b] Ref 57.

tions on the heterodimer of formic acid and formamide have not been reported with similar bases.

As already pointed out in the previous theoretical works, the $C=O$ distance in formic acid and formamide moieties increases upon dimerization irrespective of the basis set extension, though at a different rate. At the 6-31G** level, the $C=O$ elongation in the three dimers is equal to 0.012 Å and 0.014 Å in the formic acid and formamide moieties, respectively. As a counterpart to this elongation, there is a shortening in the $C-O$ and $C-N$ distances, respectively, of 0.023 Å and 0.017 Å when calculated at the 6-31G** level. Both the $O-H$ and $N-H$ distances increase, respectively, by about 0.013 Å and 0.017Å in the three dimers, as can be noticed from the 6-31G** data in Table I. Except for the STO-3G results on the formic acid dimer, the hydrogen bonds are slightly bent as noticed from the $O-H\ldots O$ and $N-H\ldots O$ angles, α_9 and α_{10}, listed in the table.

The dimer separations turn out to be very sensitive to the basis set extension. For instance, the $O-H\ldots O$ distance, r_6, in the formic acid dimer increases substantially from STO-3G to 4-31G (0.172 Å), but also from 4-31G to 6-31G** (0.081 Å). A similar trend is observed in the case of formamide where the corresponding changes are 0.257Å and 0.092 Å. In their paper on the formic acid dimer, Hayashi et al. [40] concluded at an excellent agreement between the 4-31G predictions (2.705Å) and the experimental values obtained from electron diffraction measurements (2.703 Å) [42], but this kind of agreement is not reproduced in the case of the formamide dimer: 2.878 Å theoretically and 2.948 Å experimentally [43]. At the 6-31G** level, the $O-H\ldots O$ and $N-H\ldots O$ distances are equal to 2.789 Å and 2.988 Å, respectively, in the dimers of formic acid and formamide. In the heterodimers the corresponding $O-H\ldots O$ and $N-H\ldots O$ distances are equal to 2.849 Å and 2.908 Å. The results by Sapse et al. on the formamide dimer show that the 6-31G and 6-31G* geometrical parameters of the individual moieties in the dimer are comparable with the 6-31G** results obtained in this work. However, the separation between the monomers (measured in Ref. 35 by the diagonal distance C_1N_2 between the carbon in one moiety and the nitrogen in the other) is quite different when calculated at the 6-31G (3.520 Å), 6-31G* (3.786Å) and 6-31G** (3.878 Å) levels. From these results it turns out that large basis sets including polarization functions, 6-31G* and 6-31G**, overestimate the distances between the monomers in the complexes when compared to experiment. Significant deviations occur also in the $C=O$, $C-O$ and $C-N$ bonds which are directly involved in the hydrogen bond formation. Of course one could incriminate the basis set, but this contradicts the overall agreement of the 6-31G** results on the formic acid and formamide monomers with the most recent microwave measurements by Davis et al. [44] and Hirota et al. [45]. Another explanation for these discrepancies might well be that the interactions existing in the bulk preclude a realistic comparison with theoretical calculations on the isolated dimers.

In the Cambridge data base, the only reported single-crystal x-ray diffraction measurements on the cyclic dimer of formic acid are those by Karle and Brockway [46] in 1944. More recent electron diffraction measurements have been published by Almeningen et al. [42], both are reproduced in Table I. For the formamide dimer, a

recent x-ray determination has been made by Stevens [43]. In this work, full x-ray data and geometrical parameters corrected for the bias introduced by aspherical bonding density are given. Both are reproduced in Table I, the corrected data being denoted by a superscript asterisk. Finally, Nahringbauer and Larsson [47] have published x-ray measurements on the crystal structure of the 1:1 addition compound of formic acid with formamide, $HCOOH \cdot \cdot HCONH_2$. The structure contains puckered layers of two types. The layer relevant to our study contains dimers of formamide molecules coupled by hydrogen bonds with a $N — H \ldots O$ distance equal to 2.990 Å and formic acid molecules, which cross-link the dimers by hydrogen bonds (2.602 and 2.946 Å). These measurements suggest an explanation for the systematic difference observed between 6-31G** results and experiment. In the crystal, the molecules are linked by more than one type of hydrogen bond as shown in Figure 5. For instance, in the case of formamide, a $N — H \ldots O$ hydrogen bond with a distance of 2.948 Å connects the molecules and is responsible for the dimer formation, but another hydrogen bond, $N — H \ldots O'$ with a distance of 2.885 Å, joins alternate dimers into chains [48]. These distances are nearly equal and the interactions should thus be of comparable strength, which might explain why the structural parameters for the dimers in the crystal are so different from those in the gas phase. Hinton and Harpool [49] have made investigations of $(formamide)_n$ systems to define models for the liquid state and dilute aqueous solution. Due to the size of the systems, only hydrogen bond lengths were optimized to obtain minimum energy positions. Unfortunately, the authors have not indicated the optimized distances in their
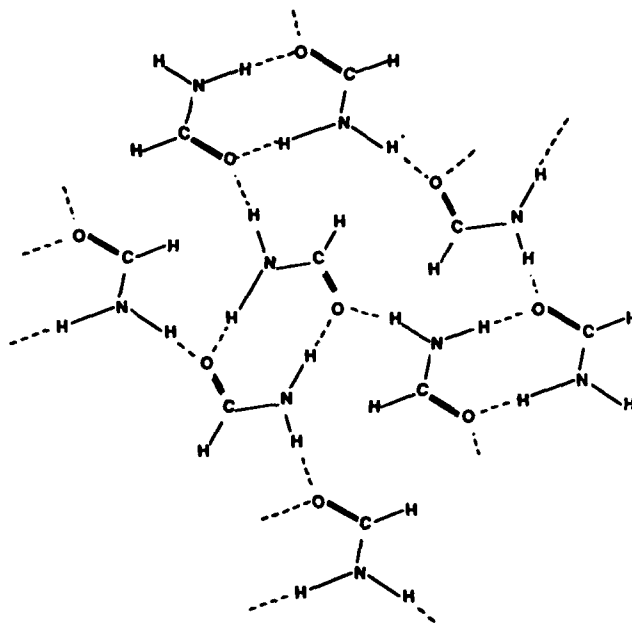

Figure 5.   Arrangement of the formamide dimers in the crystal and indication of their inter-linking through hydrogen bonds.

work, and this can only be a first step since all parameters should be optimized to allow meaningful comparison with experiment.

Work is now in progress to simulate the immediate environment of the dimers and conduct geometry optimizations to assess the incidence of the other hydrogen bonds both on the geometrical parameters of the dimer moieties and on the monomer separation in the cyclic pairs. Drastic influences on the barrier height for single or double proton transfer might result.

Net Atomic Charges, Overlap Populations, and Dipole Moments. Compared to 4-31G and 6-31G** results, the charge separation in the STO-3G basis is less pronounced, however the 6-31G** values are somewhat reduced compared to 4-31G, (see Table II). This corroborates the observation by Sapse et al. [35] who noticed that the 6-31G* charges are smaller than those of 6-31G. The trend continues with the 6-31G** values which are smaller than 6-31G*. Overlap populations reflect the geometry changes occurring in the different molecular arrangements. Except in the case of the minimal STO-3G basis, the net atomic charges are only slightly affected by the addition of the ghost functions and by their locations. It is when the dimerization occurs that the charge increase becomes more important. The carbon and hydrogen atoms involved in the dimerization become positively charged while the charges of oxygen and nitrogen atoms become more negative.

Dipole moments are more sensitive to the presence of extra basis functions and location of their centers. As usual, the minimal basis set underestimates the dipole moments. With the 4-31G and 6-31G** bases the nonvanishing dipoles are more than doubled compared to STO-3G. Notice, however, that while the dipole moment is larger in 6-31G** than in 4-31G in the case of the formic acid monomer, the reverse is true for formamide.

Stabilization Energies. Table III is very illustrative of the dependence of the counterpoise correction for basis set superposition error not only with respect to the quality of the basis set, but also with respect to the geometry of the monomers and the location of the centers of the ghost functions. In the case of the STO-3G basis and for the three dimers, large variations in the stabilization energies are found depending on the way ΔE is defined. Unrealistic stabilization energies are obtained when the counterpoise correction is applied to the minimal STO-3G basis, these differences tend to decrease for larger basis sets. As observed previously in the comparison of 6-31G and 6-31G* stabilization energies of the formamide complex, the STO-3G values for ΔE without counterpoise corrections are remarkably closer to 6-31G** than 4-31G. This was interpreted as a fortunate cancellation of errors for the minimal basis set [35, 50]. On the one hand, the small separation of charge observed for the STO-3G basis decreases the electrostatic contribution to the binding energy and thereby accounts partly for the extra stabilization the correction for superposition error would have provided. This is nicely illustrated by the artificial values of ΔE obtained when the counterpoise correction is applied to the STO-3G results. In spite of this fortuitous cancellation, the STO-3G values for ΔE (free of counterpoise correction) are remarkably and systematically close to the corresponding 6-31G** values for the three dimers.

If one considers the values of ΔE predicted with the 6-31G** basis set, it turns out that there is no unique ordering of the stabilization energies for the three isoelectronic

dimers. Even with such a large basis set, ΔE still depends on the convention chosen to calculate the monomer energies.

Average Electric Polarizability

As indicated in the Introduction, it is important to assess how the transmission of conjugation is affected by the hydrogen bonds in the bridging cyclic dimers. Real molecules will be longer than the simple dimers studied in this work and the use of large basis sets will be intractable in these studies. Qualitative trends are in use provided they are predicted consistently, and in view of our experience with conjugated chains, we could have used STO-3G results Indeed, in the case of pure hydrocarbon chains as well as chains incorporating oxygen and nitrogen[23,51,52], the STO-3G calculated longitudinal polarizabilities follow satisfactorily the 4-31G values to within a reasonably constant scaling factor, ≈ 1.45. However, the numerous warnings in the literature about hydrogen bonds on the one hand, and the sensitivity of the polarizability on the quality of the basis function on the other, suggest a cautious approach to this problem. For instance Karlström and Sadlej [53] indicate that for two interacting water molecules, the basis set superposition results in a considerable increase of the x component of the dipole moment and the xx component of the polarizability tensor for the water molecule which plays the role of the electron pair donor (x is the direction of approach of the molecules).

Our results on the average electric polarizability for the three basis sets and the various molecular arrangements schematized in Figure 4 are listed in Table IV. It can be seen that the polarizability of the distorted monomers (B, G, K and L) is systematically larger than in the equilibrium geometry (A and F), but the largest changes occur when the ghost functions are added (C, D, H, I, M, N, O, and P). A measure of the change in polarizability $\Delta\alpha$ upon dimerization can be obtained from the polarizability of the complex minus the sum of the polarizabilities of the monomers. Various possibilities exist and the results are listed in the table. The variations of $\Delta\alpha$ with respect to its various definitions are similar for each basis set, and, from left to right in the table, $\Delta\alpha$ decreases systematically. Most important is that there is no significant departure from additivity for $\langle\alpha\rangle$ due to the dimerization. This is in agreement with the results by Dykstra and Liu in their study of the electrical properties of hydrogen fluoride and the hydrogen fluoride dimer [54] for which very small changes are observed in the polarizability with the distance between the two monomers. From our results it can be observed that $\Delta\alpha$ tends to decrease when larger basis sets are used. In the simplest way to calculate $\Delta\alpha$, namely, α_E-$2\alpha_A$, α_J-$2\alpha_F$, and α_Q − $(\alpha_A + \alpha_F)$, the ratios of $\Delta\alpha$ (STO-3G)/$\Delta\alpha$(6-31G**) range from 2.33 to 2.64, and are acceptably constant. It can also be noted that $\Delta\alpha$(STO-3G) including superposition corrections for the optimized dimers (third column in the table) compares reasonably well with $\Delta\alpha$(6-31G**) in the first column.

As to the qualitative effects with which we are mostly concerned, it can be concluded the polarizability for the complexes of formic acid and formamide turns out to be an additive quantity in the monomer contributions. Irrespective of the definitions for $\Delta\alpha$, the changes remain small, the largest being observed for the STO-3G results

TABLE IV. STO-3G, 4-31G, an⌐ 6-31G** average polarizabilities α (au) for the situations represented in Fig. 4.

Formic acid: monomer and dimer

α	A	B	C	D	E
STO-3G	8.61	8.78	9.23	9.43	19.17
4-31G	13.57	13.68	14.09	14.25	27.42
6-31G**	14.75	14.84	15.24	15.36	30.24
$\Delta\alpha$	$\alpha_E-2*\alpha_A$	$\alpha_E-2*\alpha_B$	$\alpha_E-2*\alpha_C$	$\alpha_E-2*\alpha_D$	
STO-3G	1.95	1.61	0.71	0.31	
4-31G	0.28	0.06	-0.76	-1.08	
6-31G**	0.74	0.56	-0.24	-0.48	

Exp. value: 22.3[a]

Formamide: monomer and dimer

α	F	G	H	I	J
STO-3G	10.48	10.61	10.98	11.24	23.27
4-31G	16.66	16.69	17.25	17.38	34.25
6-31G**	18.07	18.13	18.60	18.67	37.13
$\Delta\alpha$	$\alpha_J-2*\alpha_F$	$\alpha_J-2*\alpha_G$	$\alpha_J-2*\alpha_H$	$\alpha_J-2*\alpha_I$	
STO-3G	2.31	2.05	1.31	0.79	
4-31G	0.93	0.87	-0.25	-0.51	
6-31G**	0.99	0.87	-0.07	-0.21	

Exp. Value: 27.53[b]

Constituents of the formic acid and formamide pair

α	K	L	M	N	O	P	Q
STO-3G	9.41	11.23	8.81	9.15	10.58	11.05	21.20
4-31G	14.27	17.38	13.70	14.09	16.75	17.26	30.83
6-31G**	15.35	18.69	14.84	15.21	18.12	18.62	33.68
$\Delta\alpha$	$\alpha_Q-(\alpha_A+\alpha_F)$	$\alpha_Q-(\alpha_B+\alpha_G)$	$\alpha_Q-(\alpha_K+\alpha_L)$	$\alpha_Q-(\alpha_M+\alpha_N)$	$\alpha_Q-(\alpha_O+\alpha_P)$		
STO-3G	2.11	1.81	1.81	1.00	0.56		
4-31G	0.60	0.46	0.38	-0.52	-0.82		
6-31G**	0.86	0.71	0.72	-0.15	-0.36		

[a]Ref. 58.
[b]Ref. 59.

without superposition corrections. Therefore (in absence of pairs of strong electron-donating and electron-accepting groups at the end of the conjugated chains which could change the picture), we may anticipate that besides the advantage of partially blocking the conformational freedom in the carbon skeleton, the hydrogen bridges will probably neither impair nor enhance the overall polarizability of the conjugated chains we are considering as interesting candidates for electro-optical applications.

Concluding Remarks

The aim of this study was to analyze the influence of dimerization on the polarizability of three cyclic pairs involving formic acid and/or formamide. The central

question was to determine if the polarizability strongly deviates from additivity and in which direction (i.e., enhancement or decrease). Strictly speaking, the idea of comparing the polarizability of a dimer to the polarizability of the contributing monomers is questionable because of the structural and electronic changes taking place during the dimerization as well as difficulties in describing the partner systems at comparable levels of rigor. To minimize these influences we have examined different possibilities for the polarizability changes. The main results of this study are as follows:

1 Geometry relaxations in the dimers are very sensitive to the basis set extension and still significant when going from the 6-31G* to the 6-31G** basis. Direct comparison of the structural parameters of the isolated dimers with experimental data do not seem to be appropriate due to hydrogen bonds interlinking the dimers in the solid state.

2 Qualitative changes in polarizability upon dimerization can be accounted for on the basis of an additivity scheme. This is true for the various definitions used to estimate these changes.

On the basis of the present results, it is reasonable — in absence of pairs of strong donor and acceptor groups incorporated in the conjugated system — to anticipate that no significant influence (adverse and/or favorable) will result on the overall polarizability contributions of conjugated chains connected by cyclic hydrogen-bonded pairs.

Acknowledgments

Bibliography

[1] R. Pethig, Dielectric and Electronics Properties of Biological Materials. (Wiley, Chichester, 1969).
[2] R. Pethig, IEEE Trans. Electrical Insulation EL-19, 453 (1984).
[3] K. E. Rieckhoff and W. L. Peticolas, Science 147, 610 (1985).
[4] S. K. Kurtz and T. T. Perry, J. Appl. Phys. 39, 3798 (1968).
[5] D. Bauerle, K. Betzler, H. Hosseb, S. Kapphan, and P. Loose, Phys. Stat. Sol. A42, K119 (1977).
[6] G. G. Roberts, Adv. Phys. 34, 475 (1985).
[7] A. S. Paranjpe and V. K. Kelkar, Mol. Cryst. Liq. Cryst. 102, 289 (1984).
[8] Y. Pei, M. Verdaguer, O. Kahn, J. Sletten, and J. P. Renard, J. Am. Chem. Soc. 108, 7428 (1986).
[9] P. L. Davies, Trans. Faraday Soc. 48, 789 (1952).
[10] K. C. Rustagi and J. Ducuing, Opt. Commun. 10, 258 (1974).
[11] H. F. Hameka, J. Chem. Phys. 67, 2935 (1977).
[12] C. Flytzanis, in Nonlinear Optical Properties of Organic Molecules and Crystals, D. S. Chemla and J. Zyss, Eds. (Academic Press, New York, 1987) pp. 121-135.
[13] L. Leiserowitz, Acta Cryst. B32, 775 (1976).
[14] L. Leiserowitz, F. Nader, Acta Cryst. B33, 2719 (1977).

[15] E. R. Davidson and D. Feller, Chem. Rev. **86**, 681 (1986).

[16] J. S. Binkley, M. J. Frisch, D. J. Defrees, K. Raghavachari, R. A. Whiteside, H. B. Schlegel, E. M. Fluder, and J. A. Pople, Gaussian 82 Program, Carnegie-Mellon University (1981).

[17] W. J. Hehre, L. Random, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, (Wiley, New York, 1986).

[18] H. D. Cohen and C. J. Roothaan, J. Chem. Phys. **51**, 2657 (1969).

[19] J. Zyss and G. Berthier, J. Chem. Phys. **77**, 3635 (1982).

[20] P. Henrici, *Elements of Numerical Analysis*, (Wiley, New York, 1964).

[21] H. Werner, and W. Meyer, Mol. Phys. **31**, 355 (1976).

[22] A. J. Sadlej, Chem. Phys, Lett. **47**, 50 (1977).

[23] V. P. Bodart, J. Delhalle, J. M. André, and J. Zyss, Canad. J. Chem. **63**, 1631 (1985).

[24] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 558 (1970).

[25] A. Johansson, P. Kollman, and S. Rothenberg, Theor. Chim. Acta **29**, 167 (1973).

[26] W. Kolos, Theor. Chim. Acta **51**, 219 (1979).

[27] Z. Latajka and S. Scheiner, J. Chem. Phys. **81**, 407 (1984).

[28] D. W. Schwenke and D. G. Truhlar, J. Chem. Phys. **82**, 2418 (1985).

[29] M. Gutowski, J. H. Van Lenthe, J. Verbeek, F. B. Van Duijneveld, and G. Chalasinski, Chem. Phys. Lett. **124**, 370 (1986).

[30] S. K. Loushin, S. Liu, and C. E. Dykstra, J. Chem. Phys. **84**, 2720 (1986).

[31] P. G. Jasien and W. J. Stevens, J. Chem. Phys. **84**, 3271 (1986).

[32] F. J. Olivares del Valle, S. Tolosa, J. J. Esperilla, E. A. Ojalvo, and A. Requena, J. Chem. Phys. **84**, 5077 (1986).

[33] M. M. Szczesniak and S. Scheiner, J. Chem. Phys. **84**, 6328 (1986).

[34] R. A. Whiteside, M. J. Frisch, J. S. Binkley, D. J. Defrees, H. B. Schelgel, K. Raghavachari, and J. A. Pople, "Carnegie-Mellon Quantum Chemistry Archives," Department of Chemistry, Carnegie-Mellon University, 1981, pp. 118

[35] A. M. Sapse, L. M. Fugler, and D. Cowburn, Intern. J. Quantum Chem. **29**, 1241 (1986).

[36] J. E. Del Bene and W. L. Kochenour, J. Am. Chem. Soc. **98**, 2041 (1976).

[37] S. Scheiner and C. W. Kern, J. Am. Chem. Soc. **101**, 4081 (1979).

[38] Y. Sugawara, Y. Hamada, A. Y. Hirakawa, M. Tsuboi, S. Kato, and K. Morokuma, Chem. Phys. **50**, 105 (1980).

[39] M. J. Wojcik, A. Y. Hirakawa, M. Tsuboi, S. Kato, and K. Morokuma, Chem. Phys. Lett. **100**, 523 (1983).

[40] S. Hayashi, J. Umemura, S. Kato, and K. Morokuma, J. Phys. Chem. **88**, 1330 (1984).

[41] C. Mijoule, M. Allavena, J. M. Leclercq, and Y. Bouteiller, Chem. Phys. **109**, 207 (1986).

[42] A. Almenningen, O. Bastiansen, and T. Motzfeldt, Acta Chem. Scand. **23**, 2848 (1969).

[43] E. D. Stevens, Acta Cryst. **B34**, 544 (1978).

[44] R. W. Davis, A. G. Robiette, M. C. L. Gerry, E. Bjarnov, and G. Winnewisser, J. Mol. Spectr. **81**, 93–109 (1980).

[45] E. Hirota, R. Sugisaki, V. J. Nielsen, and G. O. Sorensen, J. Mol. Spectr. **49**, 251 (1974).

[46] J. Karle and L. O. Brockway, J. Amer. Chem. Soc. **66**, 574 (1944).

[47] I. Nahringbauer and G. Larsson, Arkiv. För Kemi **30**, 91 (1968).

[48] H. Ohtaki and S. Itoh, Z. Naturforsch **40a**, 1351 (1985).

[49] J. F. Hinton and R. D. Harpod, J. Am. Chem. Soc. **99**, 349 (1977).

[50] J. Sauer, P. Hobza, and R. Zahradnik, J. Phys. Chem. **84**, 3318 (1980).

[51] A. Chablo and A. Hinchliffe, Chem. Phys. Lett. **72**, 149 (1980).

[52] E. Younang, J. Delhalle, and J. M. André, New J. Chem., **11**, 403 (1987)

[53] G. Karlström and A. J. Sadlej, Theoret. Chim. Acta **61**, 1 (1982).

[54] C. E. Dykstra and S. Liu, J. Mol. Struct. **135**, 357 (1986).

[55] Landolt-Börnstein, *Zahlenwerte und Functionen aus Naturwissen Scaften und Technik*, K. M. Hellewege Ed. (Springer, Berlin, 1967), Neue Serie II, Vol 4, p. 142.

[56] J. L. Derissen and P. H. Smith, Mol. Cryst. **A34**, 842 (1978).

[57] M. Dreyfus, B. Maigret, and A. Pullman, Theor. Chim. Acta **17**, 109 (1970).

[58] Landolt-Börnstein, *Zahlenwerte und Functionen*, K. M. Hellewege, Ed. (Springer, Berlin, 1962), 8 Teil, Bd. II, p. 5.
[59] J. Applequist, J. R. Carl, and K. K. Fung, J. Am. Chem. Soc. **94**, 2952 (1972).

Molecular Similarity Based on Electrostatic Potential and Electric Field

EDWARD E. HODGKIN AND W. GRAHAM RICHARDS

Physical Chemistry Laboratory, South Parks Road, Oxford OX1 3QZ, United Kingdom

Abstract

A new definition of molecular similarity in terms of electron density is proposed and a method for calculating similarity based on molecular electrostatic potential and molecular electric field introduced. It is applied to some simple isosteres.

Introduction

Replacement of a functional group in a drug molecule gives rise to a change in the biological activity. Chemists use the concept of bioisosterism [1] to attempt to predict how successful a particular substitution might be.

This idea may be quantified theoretically by comparing the electron densities, ρ_A and ρ_B, of two molecules A and B, and calculating an index of similarity, R_{AB} [Eq. (1)], as first introduced by Carbo et al. [2].

$$R_{AB} = \frac{\int \rho_A \rho_B \, dv}{\left(\int \rho_A^2 \, dv\right)^{1/2} \left(\int \rho_B^2 \, dv\right)^{1/2}} \tag{1}$$

with the integrations being over all space.

The denominator is a normalizing constant and R_{AB} varies in the range 0 to 1. Such an index of similarity is required to have a value of 1 when the electron density distributions in the two molecules are identical. However, substitution of $\rho_A = n\rho_B$ into Eq. (1), where n is a constant gives an index of unity. Thus the Carbo index represents the similarity of the shapes of the density distributions but not of the magnitudes as well. This formula has been applied to electron density in a variety of ways [3–5].

We now propose an alternative definition of molecular similarity, H_{AB} [Eq. (2)].

$$H_{AB} = \frac{2 \int \rho_A \rho_B \, dv}{\int \rho_A^2 \, dv + \int \rho_B^2 \, dv} \tag{2}$$

H_{AB} can take values between 0 and 1. Substitution of $\rho_A = n\rho_B$ gives a value for H_{AB} of $2n/(1 + n^2)$. Thus the formula gives a total similarity of both shape and magnitude of the density distribution.

Similarity of Potential and Field

The calculation of molecular similarity need not however be restricted to electron density. Both styles of index can, in principle, be applied to any property of a molecule. Molecular electrostatic potential (MEP) and molecular electric field (MEF) are well suited to this approach although since these are signed quantities the values will run from -1 to $+1$. Molecular electrostatic potential is widely used in techniques of molecular design and structure-activity relations (QSAR) studies. The MEF is less frequently used, largely as a result of the difficulty of displaying a vector quantity in three dimensions, but it is important because the scalar product of the field and a dipole gives the energy of the dipole at a given point [6]. Dipolar interactions are important in ligand-macromolecule binding and in solvation.

The use of the H_{AB} index is particularly important for calculating the MEP and MEF similarity because these properties may be of similar shape for a pair of molecules, while their absolute values are critical importance. Both properties can be calculated using Mulliken population charges [7] from a suitable molecular orbital package such as MOPAC [8]. The integration is performed numerically over a three-dimensional grid. While MEP values are simply multiplied at each grid point, the product of MEF vectors is taken as the scalar product. The definition of the grid (i.e., extent and density) is important. It was found that a three-dimensional grid extending 10 Å around the molecules on all sides with grid points 1 Å apart was sufficient to approach very close to the limit of the similarity (within 1% of the similarity obtained using a very large, finely meshed grid) for both MEP and MEF. In addition, this grid definition allows a calculation to be performed in a reasonable amount of computer time. The value of the similarity index is virtually insensitive to movement of the grid.

Clearly, there is a need to exclude the volume contained within the molecules in some way so as to avoid singularities. It was decided that the molecular volumes should be defined by the Van der Waals surfaces and that the volume lying within both molecules should be excluded altogether from the calculation. This problem is tackled by allocating values of zero to the MEP or MEF at grid points inside the molecule. This is illustrated by Figure 1. The shaded area (AUB) is inside both
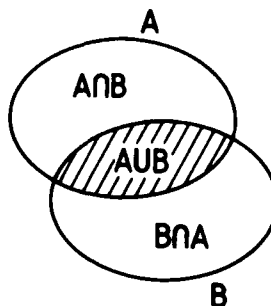

Figure 1.   Intersection of the Van der Waals surfaces of two superimposed molecules.

molecules and does not contribute to the similarity. The remaining areas (A ∩ B and B ∩ A) contribute only to the denominator of Eq. (2). The rationale behind excluding the Van der Waals volume to twofold. First, the point-charge approximation of MEP and MEF gives only realistic numbers outside the Van der Waals surface [9], and second, the volume within the molecule is of no relevance to interactions with the environment.

Superimposition

In addition to supplying information about similarity, the index can be used as a criterion for superimposing molecules in the optimum fashion so as to permit inferences about receptor topology. A variety of methods were tried for maximizing the MEP and MEF similarity. The most reliable of these proved to be a routine from the NAG mathematical library (E04CCF) employing the Simplex method [10]. Initially the two molecules may be superimposed in any relative orientation. One of the molecules is given complete freedom of motion in three dimensions and allowed to move toward the position at which the similarity is a maximum. Typically the Simplex method finds the maximum after 300 calculations of the index and consequently, is fairly costly in computer time. However, the method is robust in that a maximum is always reached which is independent of the starting point.

A series of bioisosteres, Me_2CH_2, Me_2O, Me_2S, has been the subject of some electron density similarity calculations [5, 11]. The MEP and MEF methods have been applied to this series. Atomic charges were computed using MOPAC. The results are shown in Tables I and II. The first set of values in each table shows the similarity of each pair of molecules when they are aligned along the same principal axis with the

TABLE I.   Results of MEP similarity calculations on propane, dimethyl ether, and dimethyl thioether.

	(i) Superimposition of central atoms		(ii) Position of maximum similarity	
	R_{AB}	H_{AB}	R_{AB}	H_{AB}
Me_2CH_2/Me_2O	0.67	0.022	0.70	0.023
Me_2CH_2/Me_2S	0.62	0.026	0.69	0.029
Me_2O/Me_2S	0.92	0.90	0.93	0.92

TABLE II.   Results of MEF similarity calculations on propane, dimethyl ether, and dimethyl thioether.

	(i) Superimposition of central atoms		(ii) Position of maximum similarity	
	R_{AB}	H_{AB}	R_{AB}	H_{AB}
Me_2CH_2/Me_2O	0.30	0.017	0.33	0.022
Me_2CH_2/Me_2S	0.21	0.016	0.32	0.027
Me_2O/Me_2S	0.81	0.78	0.85	0.83

central atoms superimposed. The striking feature in both tables of values is the difference between the Carbo and new formulae. In the comparisons involving propane the two indices differ greatly. This reflects the magnitudes of the atomic charges in propane which are very small compared with those in dimethyl ether and dimethyl thioether. For example the central carbon in propane has a charge of -0.05, the sulfur of dimethyl thioether -0.20 and the oxygen of dimethyl ether -0.34. However, the pattern of atomic charges in each molecule is similar and consequently their MEPs and MEFs are similarly shaped. Thus it is possible for the Carbo index to take an unrealistically high value. The Me_2O/Me_2S comparison gives good agreement between the Carbo and H_{AB} indices, reflecting the real similarity between their MEP and MEF distributions. The second set of values in each table shows the result of the similarity being maximized. The relative positions of the molecules after this maximization are *not shown. However, they do not move far from the position in which the central* atoms are overlapping (typically 0.2 Å). It is interesting to note that in all cases the molecules come to rest in different planes.
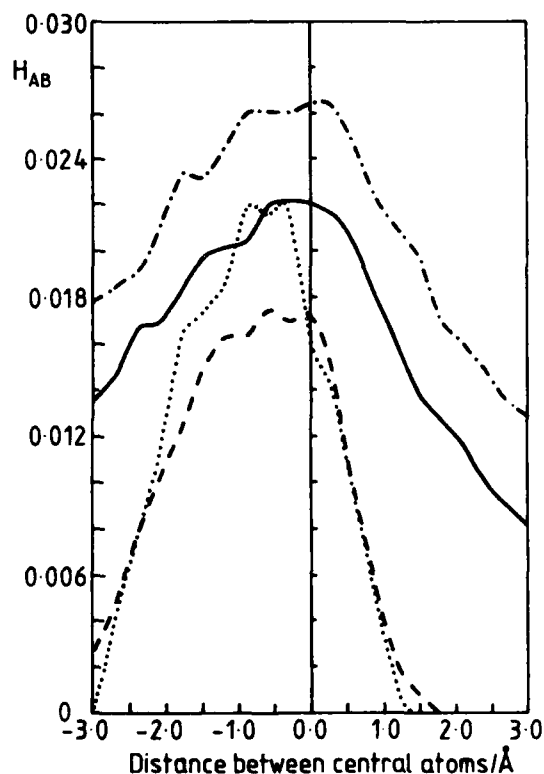
Figure 2.   Hodgkin similarity indices for the comparisons: Me_2CH_2/Me_2O (MEP: ———, MEF: — — — —) and Me_2CH_2/Me_2S (MEP: –·–·–·–, MEF: ....).

In addition, we observed the effect of moving the molecules relative to one another along the principal axis. The behavior of the MEP and MEF similarities was observed for propane/dimethyl ether, propane/dimethyl thioether (Fig. 2) and dimethyl ether/dimethyl thioether (Fig. 3) using the H_{AB} index. In general, the MEF similarity decreases more rapidly than the MEP similarity, as the molecuies are moved apart. This is expected because the electric field itself decays more rapidly than the electrostatic potential.

The use of MEP and MEF similarity calculations is very attractive in QSAR studies, given the importance that medicinal chemists attach to these properties. Previously [12] molecular shape has been shown to be a good descriptor in terms of correlations with activity. This work extends that idea so as to include electronic effects. It is hoped that the techniques described will be used to study bioactive molecules in order to relate similarity to biological activity or toxicity.
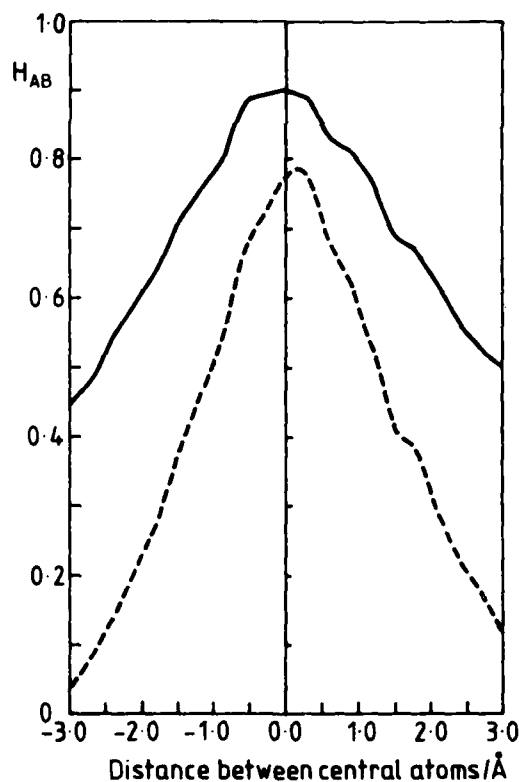
Figure 3. Hodgkin similarity indices for the comparison: Me_2O/Me_2S (MEP: ———, MEF: — — — —).

Acknowledgments

Bibliography

[1] C. W. Thornber, Chem. Soc. Rev. **8**, 563 (1979).
[2] R. Carbo, L. Leyda, and M. Arnau, Int. J. Quant. Chem. **17**, 1185 (1980).
[3] M. Martin and F. Sanz, Int. J. Quant. Chem. **23**, 1627, 1643 (1983).
[4] P. E. Bowen-Jenkins, D. L. Cooper, and W. G. Richards, J. Phys. Chem. **89**, 2195 (1985).
[5] E. E. Hodgkin and W. G. Richards, J. Chem. Soc., Chem. Commun. 1342 (1986).
[6] B. Pullman, J. Biomolec. Str. Dynam. **1**, 773 (1986).
[7] R. S. Mulliken, J. Chem. Phys. **23**, 1833 (1955).
[8] J. J. P. Stewart, QCPE Bull. **3**, 43 (1983).
[9] D. M. Hayes and P. A. Kollman, J. Am. Chem. Soc. **98**, 3335 (1976).
[10] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).
[11] P. E. Bowen-Jenkins and W. G. Richards, J. Chem. Soc., Chem. Commun. 133 (1986).
[12] A. J. Hopfinger, J. Am. Chem. Soc. **102**, 7196 (1980).

Library of Cumulative Atomic Multipole Moments. I. Nucleic Acid Bases

W. A. SOKALSKI*[†]

Institute of Organic and Physical Chemistry, I-4, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

P. C. HARIHARAN AND JOYCE J. KAUFMAN

Department of Chemistry
The Johns Hopkins University, Baltimore, Maryland 21218, USA

Abstract

Cumulative atomic multipole moments (CAMM) have been calculated for normal, rare, and protonated forms of adenine, thymine, guanine, cytosine, uracil, and 2-aminopurine from *ab initio* LCAO-MO-SCF wave function obtained from all-valence MODPOT basis set with *ab initio* effective core potentials. CAMM may be used in calculating electrostatic molecular potentials, electric fields, field gradients, etc. as well intermolecular interaction energies. Additionally, we derived analytic expressions for the point charge assemblages representing simultaneously all atomic and molecular moments. Convergence of atomic versus molecular multipole expansion has been illustrated in the Appendix.

Introduction

There is no general consensus as to how to define atomic charges in an unambiguous way. Fortunately, its arbitrariness and basis set dependence could be compensated by including higher terms in atomic multipole expansion, at least up to quadrupole moment [1]. Formally, all available multicenter multipole expansions [1–11] are equivalent, and practically, they can differ only in number of expansion centers, convergence, and complexity. The cumulative atomic multipole expansion [1] employed in this study is perhaps one of the most straightforward in practical implementation and could be regarded as natural supplement to Mulliken population analysis [12]. Its convergence has been illustrated in the Appendix. The electrostatic multipole term calculated directly from atomic multipole moments constitutes the most specific and orientation-dependent contribution to the intermolecular interaction energy for polar or ionic systems. This allowed successful prediction of the detailed structure of small hydrogen-bonded dimers [13–17] as well other van der Waals complexes involving aromatic molecules [18]. The remaining components of intermolecular interaction energy are more transferrable and could be represented by more or less universal atom–atom potentials derived from *ab initio* calculations [19–21]. Therefore, further

---

*To whom reprint requests should be addressed.

†Visiting scientist, The Johns Hopkins University.

applications of the above mentioned atom–atom potentials to biologically important systems (nucleic acids, proteins, enzymes, etc.) will require the knowledge of atomic multipoles for the corresponding building blocks, namely nucleic bases and aminoacids.

Despite the significant interest in studies on properties and interactions of nucleic acid bases, the only explicitly published set of atomic multipole moments has been obtained within the semiempirical IEHT approach [2]. Therefore, the *ab initio* atomic multipoles presented in this contribution constitute an indispensable supplement to nonempirical atom–atom potentials [20] enabling calculation of intermolecular interaction energy and its components as well studying other properties of DNA bases available from multipole expansion. Besides normal nitrogen DNA bases our calculations have been extended to include the corresponding protonated and rare forms as well as 2-aminopurine analogs. Their interactions within DNA may be responsible for the observed sequence-specific mutation "hot spots" and will be the subject of our forthcoming study.

In more simplified models, the anisotropy of local charge distribution is frequently accounted for by locating additional point charges outside atomic nuclei. However, previously introduced models of this kind [22–31] have been obtained in an arbitrary and unsystematic manner, either by fitting point charge values and locations to approximate molecular multipole moments [23–25] or molecular electrostatic potentials [26–31]. So the above mentioned models [22–31] depend on the arbitrary choice of number, location, and value of point charges. While reproducing some global properties it does not always correctly represent the anisotropy of local charge distributions. Sometimes such a procedure even leads to unphysically large point charges [28]. With the aid of CAMM we are in a position to construct analytically point charge models of desirable precision in a much more systematic manner. In contrast to methods based on global criteria [22–31], our model allows the preservation of the local anisotropies of charge distribution represented by atomic multipoles and all corresponding molecular moments.

Cumulative Atomic Multipole Moments

Theoretical molecular multipole moments are determined within the LCAO MO SCF approach as expectation values of the operator $u^k v^l w^m$ ($u, v, w = x, y, z$)

$$\langle u^k v^l w^m \rangle = \sum_i^{\text{atoms}} Z_i u_i^k v_i^l w_i^m - \sum_{l \in i}^{\text{AO}} \sum_J^{\text{AO}} P_{IJ} \langle I \mid u^k v^l w^m \mid J \rangle$$

$$= \sum_i^{\text{atoms}} \langle u^k v^l w^m \rangle_i \tag{1}$$

where Z_i denotes nuclear charge, $\langle I \mid u^k v^l w^m \mid J \rangle$ one electron multipole moment integral and P_{IJ} density matrix element. Transforming each atomic multipole moment $\langle u^k v^l w^m \rangle_i$ to a local coordinate system with origin at the i-th atomic center (u_i, v_i, w_i) we obtain cumulative atomic multipole moments (CAMM) M_i^{klm}

$$M_i^{klm} = \langle u^k v^l w^m \rangle - \sum_{k'>0}^{k} \sum_{l'>0}^{l} \sum_{m'>0}^{m} \binom{k}{k'}\binom{l}{l'}\binom{m}{m'} u^{k-k'} v^{l-l'} w^{m-m'} M_i^{k'l'm'} \tag{2}$$

$$klm \neq k'l'm'$$

In this formulation M_i^{000} is equivalent to the Mulliken atomic charge [12]. Each of the higher multipoles corresponds to contributions not included in the lower moments. Cumulative character of the consecutive moments of $(k + l + m)$th order allows the stepwise refinement of local charge anisotropy and multipole expansion. Cumulative second-order moments M_i^{klm} $(k + l + m = 2)$ can be additionally transformed into traceless quadrupole tensors $\mathbf{Q}_i^{uv}$

$$\mathbf{Q}_i^{uv} = 0.5(3M_i^{uv0} - \delta_{uv} \sum_t^{x,y,z} M_t^{tt0}) \qquad (u, v, w = x, y, z) \tag{3}$$

The molecular wavefunctions of nucleic acid bases have been calculated within *ab initio* SCF-LCAO-MO method using the minimal all-valence MODPOT basis set $(3s3p/3s) \rightarrow [1s1p/1s]$ with the *ab initio* effective core model potentials [32]. This well balanced basis set yielded for several hydrogen-bonded complexes intermolecular interaction energies very close to results obtained in extended 6-31G* and 6-31 G** basis sets [33]. It must be noted that MODPOT results available at a fraction of the cost compare favorably with those obtained in more expensive basis sets [33, 34] which underestimate molecular dipole moments and intermolecular interactions. The sample cartesian coordinates (X, Y, Z), atomic monopoles (Q), atomic dipoles (DX, DY, DZ), and traceless atomic quadrupoles $(QXX, QYY, QZZ, QXY, QXY, QYZ)$ for 2-aminopurine have been presented in the Table I. Analogous data for normal, rare, and protonated forms of adenine, guanine, uracil, cytosine, and thymine (Figs. 1 and 2) are available on request [either in printed form or on one DSDD 5.25" floppy diskette mailed to one of authors (WAS)].

Analytic Point Charge Representation of Atomic and Molecular Multipoles

Each multipole can be represented by a minimal set of n properly arranged point charges q_p

$$M_i^{klm} = \sum_{p=1}^{n} q_p(u_p - u_i)^k (v_p - v_i)^l (w_p - w_i)^m \tag{4}$$

There is no *a priori* way of determining the values (q_p) and locations (u_p, v_p, w_p) of the point charges without additional constraints. In this study we introduce only one quite natural assumption: that the charge located at nucleus (u_i, v_i, w_i) is equal to the corresponding atomic core charge. In the case of MODPOT basis set, the core charge equals the number of valence electrons of corresponding atom. To preserve this condition all other off-nuclear charges have to be located on a sphere of radius R_i

$$R_i = \sqrt{-Tr(T_iM_iT_i')/(Z_i - M_i^{000})} \tag{5}$$

where $Tr(T_iM_iT_i')$ stands for the trace of the (diagonal) atomic second-moment tensor M_i^{tt} transformed by rotation matrix T_i into the corresponding principal axes.

Table II gives analytic expressions for point charge values q and their locations (x_p, y_p, z_p) for all atomic and molecular multipole moments up to monopole $(k + l + m = n = 0)$, dipole $(n = 1)$, or quadrupole $(n = 2)$. Trivially, the first expansion $(n = 0)$ is equivalent to the Mulliken population analysis. The second

TABLE I. Atomic coordinates and cumulative multipoles (in au units) 2-aminopurine (amino form) [47] (3-21G opt).

ATOM	X	Y	Z	Q	DX	DY	DZ	QXX	QYY	QZZ	QXY	QXZ	QYZ
N(1)	.0000	.0000	.0000	-.3588	.3760	.0000	.2691	-.0878	-.2001	.2879	.0000	-.3792	.0000
C(2)	.0000	.0000	2.5511	.3970	.0702	.0000	.0069	.7200	-1.5014	.3315	.0000	-.4839	.0000
N(3)	2.0678	.0000	3.9963	-.3936	.0172	.0000	-.4589	.4705	-.2141	-.2563	.0000	-.0129	.0000
C(4)	4.2130	.0000	2.7307	.2505	.0893	.0000	.0594	1.4084	-1.2311	-.1773	.0000	-.1250	.0000
C(5)	4.4221	.0000	.1029	-.0174	.2386	.0000	.1143	1.3284	-1.4914	.1630	.0000	.3543	.0000
C(6)	2.1802	.0000	-1.2219	-.0708	-.1165	.0000	.6650	.5153	-.0704	-.4450	.0000	-.2262	.0000
N(7)	6.9840	.0000	-.5798	-.3209	-.2023	.0000	.4337	.1453	-.0727	-.0726	.0000	.0585	.0000
C(8)	8.1987	.0000	1.4857	.0389	-.8988	.0000	.0178	.0714	-.5405	.4690	.0000	-.1515	.0000
N(9)	6.6122	.0000	3.5910	-.4973	-.1156	.0000	-.3792	.3539	-.5375	.1836	.0000	-.0005	.0000
H(6)	2.1290	.0000	-3.2433	.2736	.0010	.0000	-.0971	.0331	-.1022	.0690	.0000	.0029	.0000
H(9)	7.0798	.0000	5.4123	.3777	.0068	.0000	.0823	.0046	-.0896	.0851	.0000	.0441	.0000
H(8)	10.1982	.0000	1.6786	.2714	.1206	.0000	.0108	.0472	-.0778	.0307	.0000	.0046	.0000
N(2)	-2.2839	.0000	3.6621	-.6495	.2865	.0000	-.2172	.3124	-.3137	.0013	.0000	-.4020	.0000
H(N21)	-2.1166	.0000	5.5336	.3426	.0602	.0000	.0476	-.0429	-.1228	.1657	.0000	-.0580	.0000
H(N22)	-3.9737	.0000	2.8405	.3566	-.0791	.0000	-.0383	.0900	-.0883	-.0017	.0000	.0621	.0000
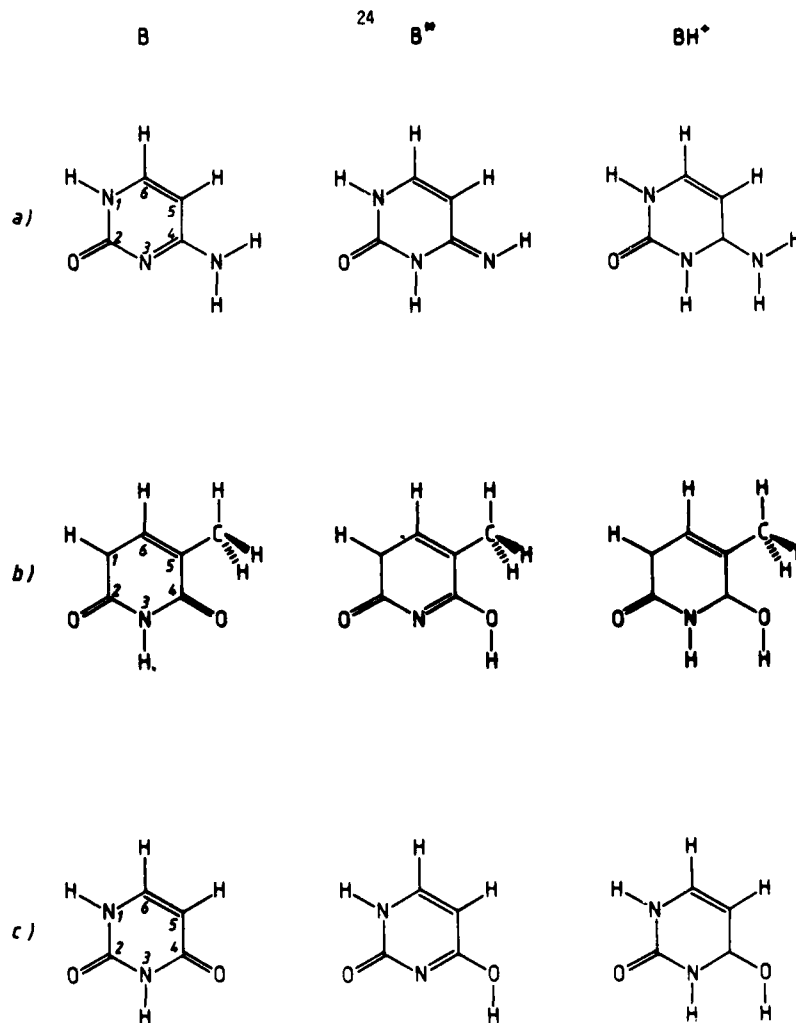
Figure 1. Pyrimidine nucleic acid bases (B) and their protonated (BH^+) and rare tau-
tomeric forms (B*). (a) cytosine; (b) thymine; (c) uracil.

model (n = 1) reproducing atomic and molecular moments has been utilized in our
earlier calculations [35, 36]. Equivalent, but apparently more perplexed point charge
models preserving molecular dipole moments have been proposed elsewhere [24, 37].
However, our later calculations including also quadrupole and octopole moments
[1, 33, 34] indicate that atomic quadrupole moments play an essential role in
the description of intermolecular interactions and other properties. Therefore, we
recommend here the use of at least quadrupole model (n = 2). The importance of
quadrupole contribution can be judged directly by comparison of magnitudes of point
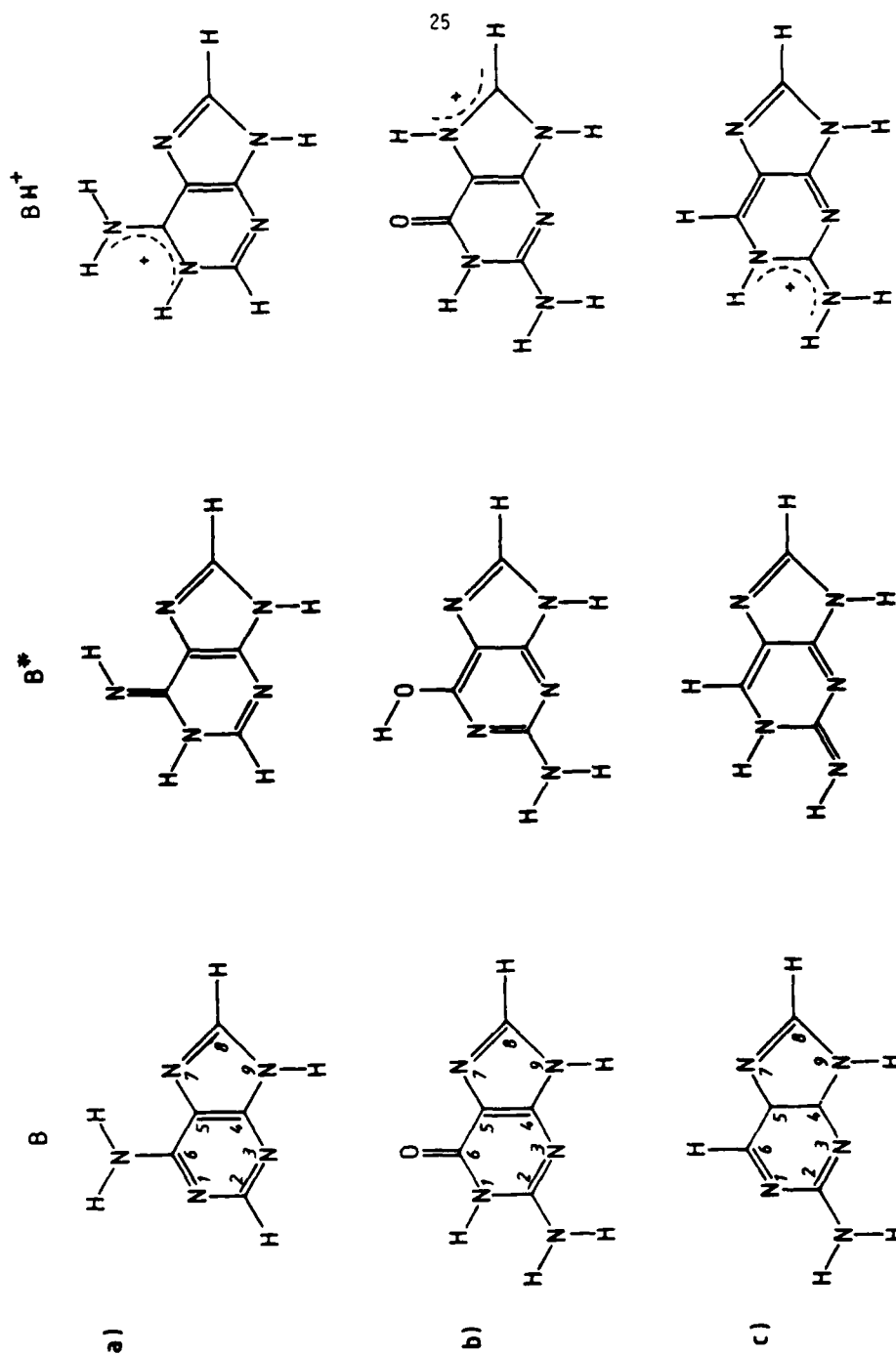charge values derived from atomic dipoles (p = 2, 3) and quadrupoles (p =

Figure 2. Purine nucleic acid bases (B) and their protonated (BH⁺) and rare tautomeric forms (B*). (a) adenine; (b) guanine; (c) 2-aminopurine.

TABLE II. Point charge values q and their locations u_p ($u_p = x, y, z$) equivalent to atomic and molecular multipole expansions truncated at monopole ($n = 0$), dipole ($n = 1$), and quadrupole ($n = 2$) level.

Expansion level n	Point charge P	Point charge value q_p	Point charge location u_p
0	1	M_i^{000}	u_i
1	1	M_i^{000}	u_i
	2	$\|\mu_i\|/(2R_i)$	$u_i - M_i^{u00}R_i/\|\mu_i\|$
	3	$-\|\mu_i\|/(2R_i)$	$u_i + M_i^{u00}R_i/\|\mu_i\|$
2	1	Z	u_i
	2	$\|\mu_i\|/(2R_i)$	$u_i - M_i^{u00}R_i/\|\mu_i\|$
	3	$-\|\mu_i\|/(2R_i)$	$u_i + M_i^{u00}R_i/\|\mu_i\|$
	4	$M_i^{11}/(2R_i^2)$	$u_i - T_i^{u1}R_i$
	5	$M_i^{11}/(2R_i^2)$	$u_i + T_i^{u1}R_i$
	6	$M_i^{22}/(2R_i^2)$	$u_i - T_i^{u2}R_i$
	7	$M_i^{22}/(2R_i^2)$	$u_i + T_i^{u2}R_i$
	8	$M_i^{33}/(2R_i^2)$	$u_i - T_i^{u3}R_i$
	9	$M_i^{33}/(2R_i^2)$	$u_i + T_i^{u3}R_i$

$$\mu_i = \sqrt{\sum^{xyz}(M_i^{u00})^2} \qquad M_i^n = T_i M_i^{uv} T_i^+$$

$4, 5, 6, 7, 8, 9$). Typically, the latter are about an order of magnitude larger than the former. The corresponding point charge representation of molecular and atomic multipoles for 2-aminopurine is presented in Table III at quadrupole level ($n = 2$). For each atom the corresponding cartesian coordinates and values of 9 point charges have been given there. NUCLEI represents the integer charge of atomic core ($p = 1$), DIPOLE: two-point charges ($p = 2, 3$) representing atomic dipole and QUADRU-POLE: six-point charges ($p = 4, 5, 6, 7, 8, 9$) representing atomic quadrupole moment. The lower level representations ($n = 0, 1$) may be easily obtained from the data presented in this paper.

Although such point charge models are inferior in respect to the corresponding atomic multipole models (due to overlapping atomic spheres), they yield valuable conceptual information. They allow one to illustrate directly the relationship between multipoles and various chemical concepts of bonding, arrangement of lone electron pairs, etc. In addition, it is evident that models with smaller number of point charges may be deficient. The point charge representations for the remaining molecules are available on request [either in printed form or on one DSDD 5.25" floppy diskette mailed to one of the authors (WAS)].

SOKALSKI ET AL.

TABLE III. Point charge representation 2-aminopurine (amino form).

Atom	X(au)	Y(au)	Z(au)	Charge(au)	
N(1)	.000000	.000000	.000000	5.000000	NUCLEI
N(1)	−1.238989	.000000	−.886874	−.151723	DIPOLE
N(1)	1.238989	.000000	.886874	.151723	DIPOLE
N(1)	1.294647	.000000	.803447	−.939526	QUADRUPOLE
N(1)	−1.294647	.000000	−.803447	−.939526	QUADRUPOLE
N(1)	.000000	−1.523693	.000000	−.921877	QUADRUPOLE
N(1)	.000000	1.523693	.000000	−.921877	QUADRUPOLE
N(1)	−.803447	.000000	1.294647	−.818018	QUADRUPOLE
N(1)	.803447	.000000	−1.294647	−.818018	QUADRUPOLE
C(2)	.000000	.000000	2.551131	4.000000	NUCLEI
C(2)	−1.630809	.000000	2.391209	−.021537	DIPOLE
C(2)	1.630809	.000000	2.711052	.021537	DIPOLE
C(2)	.000000	−1.638631	2.551131	−.731023	QUADRUPOLE
C(2)	.000000	1.638631	2.551131	−.731023	QUADRUPOLE
C(2)	−.917801	.000000	1.193650	−.599961	QUADRUPOLE
C(2)	.917801	.000000	3.908611	−.599961	QUADRUPOLE
C(2)	−1.357481	.000000	3.468931	−.470505	QUADRUPOLE
C(2)	1.357481	.000000	1.633330	−.470505	QUADRUPOLE
N(3)	2.067806	.000000	3.996336	5.000000	NUCLEI
N(3)	2.010491	.000000	5.527925	−.149810	DIPOLE
N(3)	2.125121	.000000	2.464747	.149810	DIPOLE
N(3)	2.040694	.000000	2.463915	−.935336	QUADRUPOLE
N(3)	2.094918	.000000	5.528758	−.935336	QUADRUPOLE
N(3)	2.067806	−1.532661	3.996336	−.929317	QUADRUPOLE
N(3)	2.067806	1.532661	3.996336	−.929317	QUADRUPOLE
N(3)	.535385	.000000	4.023448	−.832141	QUADRUPOLE
N(3)	3.600228	.000000	3.969224	−.832141	QUADRUPOLE
C(4)	4.212948	.000000	2.730733	4.000000	NUCLEI
C(4)	2.846673	.000000	1.822821	−.032695	DIPOLE
C(4)	5.579223	.000000	3.638646	.032695	DIPOLE
C(4)	4.212948	−1.640431	2.730733	−.777406	QUADRUPOLE
C(4)	4.212948	1.640431	2.730733	−.777406	QUADRUPOLE
C(4)	4.084835	.000000	1.095313	−.648090	QUADRUPOLE
C(4)	4.341061	.000000	4.366154	−.648090	QUADRUPOLE
C(4)	2.577527	.000000	2.858846	−.449250	QUADRUPOLE
C(4)	5.848368	.000000	2.602620	−.449250	QUADRUPOLE
C(5)	4.422072	.000000	.102873	4.000000	NUCLEI
C(5)	2.895037	.000000	−.628821	−.078135	DIPOLE
C(5)	5.949108	.000000	.834567	.078135	DIPOLE
C(5)	4.422072	1.693285	.102873	−.842947	QUADRUPOLE
C(5)	4.422072	−1.693285	.102873	−.842947	QUADRUPOLE
C(5)	3.965275	.000000	1.733379	−.662146	QUADRUPOLE
C(5)	4.878870	.000000	−1.527633	−.662146	QUADRUPOLE
C(5)	2.791566	.000000	−.353924	−.503586	QUADRUPOLE
C(5)	6.052578	.000000	.559670	−.503586	QUADRUPOLE
C(6)	2.180201	.000000	−1.221910	4.000000	NUCLEI
C(6)	2.492399	.000000	−3.004560	−.186526	DIPOLE
C(6)	1.868002	.000000	.560740	.186526	DIPOLE

TABLE III. (*Continued.*)

Atom	X(au)	Y(au)	Z(au)	Charge(au)	
C(6)	1.784971	.000000	−2.988008	−.728909	QUADRUPOLE
C(6)	2.575430	.000000	.544188	−.728909	QUADRUPOLE
C(6)	2.180201	−1.809781	−1.221910	−.685634	QUADRUPOLE
C(6)	2.180201	1.809781	−1.221910	−.685634	QUADRUPOLE
C(6)	.414103	.000000	−.826681	−.620877	QUADRUPOLE
C(6)	3.946299	.000000	−1.617139	−.620877	QUADRUPOLE
N(7)	6.983974	.000000	−.579754	5.000000	NUCLEI
N(7)	7.632219	.000000	−1.969702	−.156017	DIPOLE
N(7)	6.335729	.000000	.810194	.156017	DIPOLE
N(7)	6.609711	.000000	.907561	−.899200	QUADRUPOLE
N(7)	7.358237	.000000	−2.067069	−.899200	QUADRUPOLE
N(7)	6.983974	1.533681	−.579754	−.897116	QUADRUPOLE
N(7)	6.983974	−1.533681	−.579754	−.897116	QUADRUPOLE
N(7)	5.496659	.000000	−.954017	−.864146	QUADRUPOLE
N(7)	8.471289	.000000	−.205491	−.864146	QUADRUPOLE
C(8)	8.198682	.000000	1.485707	4.000000	NUCLEI
C(8)	9.949063	.000000	1.450991	−.256749	DIPOLE
C(8)	6.448301	.000000	1.520423	.256749	DIPOLE
C(8)	8.198682	−1.750725	1.485707	−.718968	QUADRUPOLE
C(8)	8.198682	1.750725	1.485707	−.718968	QUADRUPOLE
C(8)	9.857441	.000000	2.045668	−.657986	QUADRUPOLE
C(8)	6.539922	.000000	.925747	−.657986	QUADRUPOLE
C(8)	7.638721	.000000	3.144467	−.603617	QUADRUPOLE
C(8)	8.758642	.000000	−.173052	−.603617	QUADRUPOLE
N(9)	6.612196	.000000	3.591045	5.000000	NUCLEI
N(9)	7.061668	.000000	5.065895	−.128566	DIPOLE
N(9)	6.162724	.000000	2.116195	.128566	DIPOLE
N(9)	6.612196	−1.541819	3.591045	−.991587	QUADRUPOLE
N(9)	6.612196	1.541819	3.591045	−.991587	QUADRUPOLE
N(9)	6.607928	.000000	2.049231	−.890461	QUADRUPOLE
N(9)	6.616465	.000000	5.132858	−.890461	QUADRUPOLE
N(9)	5.070383	.000000	3.595313	−.866586	QUADRUPOLE
N(9)	8.154010	.000000	3.586776	−.866586	QUADRUPOLE
H(6)	2.129035	.000000	−3.243270	1.000000	NUCLEI
H(6)	2.071250	.000000	−1.853238	−.034938	DIPOLE
H(6)	2.186819	.000000	−4.633301	.034938	DIPOLE
H(6)	2.129035	1.391232	−3.243270	−.138669	QUADRUPOLE
H(6)	2.129035	−1.391232	−3.243270	−.138669	QUADRUPOLE
H(6)	3.515785	.000000	−3.354842	−.115404	QUADRUPOLE
H(6)	.742284	.000000	−3.131697	−.115404	QUADRUPOLE
H(6)	2.017462	.000000	−4.630021	−.109141	QUADRUPOLE
H(6)	2.240607	.000000	−1.856519	−.109141	QUADRUPOLE
H(9)	7.079802	.000000	5.412250	1.000000	NUCLEI
H(9)	6.974798	.000000	4.131721	−.032143	DIPOLE
H(9)	7.184806	.000000	6.692779	.032143	DIPOLE
H(9)	7.079802	1.284826	5.412250	−.121817	QUADRUPOLE
H(9)	7.079802	−1.284826	5.412250	−.121817	QUADRUPOLE
H(9)	8.255218	.000000	4.893426	−.106723	QUADRUPOLE

TABLE III. (Continued.)

Atom	X(au)	Y(au)	Z(au)	Charge(au)	
H(9)	5.904387	.000000	5.931074	-.106723	QUADRUPOLE
H(9)	6.560978	.000000	4.236835	-.082604	QUADRUPOLE
H(9)	7.598626	.000000	6.587665	-.082604	QUADRUPOLE
H(8)	10.198180	.000000	1.678589	1.000000	NUCLEI
H(8)	8.773540	.000000	1.550927	-.042333	DIPOLE
H(8)	11.622820	.000000	1.806251	.042333	DIPOLE
H(8)	10.198180	1.430347	1.678589	-.134121	QUADRUPOLE
H(8)	10.198180	-1.430347	1.678589	-.134121	QUADRUPOLE
H(8)	9.835962	.000000	3.062313	-.116637	QUADRUPOLE
H(8)	10.560400	.000000	.294865	-.116637	QUADRUPOLE
H(8)	8.814455	.000000	1.316372	-.113558	QUADRUPOLE
H(8)	11.581900	.000000	2.040806	-.113558	QUADRUPOLE
N(2)	-2.283915	.000000	3.662111	5.000000	NUCLEI
N(2)	-3.542500	.000000	4.616081	-.113837	DIPOLE
N(2)	-1.025329	.000000	2.708142	.113837	DIPOLE
N(2)	-2.283915	-1.579271	3.662111	-.983501	QUADRUPOLE
N(2)	-2.283915	1.579271	3.662111	-.983501	QUADRUPOLE
N(2)	-3.176697	.000000	2.359408	-.978224	QUADRUPOLE
N(2)	-1.391132	.000000	4.964814	-.978224	QUADRUPOLE
N(2)	-3.586618	.000000	4.554893	-.862997	QUADRUPOLE
N(2)	-.981212	.000000	2.769328	-.862997	QUADRUPOLE
H(N21)	-2.116560	.000000	5.533598	1.000000	NUCLEI
H(N21)	-3.091908	.000000	4.762494	-.030835	DIPOLE
H(N21)	-1.141212	.000000	6.304702	.030835	DIPOLE
H(N21)	-2.116560	-1.243345	5.533598	-.136044	QUADRUPOLE
H(N21)	-2.116560	1.243345	5.533598	-.136044	QUADRUPOLE
H(N21)	-.913016	.000000	5.845670	-.122055	QUADRUPOLE
H(N21)	-3.320103	.000000	5.221526	-.122055	QUADRUPOLE
H(N21)	-2.428632	.000000	6.737142	-.070596	QUADRUPOLE
H(N21)	-1.804488	.000000	4.330054	-.070596	QUADRUPOLE
H(N22)	-3.973713	.000000	2.840495	1.000000	NUCLEI
H(N22)	-2.803970	.000000	3.407069	-.033802	DIPOLE
H(N22)	-5.143457	.000000	2.273922	.033802	DIPOLE
H(N22)	-3.973713	1.299733	2.840495	-.124659	QUADRUPOLE
H(N22)	-3.973713	-1.299733	2.840495	-.124659	QUADRUPOLE
H(N22)	-4.559113	.000000	4.000932	-.113747	QUADRUPOLE
H(N22)	-3.388314	.000000	1.680059	-.113747	QUADRUPOLE
H(N22)	-5.134150	.000000	2.255096	-.083287	QUADRUPOLE
H(N22)	-2.813277	.000000	3.425894	-.083287	QUADRUPOLE

Source: From Ref. 47.

Electrostatic Interaction Energy in DNA Complementary Bases

Cumulative atomic multipoles presented in this study can be used to evaluate the multipole component of the electrostatic interaction energy. Such results are presented in Table IV for hydrogen-bonded adenine-thymine and guanine-thymine base pairs at the corresponding experimentally determined geometries [38] and compared with other nonempirical estimates [31, 39].

TABLE IV.   Nonempirical estimates of multipole electrostatic interaction energy in DNA complementary base pairs (in [kcal/mol]). Total interaction energies in parentheses.

Base pair	Basis set	minimal [40]	STO-3G[41]	MODPOT [32]	
Base pair		(7s3p/3s)	(6s3p/3s)	(3s3p/3s)	exp
Adenine-thymine		$-13.9(-12.9)^a$	-8.96^b	$-17.0(-13.8)^c$	$(-13.0)^d$
Guanine-cytosine		$-25.4(-23.7)^a$	-21.53^b	$-33.5(-23.2)^c$	$(-21.0)^d$

[a] Ref. 39, up to quadrupole–quadrupole term.

[b] Ref. 31 (approximate point charge model fitted to reproduce electrostatic molecular potentials).

[c] This work, up to monopole–quadrupole term. Total interaction energy calculated with the use of nonempirical atom-atom potentials described in Ref. 20.

[d] Experimental values Ref. 42.

Electrostatic multipole energies predicted with MODPOT basis set are much larger than values obtained in minimal $(7, 3/3) \rightarrow [2s1p/1s]$ [40] and STO-3G [41] sets. This is consistent with our results obtained for 12 smaller hydrogen-bonded dimers [33]. MODPOT interaction energies obtained in above mentioned study [33] have been very close to the reference 6-316* and 6-316** values in contrast to the severely underestimated STO-3G results. Also the MODPOT basis gives results very close to the all-electron calculations using the same valence basis set as the MODPOT (and the same inner shell basis set from which the MODPOT parameters were determined) [33, 34]. Our total interaction energies match closely those predicted within more complex nonempirical potentials [39] and recently available experimental values [42].

Summary

Molecular charge distributions obtained in LCAO-MO-SCF calculations can be analytically decomposed into a set of cumulative atomic multipole moments. With the aid of CAMM one may estimate electrostatic molecular potentials, electric fields, and electrostatic interaction energies with much better accuracy than within one-center molecular expansion (see Appendix). Expansions up to at least atomic quadrupoles are preferable.

Furthermore, atomic multipoles may be analytically decomposed into a set of off-nuclear point charges reproducing simultaneously all atomic and molecular moments. In contrast to other available schemes of this kind our approach allows systematic refinement by including higher multipole moments. Recently our approach has been extended to include correlation effects and the lowest molecular multipole moments match closely the available experimental results for small molecules [43].

The results presented in this contribution enable inexpensive nonempirical studies of electrostatic interactions between nucleic acid components. The remaining interaction energy contributions may be estimated from existing nonempirical potential functions [20]. A recent ab initio study of Aida and Nagata [44] indicates that the electrostatic energy has an almost linear relation to the total stacking energy of A-DNA and B-DNA compounds. This should allow estimation of the relative stacking energies from atomic multipole moments, in accordance with Langlet et al. [39].

$$(E^{(J)}_{EL,appr} - E^{(1)}_{EL,exact}) / E^{(1)}_{EL,exact} \cdot 100\%$$

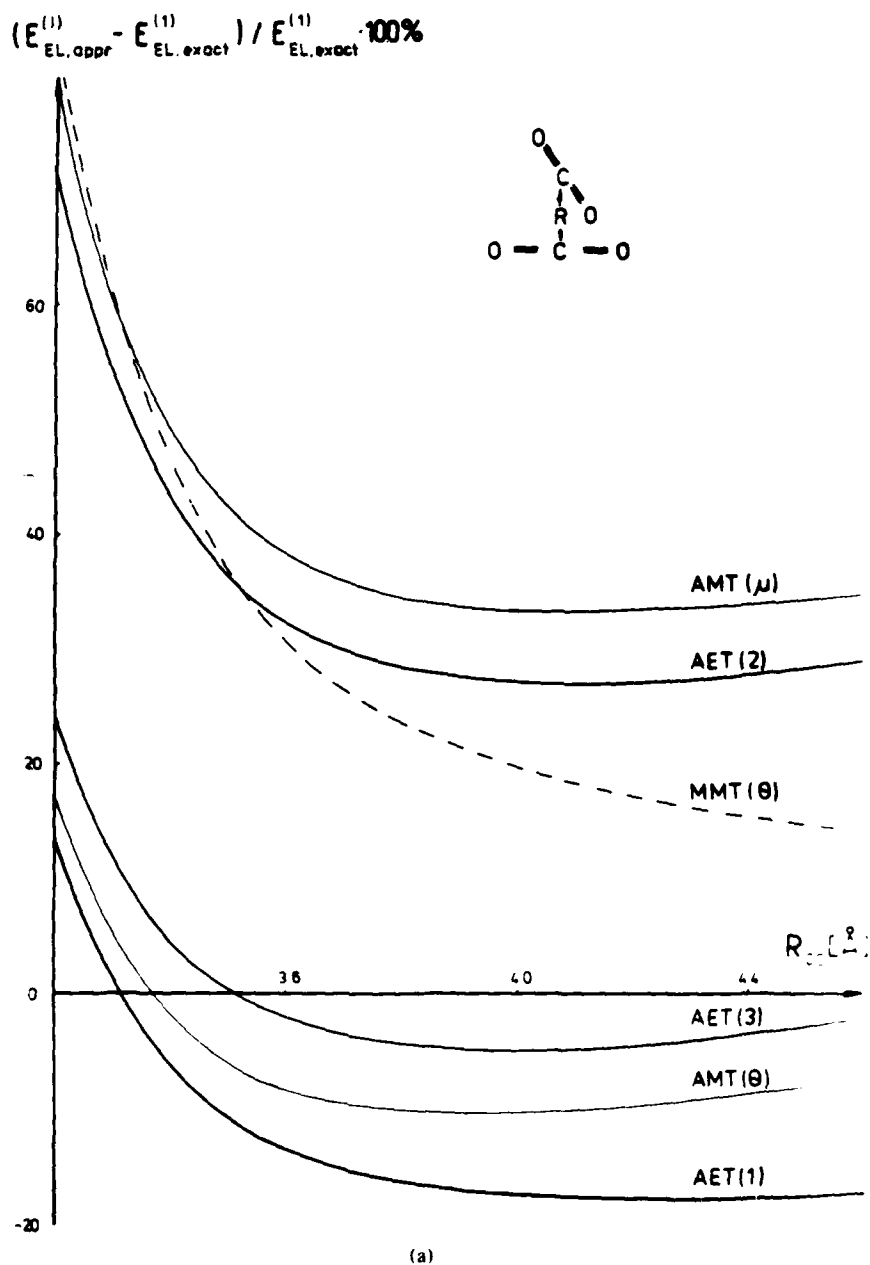AMT (μ)

AET (2)

MMT (θ)

AET (3)

AMT (θ)

AET (1)

(a)

Figure 3. The relative errors of various multipole or estimates of the electrostatic interaction energy calculated in the minimal all-valence MODPOT basis set [32] as a function of intermolecular distance R. (a) $(CO_2)_2$ dimer; (b) hydrogen-bonded HOH...FH dimer. AMT–Atomic Multipole Truncated expansion; AET–Atomic Exponent Truncated expansion; MMT–Molecular Multipole Truncated expansion; MET–Molecular Exponent Truncated expansion.

(b)
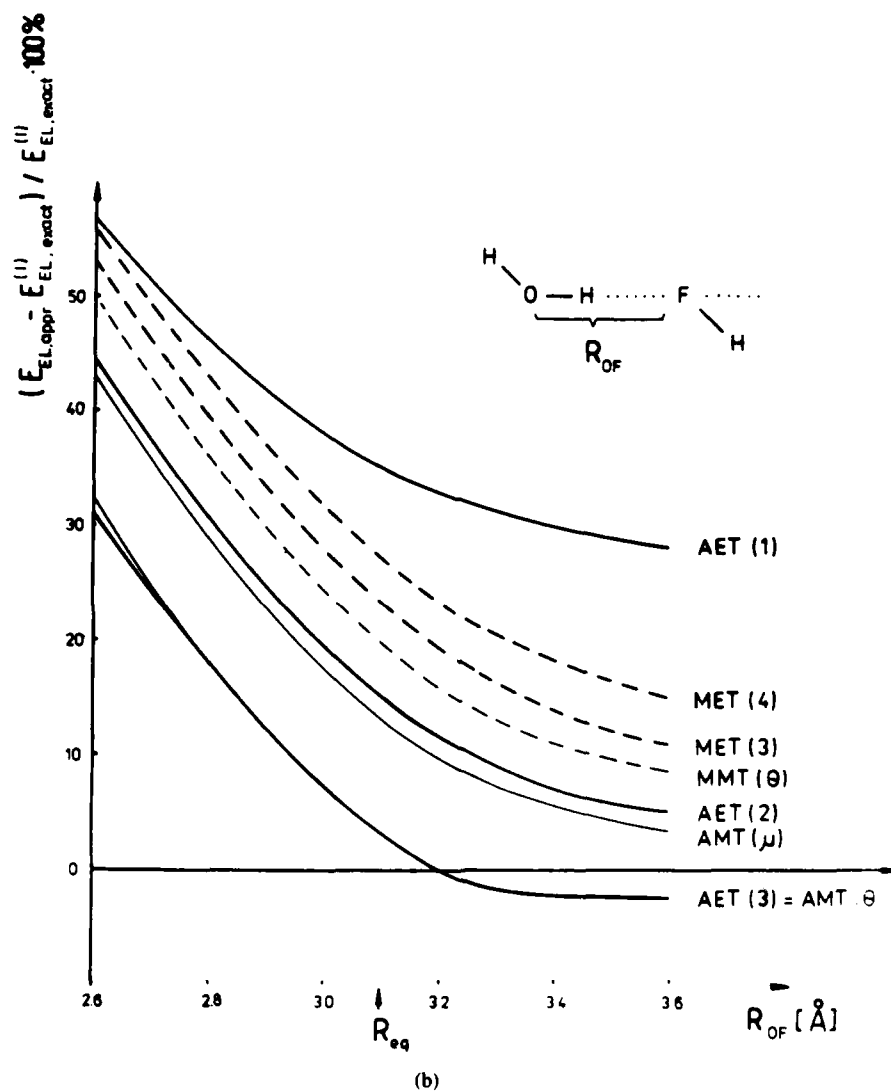
Appendix

Convergence of Atomic Versus Molecular Multipole Expansion

The convergence of multipole expansion could be illustrated by the relative error $(X_{appr} - X_{exact})/X_{exact} \cdot 100\%$ of property X_{appr} estimated from multipole series. As the sample properties X, we use here electrostatic molecular potentials V, electric fields E calculated for CO_2 molecule in minimal all electron ($6s3p$) basis set [32], and electro-

static interaction energies $E_{EL}^{(1)}$ for $(CO_2)_2$ and $HF \cdot HOH$ dimers obtained in an SCF decomposition scheme [33] in minimal all-valence MODPOT $(3s3p) \rightarrow |1s1p/1s|$ basis set [32]. Traditionally, the multipole series is truncated at the term containing highest multipole moment $m = q, \mu, \theta, \Omega$ |moment truncated, MT(m)| and it includes all terms involving lower moments. However, in the case of electrostatic interaction energy evaluated from superposition of two multipole series the results depend strongly on the way the entire expansion is truncated. In contrast to multipole truncated (MT) expansion, much better results can be obtained |45, 46| when it is terminated at terms having the same R dependence |exponent truncated, ET(k)| and contains all terms involving lower exponents. In order to compare the convergence of (A)tomic (AMT, AET) and (M)olecular (MMT, MET) expansions the corresponding relative errors have been plotted in Figures 3–5. The exponentially decreasing relative errors at short distances [Fig. 3(a) and (b) are due to penetration effects. They can be estimated from nonempirical atom–atom potentials |20|. As it may be seen on Figure 3(a) and (b) the remaining multipole component of electrostatic interaction energy can be reasonably estimated from cumulative atomic expansion even for separations typical for hydrogen-bonded dimers.

It can be seen that AET(3) expansion including all R^{-3} terms yields much better results than frequently used AET(1) scheme based on Mulliken charges only $(\Sigma_{a\in A} \Sigma_{b\in B} q_a q_b R_{ab}^{-1})$. Besides, AET(1) displays strong basis set dependence compensated only partly on higher order expansions.
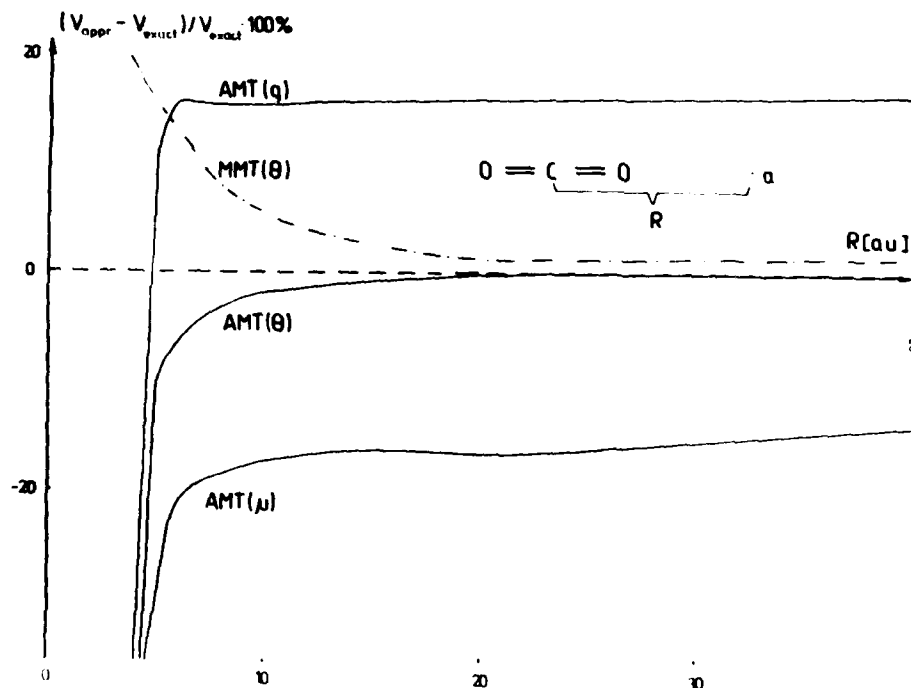

Figure 4. Relative errors of various multipolar estimates of the molecular electrostatic potential V calculated in minimal all-electron $(6s3p/3s)$ basis set [32] as a function of distance R for CO_2 molecule.
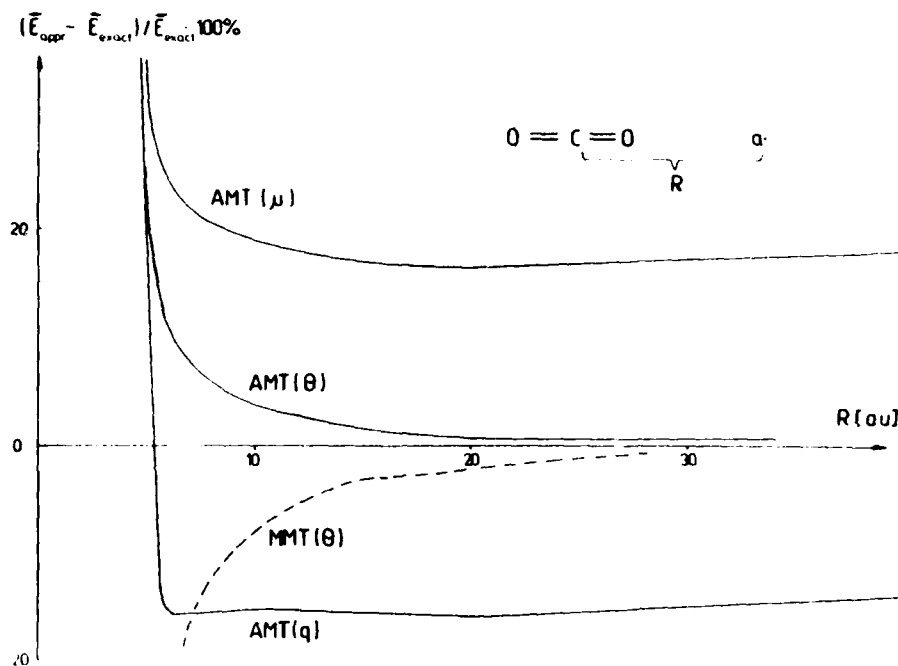
Figure 5. Relative errors of various multipolar estimates of electrical field values $\vec{E}$ calculated in minimal all electron $6s3p$ basis set [32] as a function of distance R for CO_2 molecule.

Acknowledgments

Bibliography

[1] W. A. Sokalski and R. A. Poirier, Chem. Phys. Lett. 98, 86 (1983).

[2] R. Rein, Adv. Quantum Chem. 7, 335 (1973).

[3] R. Dovesi, C. Pisani, F. Ricca, and C. Roetti, J. Chem. Soc. Faraday. Trans. II 70, 1381 (1974).

[4] P. Claverie, in Localization and Delocalization in Quantum Chemistry, Vol. 2, Chalvet R. Daudel, S. Diner and J. P. Malrieu, Eds. (Dordrecht, Reidel 1976).

[5] M. Mezei and E. S. Campbell, Theor. Chim. Acta 43, 227 (1977).

[6] F. L. Hirshfeld, Theor. Chim. Acta. 44, 129 (1977).

[7] A. Pullman and D. Perahia, Theor. Chim. Acta. 48, 29 (1978).

[8] J. Bentley, in The Role of Electrostatic Potential in Chemistry, Proceedings of 1980 ACS Sympo-sium, Houston.

[9] A. J. Stone, Chem. Phys. Lett. 83, 233 (1981).

[10] G. Karlstrom, 5th Seminar on Computational Methods in Quantum Chemistry, 1981

[11] J. Fernandez Rico, J. L. Alvarez-Collado, and M. Paniagua, Mol. Phys. 56, 1145 (1985)

[12] R. S. Mulliken, J. Chem. Phys. 23, 1833 (1955).

[13] A. D. Buckingham and P. W. Fowler, J. Chem. Phys. 79, 6426 (1983)

[14] A. D. Buckingham and P. W. Fowler, Can. J. Chem. **63**, 2018 (1985).

[15] G. J. B. Hurst, P. W. Fowler, A. J. Stone, and A. D. Buckingham, Int. J. Quantum Chem. **29**, 1223 (1986).

[16] D. L. Cummings, A. P. L. Rendell, D. J. Swanton, G. B. Bacskay, and N. S. Hush, Int. Rev. Phys. Chem. **5**, 139 (1986).

[17] M. A. Spackman, J. Chem. Phys. **85**, 6587 (1986).

[18] S. L. Price and A. J. Stone, J. Chem. Phys. **86**, 2859 (1987).

[19] N. Gresh, P. Claverie and A. Pullman, Int. J. Quantum Chem.: Quantum Chem. Symp. **13**, 243 (1979).

[20] W. A. Sokalski, A. H. Lowrey, S. Roszak, V. Lewchenko, J. M. Blaisdell, P. C. Hariharan, and J. J. Kaufman, J. Comp. Chem. **7**, 693 (1986).

[21] M. A. Spackman, J. Chem. Phys. **85**, 6579 (1986).

[22] R. M. Berns and A. van der Avoird, J. Chem. Phys. **72**, 610 (1980).

[23] J. T. Brobjer and J. N. Murrell, J. Chem. Soc. Faraday Trans. II **78**, 1853 (1982).

[24] T. Thole and P. T. van Duijnen, Ther. Chim. Acta **63**, 209 (1983).

[25] J. T. Brobjer and J. N. Murrell, Mol. Phys. **50**, 885 (1983).

[26] F. A. Momany, J. Phys. Chem. **82**, 592 (1978).

[27] P. H. Smit, J. L. Derissen, and F. B. van Duijneveldt, Mol. Phys. **87**, 521 (1979).

[28] S. R. Cox and P. A. Williams, J. Comp. Chem. **2**, 304 (1981).

[29] A. Agresti, R. Bonaccorsi, and J. Tomasi, Theor. Chim. Acta **53**, 215 (1979).

[30] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Protela, Jr., and P. Weiner, J. Am. Chem. Soc. **106**, 765 (1984).

[31] N. K. Ray, M. Shibata, G. Bolis, and R. Rein, Chem. Phys. Lett. **109**, 352 (1984).

[32] H. E. Popkie and J. J. Kaufman, Int. J. Quantum Chem.: Quantum Chem. Symp. **10**, 47 (1976).

[33] W. A. Sokalski, P. C. Hariharan, and J. J. Kaufman, J. Phys. Chem. **87**, 2803 (1983).

[34] W. A. Sokalski, P. C. Hariharan, and J. J. Kaufman, J. Comp. Chem. **4**, 506 (1983).

[35] A. Sawaryn and W. A. Sokalski, Int. J. Quantum Chem. **16**, 293 (1979).

[36] W. A. Sokalski, Int. J. Quantum Chem. **20**, 231 (1981).

[37] S. Huzinaga and S. Narita, Isr. J. Chem. **19**, 242 (1980).

[38] S. Arnott and P. W. L. Hukins, Biochem. Biophys. Acta **47**, 1504 (1972).

[39] J. Langlet, P. Claverie, F. Caron, and J. C. Boeuve, Int. J. Quantum Chem. **19**, 299 (1981).

[40] A. Pullman, H. Berthod, and N. Gresh, Int. J. Quantum Chem., Quantum Chem. Symp. **10**, 59 (1976).

[41] W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**, 2657 (1969).

[42] L. F. Sukhodub, Studia Biophysica **101**, 27 (1984).

[43] W. A. Sokalski and A. Sawaryn, J. Chem. Phys. **87**, 526 (1987).

[44] M. Aida and C. Nagata, Int. J. Quantum Chem. **29**, 1253 (1986).

[45] L. Piela and J. Delhalle, Ann. Soc. Sci. Bruxelles **92**, 42 (1978).

[46] D. Maillard and B. Silvi, Mol. Phys. **40**, 933 (1980).

[47] M. Aida, K. Yamane and C. Nagata, Mut. Res. **173**, 49 (1986).

Group Theory of Shapes
of Asymmetric Biomolecules

PAUL G. MEZEY

*Department of Chemistry and Department of Mathematics,
University of Saskatchewan, Saskatoon, Saskatchewan, Canada, S7N 0W0*

Abstract

Most of the important conformations of biomolecules possess only trivial symmetry. Consequently, symmetry groups have no roles in the characterization of the shapes of such conformations. However, an alternative group theoretical model, based on homology groups of algebraic topology, provides a detailed description of shapes for all conformations. These shape groups are useful for precise comparison of molecular shapes and are proposed as a computational tool for QSAR. A new computational method for the determination of various shape groups is described which is suitable for the simultaneous analysis of a pair or a family of molecular properties. In this note a general method is described and applied to the shape of electronic charge distribution along van der Waals surfaces.

Introduction

Molecular shape is one of the most fundamental concepts of chemistry. Chemical reactivity and most other chemical, physical, and biological properties of molecules are strongly dependent on molecular shape.

It is common to consider the arrangement of the atomic nuclei: the nuclear geometry and the associated network of formal bonds, as a descriptor of molecular shape. This is indeed common practice in the usual textbook representation of molecules. It is evident, however, that nuclear positions and lines of formal chemical bonds provide only a "skeleton" of the geometrical model of molecules, and that of the actual molecular shape. The shape of the "body" of the molecule requires additional information. It is the shape of the "body" of electronic charge distribution, or the shape of electrostatic potentials, or the shape of the van der Waals surface of the molecule, among other descriptors, which go beyond the simple skeletal description of molecular shape.

With the ever-increasing demands of the pharmaceutical industry for more efficient methods of computer-aided drug design and molecular engineering, there is a new interest in a realistic, yet simple representation of molecular shape that is suitable for taking into account the full three-dimensional nature of the "body" of molecules. This "body" however, is a quantum-mechanical entity, controlled by the properties of three-dimensional electron distributions. Fortunately, only some of the properties of electron density are relevant to molecular shape, and it is possible to extract this information and represent it in a remarkably simple algebraic form using group theory, in terms of *symmetry-independent shape groups* of electron densities [1, 2]. Of

course, a much simpler and somewhat less revealing description is possible by apply-
ing the same group theoretical technique to van der Waals surfaces.

The purpose of this article is to describe a new method for the characterization of
molecular shape in terms of two properties: electron density and van der Waals sur-
faces. Both of these properties are of great interest in drug design, and the unified
treatment we present in this study is expected to be easily applicable.

The precise characterization of the shapes of molecules which possess no more
than trivial symmetry in their most stable conformations is of fundamental impor-
tance in the analysis of biochemical problems on the molecular level. Recently a
symmetry-independent group-theoretical method has been proposed for such an anal-
ysis [1, 2]. The above method is based on a curvature analysis of various contour sur-
faces of the molecule or molecular fragment of interest, which leads to a concise
group theoretical description of the interrelations of surface domains of various cur-
vature properties. The contour surfaces considered are the equipotential contours of
the electrostatic potential generated by the molecule; the isodensity contours of elec-
tronic charge densities; contour surfaces of molecular orbitals; or van der Waals sur-
faces [1, 2]. These surfaces can be calculated by routine quantum chemical methods
or by other techniques [3-5]. The resulting groups are the homology and
cohomology groups of truncated contour surfaces; among them the one-dimensional
homology groups H^1 contain the most relevant chemical information [1, 2].

In the present article we shall describe an extension of the above technique for the
analysis of two (and by a straightforward generalization, of several) molecular prop-
erties relevant to molecular shape. In describing the method we shall consider the
problem of electronic density variations along the van der Waals surface, a problem
of obvious relevance to the study of chemical reactivity, polarizibility, solute-solvent
interactions, orientation effects, drug receptor interactions, and so on. However,
the method is general and is applicable for the analysis of the interrelations of any
two three-dimensional molecular functions, for example, HOMO contours and total
charge density.

We shall not repeat the actual derivation of homology groups of individual contour
surfaces; the method has been described earlier [1, 2], and there are excellent text-
books available for more mathematical background [6-9]. The discussion will be re-
stricted to the only novel aspect of the new method: using interpenetrating surfaces
for the development of a new family of symmetry-independent shape groups.

Interpenetrating Contour Surfaces

Consider a three-dimensional cartesian coordinate system attached to the molecule
or molecular fragment of interest and two molecular functions $f_1(\mathbf{r})$ and $f_2(\mathbf{r})$ of the
three-dimensional position variable $\mathbf{r}$. The respective contour surfaces $G_1(a_1)$ and
$G_2(a_2)$ are defined as:

$$G_1(a_1) = \{\mathbf{r} : f_1(\mathbf{r}) = a_1\} \tag{1}$$

and

$$G_2(a_2) = \{\mathbf{r} : f_2(\mathbf{r}) = a_2\}. \tag{2}$$

for some suitably chosen constants a_1 and a_2. For example, one may picture $G_1(a_1)$ as the *charge density contour* drawn about a molecule where the charge density $f_1(\mathbf{r})$ is equal to the value a_1, whereas $G_2(a_2)$ may be taken as the van der Waals contour surface. For the latter choice one may define function $f_2(\mathbf{r})$ as:

$$f_2(\mathbf{r}) = \begin{cases} 1 & \text{on the van der Waals surface} \\ 0 & \text{elsewhere,} \end{cases} \tag{3}$$

and one may choose $a_2 = 1$ in Eq. (2). Alternative definitions, continuous in $\mathbf{r}$, may also be used, which definitions may express the "penetrability" or "hardness" of the molecular neighborhood, for example, by taking a value for $f_2(\mathbf{r})$ less than 1.0 for points $\mathbf{r}$ outside of the van der Waals surface and greater than 1.0 within. However, for our present purposes definition (3) is appropriate. Since for this choice $a_2 = 1$, the argument of $G_2(a_2)$ can be omitted and we shall write simply G_2.

The methods of symmetry-independent shape groups [1, 2] are applicable for the detailed characterization of the shapes of both the charge density contour $G_1(a_1)$ and the van der Waals contour G_2. However, our present purpose is to analyze the *interrelations between two molecular properties*. In the actual example these properties are the shape of the electronic distribution and the shape of the van der Waals surface. The analysis can be accomplished by generating the homology groups of the object obtained by allowing these two surfaces to interpenetrate one another, as shown schematically in Figure 1.

The pattern of mutual interpenetration of the two surfaces G_1 and G_2 can be characterized by either one of four *truncated* surfaces:

$$G_1(f_2 \geq a_2) \tag{4}$$

$$G_1(f_2 < a_2) \tag{5}$$

$$G_2(f_1 \geq a_1) \tag{6}$$

$$G_2(f_1 < a_1) \tag{7}$$

In the above sets only those points of the original point set G_i are retained which satisfy the condition for the *other* function f_j, as stated in the parentheses. It is sufficient to choose only one of the above representations for a shape characterization, for example, in Figure 1 the truncated surface $G_2(f_1 \geq a_1)$ is chosen, that is, the collection of all those points of the van der Waals surface G_2 where the electronic charge density $f_1(\mathbf{r})$ is greater than or equal to the value a_1.

It is clear that an actual interpenetration of two general contour surfaces $G_1(a_1)$ and $G_2(a_2)$ may occur only for restricted ranges of parameters a_1 and a_2; no interpenetration occurs if one of these contour surfaces completely surrounds the other. For example, in the case of an electronic density contour $G_1(a_1)$ and the van der Waals surface G_2 no interpenetration occurs if the charge density contour value a_1 is chosen as either too large or too small. For a large a_1 value the contour $G_1(a_1)$ lies much too close to the nuclei, and hence the entire $G_1(a_1)$ surface is contained within the volume enveloped by the van der Waals surface G_2. On the other hand, for a low value of charge density parameter a_1 the surface $G_1(a_1)$ lies far from the nuclei, and the entire
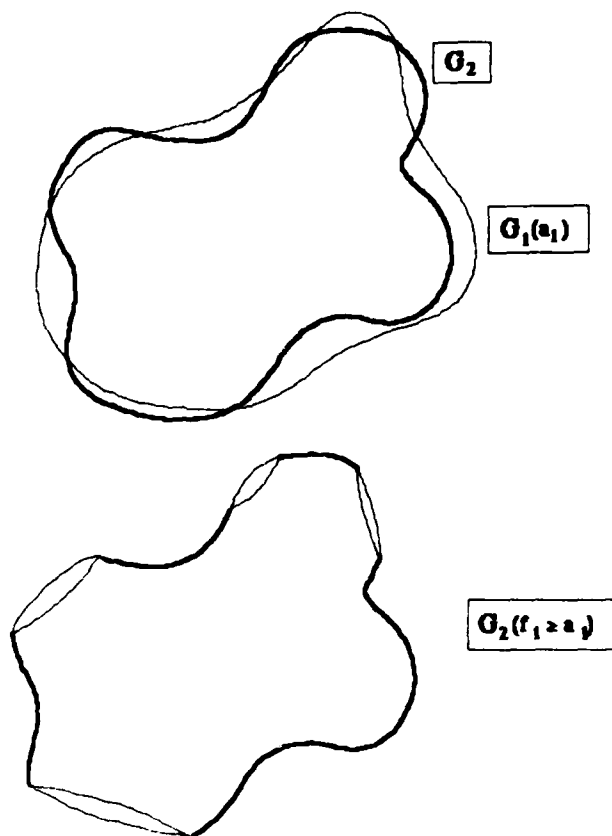
Figure 1. Interpenetrating contour surfaces $G_1(a_1)$ and G_2. Elimination of low charge density domains from the van der Waals surface G_2 along the lines of intersection leads to truncated (punctured) surface $G_2(f_1 \geq a_1)$. The resulting homology groups $H^1(G_2(f_1 \geq a_1)$ depend on the charge density contour parameter a_1, and give a concise description of the shapes of high and low density domains along the van der Waals surface.

van der Waals surface G_2 is "nested" within $G_1(a_1)$. In both of such extreme cases the two contour surfaces have no common points and a formal truncation according to conditions (4)–(7) either leaves the original contour surface unchanged, or eliminates it altogether.

In Figure 1 a chemically more interesting intermediate case is shown where the two contour surfaces G_1 and G_2 do have common points, hence a truncation does lead to topologically significant changes. The truncated surface $G_2(f_1 \geq a_1)$ identifies the domains of high and low electron density along the van der Waals surface G_2. The topological pattern of these domains can be described by the one-dimensional homology group of the truncated surface, denoted by

$$H^1(G_2(f_1 \geq a_1)) \tag{8}$$

Using a construction identical to that described in detail in Ref. 1. in the case of the example shown in Figure 1 the homology group H^1 obtained is isomorphic to the

Abelian group of three free generators g_1, g_2, g_3, and has elements of the form:

$$k_1 g_1 + k_2 g_2 + k_3 g_3 \qquad (k_i \text{ are integers}) \qquad (9)$$

The mutual interrelations among the low and high electron density domains along the van der Waals surface are described in detail by the *family* of one-dimensional homology groups, as the parameter value a_1 sweeps over a chemically significant parameter range $[a'_1, a''_1]$, where a'_1 and a''_1 are some low and high charge density values where no interpenetration of G_1 and G_2 occurs (corresponding to the two possible nesting arrangements: G_2 within G_1 and G_1 within G_2, respectively). For the value $a_1 = a'_1$ the surface $G_1(a'_1)$ completely surrounds the van der Waals surface G_2 hence

$$G_2(f_1 \geq a'_1) = G_2 \qquad (10)$$

and the homology group

$$H^1(G_2(f_1 \geq a'_1)) = H^1(G_2) \qquad (11)$$

is the trivial group. For the same value $a_1 = a'_1$ the zero- and two-dimensional homology groups are isomorphic to one another and to the additive group of integers:

$$H^0(G_2(f_1 \geq a'_1)) = H^0(G_2) \cong H^2(G_2(f_1 \geq a'_1)) = H^2(G_2) \qquad (12)$$

If a_1 is increased to a higher value at which a *single* interpenetration occurs, leading to the truncation of a single, simply connected domain of *low* electron density along the van der Waals surface G_2, then Eq. (10) no longer holds, and Eq. (11) is replaced by the isomorphism

$$H^1(G_2(f_1 \geq a_1)) \cong H^1(G_2) \qquad (13)$$

Whereas for this value a_1 the truncated surface has a one-dimensional homology group isomorphic to that for value a'_1, the two-dimensional homology group has changed from one isomorphic to the additive group of integers to a trivial group.

In general, a further increase in the value of a_1 leads to additional interpenetrations and truncations, which leave the zero- and two-dimensional homology groups invariant but changes the one-dimensional homology group $H^1(G_2(f_1 \geq a_1))$ to more complicated groups which are Abelian groups of several free generators. One example is shown in Figure 1 where the resulting one-dimensional homology group has three free generators. By continuing the increase of value a_1, a larger part of $G_1(a_1)$ appears within the volume enclosed by the van der Waals surface G_2 and fewer but larger domains of G_2 are truncated, leading to simpler one-dimensional homology groups of fewer generators. For a large enough charge density value a_1, for example, for the value $a_1 = a'_1$, the entire charge density contour surface $G_1(a''_1)$ is enclosed by the van der Waals surface G_2, and truncation condition (6) eliminates the entire surface G_2. In this latter extreme case there is no topological object left to be analyzed and no homology groups are generated.

The above homology groups, combined with the shape groups of original objects $G_1(a_1)$ and $G_2(a_2)$, described in earlier studies [1, 2], give a complete characterization of both the shapes and the interrelations of the corresponding molecular functions, for example, of charge densities and van der Waals surfaces.

Summary

A method is described to extend the application of symmetry-independent shape groups of molecules or molecular fragments to the characterization of the interplay of two or several molecular properties, such as electronic charge densities and van der Waals surfaces. In the example used as illustration of the method, the homology groups give a concise description of the interrelations between domains of high and low charge density along van der Waals contour surfaces of molecules. These domains and their shapes are of fundamental importance in controlling intermolecular interactions, such as solvent-solute and drug-receptor interactions.

Acknowledgment

Bibliography

[1] P. G. Mezey, Int. J. Quantum Chem., Quant. Biol. Symp. 12, 113 (1986).
[2] P. G. Mezey, J. Comput. Chem. 8, 462 (1987).
[3] M. L. Conolly, J. Am. Chem. Soc. 107, 1118 (1985).
[4] A. Y. Meyer, J. Chem. Soc. Perkin Trans. 2 (in press).
[5] A. Y. Meyer, J. Mol. Struct. Theochem. 124, 93 (1985).
[6] E. H. Spanier, *Algebraic Topology* (McGraw-Hill, New York, 1966).
[7] W. S. Massey, *Homology and Cohomology Theory* (Dekker, New York, 1978).
[8] J. Vick, *Homology Theory* (Academic Press, New York, 1973).
[9] J. Milnor, *Morse Theory, Annals of . lath. Studies*, Vol. 51 (Princeton University Press, Princeton, 1973).

A Method for the Characterization of Molecular Conformations

GUSTAVO A. ARTECA* AND PAUL G. MEZEY[†]

Department of Chemistry, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0

Abstract

In the course of conformational motions of molecules the changes in shapes of electronic charge distributions follow that of the nuclear framework. However, this coupling between the changes in the nuclear geometry and electron density may depend on the actual nuclear displacement; the coupling may be weak or strong for a given conformational motion. It is of some interest to analyze how faithfully the charge density variations follow the nuclear displacements in a family of conformational rearrangements. In certain cases small conformational changes may induce large changes in the shape of charge density distributions, while in other cases large and qualitatively important conformational changes may involve qualitatively inessential distortions in the shape of electron distributions. In this article we describe a new classification of conformations based on those domains of nuclear configuration space within which the "shape groups" (symmetry independent homology groups) of the electric charge density remain invariant. Such an analysis might be valuable when seeking correlations between molecular structure and certain biological or biochemical activities.

Introduction

The description of dynamical changes in molecules is of utmost importance in several fields. In particular, the variation of certain physical properties under conformational changes constitutes a problem of general interest. The extent of such variations may serve, in fact, as a criterion to classify the conformational rearrangements. This paper is essentially devoted to study this latter possibility.

One of the useful and intuitively simple properties to characterize a molecule, and the processes it undergoes, is the molecular shape (MS). It is commonly assumed that at least a restricted set of molecular properties can be explained in terms of the MS. Most biological and biochemical activities are among the characteristics correlating directly with the MS. However, the correlation of these activities with conformational rearrangements is comparatively not so clear. The basic reason for this difference is that a conformational change may not be followed by a significant and meaningful modification in shape. In other words, there exists in general an "uncoupling" between the changes in configurational space and the changes in the three-dimensional space where the MS is described. Some rearrangements in the former space will be

---

more important than others, because they may lead to qualitatively significant shape modifications. According to these modifications, one might be able to introduce a sort of decompostiion of the configurational space into subsets. This decomposition would provide a detailed shape characterization of conformational changes. In this paper we provide a possible method to solve this problem.

There is no unique way to evaluate and characterize the MS. Several alternatives have been proposed to accomplish this goal. First, isodensity contours [see e.g., Refs. 1–3] and electrostatic isopotential contour diagrams [4–8, and references quoted therein] are a common source of data qualitatively representative of MS. On the other hand, algebraic and geometrical topology supply appropriate tools for their characterization [7, 8]. This is particularly important if, as most commonly, the application of standard point group theory results in no insight, due to total lack of symmetry.

This article shows how the MS, defined with the aid of algebraic topology, is dependent on the conformational rearrangement. The central idea is very simple and can be summarized as follows. Let us suppose that a physically meaningful quantity is selected to describe the MS, say, for instance, the electronic charge distribution. In this case, an isodensity contour defines a surface in 3-space. In the overwhelming majority of cases, this surface will be a topological 2-sphere or a collection of them; however, tori and manifolds with higher genus are possible. Following an appropriate procedure, that we discuss later, it is possible to associate with this contour a series of new surfaces, no longer topologically spherical. Each of these new surfaces can be completely characterized by the symmetry-independent homology or cohomology groups ("shape groups") [see e.g., Refs. 7–10] Accordingly, the main features of the MS can be described, in general, by a set of homology groups.

In the above approach, the MS is associated with a given conformation. By extension, a set of homology groups is assigned to each nuclear configuration. Based on this assignment, we propose to describe the coupling between the shape of the charge density distributions and the molecular rearrangements in the following way: the coupling will be nonessential in those subsets of the configurational space where the shape groups are invariant. According to this model, only the rearrangements leading to transitions between "shape domains" would imply an actual essential modification in shape-dependent properties.

In order to perform the above analysis the paper has been organized as follows: in the Methods section we review the basic principles to apply the method. Several criteria that can be followed to describe important features of the MS undergoing conformational changes are discussed. Using any of these criteria the shape analysis can be accomplished, but in general they will lead to different, perhaps conplementary descriptions. A brief explanation is given about the construction of the shape groups. In "Shape Regions in Configuration Space," the method is illustrated first by studying the conformational distribution of shape groups in the particular case of a triatomic system. Some connectedness properties of the domains in configuration space are also stressed. The application to larger systems is illustrated by considering a subset of the configuration space composed by a stretching and internal rotation coordinates in a substituted benzene derivative. Further comments and conclusions are found in the final section.

Method

Let us consider a system of N nuclei, and the corresponding configuration space ^{3N}R, where the position vector for each nucleus is specified in terms of three cartesian coordinates. In order to discuss the MS we will attach a three-dimensional object to each point in ^{3N}R. If any of such objects is to be appropriate to characterize the MS, then it will have to be translationally and rotationally invariant. To this end, we will study here the conformational changes in the reduced nuclear configuration space M [11, 12]. Space M is the quotient space of ^{3N}R with equivalence relation of superposition under rigid translation and rotation. Furthermore, it can be proved that M, as well as ^{3N}R, is a metric space with an appropriate metric [11, 12]. We will indicate with $\mathbf{R}$ an element of that space ($\mathbf{R} \in M$), representing a nuclear configuration. In turn, $\mathbf{r}$ will stand for a vector in 3-space ($\mathbf{r} \in {}^3R$), representing a point on the object chosen to describe the shape. Vector $\mathbf{r}$ will not in general represent any nuclear position.

There exist several alternatives that can be followed to define a three-dimensional entity, useful as a model for the MS. Let us consider a function $f(\mathbf{r}, \mathbf{R})$, with the property:

$$\lim_{\|\mathbf{r}\| \to \infty} |f(\mathbf{r}, \mathbf{R})| = 0, \qquad \mathbf{R} \in M. \qquad (1)$$

Function $f(\mathbf{r}, \mathbf{R})$ depends parametrically on the nuclear coordinates, and it is real, and single-valued, for each point in 3-space. Some one-electron, marginal properties fulfill these requisites, together with Eq. (1); among them the total electronic charge density and the effective electrostatic potential are the most straightforward examples [1–8].

We will introduce now a *closed* contour surface in ^3R, depending parametrically on $\mathbf{R}$ and on a real number a related to the function f as follows:

$$G(a, \mathbf{R}) = \{\mathbf{r} \in {}^3R: f(\mathbf{r}, \mathbf{R}) = a\}, \qquad \mathbf{R} \in M. \qquad (2)$$

The number a can be regarded as a parameter, or it can be fixed beforehand, or, as we will see below, some physically significant values for a can be determined implicitly in terms of some physical constraints. $G(a, \mathbf{R})$ can be a single closed surface or a collection of a number of disjoint closed surfaces.

Let $G^*(a, \mathbf{R})$ denote the set of points enclosed by the surface $G(a, \mathbf{R})$, and let us define a new set:

$$F(a, \mathbf{R}) = G^*(a, \mathbf{R}) \cup G(a, \mathbf{R}), \qquad (3)$$

that is, in fact, a level set of one of two types. Depending on the function chosen, $G^*(a, \mathbf{R})$ may contain points in ^3R where $f(\mathbf{r}, \mathbf{R})$ is larger than, or smaller than the parameter a. Without any loss of generality, the following treatment is applicable to all functions $f(\mathbf{r}, \mathbf{R})$ by an appropriate choice of sign: we will consider from now on that $f(\mathbf{r}, \mathbf{R})$ is the charge density; consequently, the set (3) will contain points for which the density is not smaller than a. Furthermore, the origin of the cartesian system of space ^3R will be considered to be within $F(a, \mathbf{R})$. With these definitions, G is a constant contour surface of function $f(\mathbf{r}, \mathbf{R})$ in ^3R and F its corresponding level set [7, 8].

As mentioned above the constant a can be allowed to take values in a given interval; in this case we obtain a *family* of isodensity contours for shape characterization

of the molecule in the conformation $\mathbf{R}$ [8]. However, there exists a series of alternative choices for a that might prove valuable in other cases. We summarize here some of the more interesting possibilities:

First, let us consider an arbitrary contour $G(a', \mathbf{R})$, with $a' > 0$, some constant. One may set a numerical value for the volume $V(\mathbf{R})$ enclosed by such a surface, and determine the value of the parameter a in Eqs. (2) and (3) implicitly from the chosen fixed value for the volume:

$$V(\mathbf{R}) = \int_{G(a, \mathbf{R})} d\mathbf{r}. \tag{4}$$

As stated above, Eq. (4) indicates the integration in ^3R restricted to the interior of a contour surface. If we now request that $V(\mathbf{R}) = V_0 = \text{const.}$ for all $\mathbf{R} \in M$, then the charge density value a will depend implicitly on V_0, and will vary parametrically with $\mathbf{R}$: $a = a(V_0; \mathbf{R})$.

The shape of a density contour enclosing a given volume is a physically appealing description of the MS. It may be linked to a van der Waals surface description of a molecule. In principle, as $F(a(V_0; \mathbf{R}); \mathbf{R})$ will change with the conformations, the value a defining the contour will also change. Consequently, the shape analysis obtained following this criterion will be different, and complementary to the one commented previously.

Second, one may define the constant a not in terms of a volume but in terms of regions within which the number of electrons N_e is conserved for all conformations. This can be simply done by introducing the constraint on the integrated density:

$$N_e(\mathbf{R}) = \int_{G(a, \mathbf{R})} f(\mathbf{r}, \mathbf{R}) d\mathbf{r}. \tag{5}$$

where the meaning of the integration domain is the same as in Eq. (4). Requiring $N_e(\mathbf{R}) = N_e^0 = \text{const}$, we obtain $a = a(N_e; \mathbf{R})$. The idea of using subregions in ^3R with fractional number of electrons has proved to be useful in other contexts [Ref. 13, and others quoted therein]. However, it has never been exploited to provide a complete shape characterization of the overall molecule undergoing conformational rearrangements, even though it would provide a conceptually clear physical approach. From the present viewpoint, the evolution of the MS with changes in space M would be followed by focusing the attention to the way in which the isodensity surface is modified in order to keep constant part of the total electronic charge enclosed.

It would be expected that under certain conditions reactivity and other properties could be correlated to the steric or "volumetric" effects in molecules. In other cases the molecular interactions might depend on the total charge enclosed in a given region of space. The contribution of some configurational reordering to the above effects could be estimated by carrying out the shape analysis with the appropriate choice for constant a.

Third, let us consider now the following situation: for a given (initial) configuration $\mathbf{R}$ the density contour $G(a, \mathbf{R})$ is chosen for some constant a. Then, the integrated density is evaluated, leading to a "fraction" of the total electronic charge N_e^0. As the next step, some underlying structure of the configuration space may be

revealed upon finding those trajectories in space M along which both N_c^o and a are conserved. These trajectories should be highly significant for they correspond to the conservation of an integrated quantity as well as a boundary condition. Trajectories along which both a and N_c^o are exactly conserved may eventually not exist for a given initial configuration R. In this case we simply look for conservation of a and minimum change in $N_c(R)$ with respect to N_c^o. The changes in the MS along the above paths are interesting, since they correspond to configuration changes with the least fluctuation of charge enclosed by a chosen density contour.

In what follows we will restrict ourselves to the first and simplest criterion, for the sake of simplicity and brevity. Nonetheless, it should be kept in mind that it is only one among many other possible choices.

To proceed with the MS characterization we will essentially associate a regular simplicial complex (or a collection of them) to $G(a, R)$ and determine the homology (or cohomology) groups induced on it by cellular decomposition. Details of the method have been discussed previously in the literature for frozen nuclear conformations [7, 8]; accordingly, we will include here only some basic steps adapted to our needs in order to include conformational changes.

Let us assume that the potential energy hypersurface is pathwise connected; we can define consequently a surjective mapping $g: [0, 1] \mapsto M$, so that the $g(t) = R = R(t)$, $t \in [0, 1]$, nuclear configurations collectively describe a path of a possible conformational change. From now on the parameter t will measure the evolution of the transformation in the metric space M. $R(0)$ and $R(1)$ stand for the initial and final configurations, respectively. The contour on which the analysis will be performed [Eq. (2)] becomes now $G(a, R(t))$ [and $F(a, R(t))$ in Eq. (3)].

In order to transform $G(a, R(t))$ into a cell complex (or a collection of them) we will develop first a contour partitioning of the surface into domains of specified curvature properties [7, 8]. To this end, we introduce a real parameter b, that will take the role of a "reference curvature" to determine subregions on $G(a, R(t))$. The use and meaning of this new parameter is as follows: consider first a point on the contour $G(a, R(t))$, say r_o, and the straight line defined by the normal to the surface at that point. Further, consider a sphere, whose center r_o' is on the above straight line. The sphere is tangent to the surface at point r_o, and it has radius $k = \|r_o - r_o'\|$. We may now introduce the curvature parameter b related to the radius k in the following way:

(1) $b = 0$ if the radius is infinite (the tangent sphere is a tangent plane [7]).

(2) $b = -1/k < 0$ if the center r_o' lies in the half-straight line extending from r_o toward the interior of $G(a, R(t))$. (Notice, however, that the center may or may not lie within $F(a, R(t))$.)

(3) $b = 1/k > 0$ if the center r_o' lies in the half-straight line extending from r_o to the exterior of $G(a, R(t))$.

As already discussed [8], one may determine a series of curvature domains on $G(a, R(t))$ using the above "tangent spheres" and "tangent planes" for different b values. These domains of curvature represent a generalization of concave, saddle, and convex $D_\mu(b)$ domains $(D_\mu(b) \subset G(a, R(t)))$ as follows [7, 8]:

a. $r_O \in D_0(b)$ ("concave" domain) if an infinitesimal spheric sector d_O of the tangent sphere with curvature b, centered about r_O, fulfills the property: $d_O \subset F(a, \mathbf{R}(t))$.

b. $r_O \in D_1(b)$ ("saddle" domain) if for the corresponding value of b there exist subsets $d_O' \subset d_O$ and $d_O'' \subset d_O$, so that: $d_O' \subset F(a, \mathbf{R}(t))$ and $d_O'' \not\subset F(a, \mathbf{R}(t))$.

c. $r_O \in D_2(b)$ ("convex" domain) if $d_O \not\subset G^*(a, \mathbf{R}(t))$, for the chosen value of b.

The notation $D_\mu(b)$ ($\mu = \mu(b) = 0, 1, 2$) follows the convention, previously used in the literature, of counting the number of negative eigenvalues of the two-dimensional Hessian matrix on $r_O \in G(a, \mathbf{R}(t))$, for $b = 0$ [7]. If in our case $b \neq 0$, μ will simply count how many of the above eigenvalues are smaller than the value b [8].

After completing the analysis of curvature explained above, we have a set $\{D_\mu(b)\}$ of domains associated to $G(a, \mathbf{R}(t))$. These domains are not necessarily connected. As a result, the previous contour partitioning can be used to obtain objects topologically different from a sphere proceeding this way: for a given value of b, choose a particular index value $\mu = \nu(b)$ ($\nu = 0$, 1, or 2) and *cut* the $D_\nu(b)$ domains from $G(a, \mathbf{R}(t))$. After this, we end up with a collection of objects:

$$A = \{[G_{\nu(b)}(a, \mathbf{R}(t))]^{(i)}, i = 1, 2, \ldots, J\}. \tag{6}$$

$G_{\nu(b)}$ is disjoint if $J > 1$. In this case, $[G_{\nu(b)}(a, \mathbf{R}(t))]^{(i)}$ stands for one of the pieces (maximum connected components) left from $G(a, \mathbf{R}(t))$. It is clear that if $D_\mu(b) \subset [G_{\nu(b)}(a, \mathbf{R}(t))]^{(i)}$, then $\mu \neq \nu$.

One may transform each subset $[G_{\nu(b)}(a, \mathbf{R}(t))]^{(i)}$ into a regular simplicial cell complex [9, 10],

$$K^{(i)}(\nu, b, a, t) = |[G_{\nu(b)}(a, \mathbf{R}(t))]^{(i)}|, \tag{7}$$

upon cell decomposition by triangulation. From now on we shall abbreviate the notation as $K^{(i)}(t) = K^{(i)}(\nu, b, a, t)$, understanding that ν, b, and a have been chosen beforehand. In this way the attention is focused to the conformational parameter t describing the present conformation in M.

The methods of homology (or cohomology) theory allow one to characterize completely the complex $K^{(i)}(t)$ upon finding some of its topological and homotopical invariants [7–10].

Using, for instance, the approach of homology theory, the central idea of the method can be summarized as follows: Let $C_p^i(t)$ be the set of all p-dimensional cells of $K^{(i)}(t)$. In our present case $p = 0$, 1, and 2. [For the basic concepts used here and henceforth the reader is referred, for example, to Refs. 7 and 8.] The pair formed by $C_p^i(t)$ and a boundary operator $\Delta_p: C_p^i(t) \rightarrow C_{p-1}^i(t)$, defined in terms of an incidence function between cells, induces a set of chain complexes on $K^{(i)}(t)$, for different values of p. Moreover, for each value of p we can build the following two Abelian subgroups of chains:

$$Z_p(K^{(i)}(t)) = \operatorname{Ker} \Delta_p, \tag{8a}$$

$$B_p(K^{(i)}(t)) = \operatorname{Im} \Delta_{p+1}; \qquad Z_p(K^{(i)}(t)) \supseteq B_p(K^{(i)}(t)). \tag{8b}$$

The group Z_p represents the set of each p-dimensional chain with zero boundary (p-cycles). Taking into consideration the operator property $\Delta_p\Delta_{p+1} = 0$ [9, 10], we know that the boundary of all $(p + 1)$-dimensional chains is a p-cycle. Eq. (8b) represents then the set of all bounding p-cycles. As a result, the p-dimensional homology group $H_p(k^{(i)}(t))$, is obtained as the quotient set (or, in the case of additive notation, the *difference* group):

$$H_p(K^{(i)}(t)) = Z_p(K^{(i)}(t))/B_p(K^{(i)}(t)), \tag{9}$$

These groups are finitely generated Abelian groups. According to the fundamental theorem of the Abelian groups [9, 10] the group (9) is isomorphic to the direct sum of a number of free cyclic Abelian groups and a number of finite cyclic (torsion) groups. The rank of the free cyclic component is the integer $b_p(K^{(i)}(t))$, known as the p-th Betti number of the homology group. In all the examples in this paper no torsion component is present. Consequently H_p has b_p generators. These numbers are topological and homotopical invariants. Following this procedure, we obtain a detailed characterization of the shape of the isodensity contour $G(a, \mathbf{R}(t))$ in terms of the set of Betti numbers of the cell complexes related to the disjoint pieces $[G_{\nu b_i}(a, \mathbf{R}(t))]^{(i)}$ [Eq. (6)]. This set can be described in terms of the J-tuple:

$$e_p(t) = (b_p(K^{(1)}(t)), b_p(K^{(2)}(t)), \ldots, b_p(K^{(J)}(t))),$$

$$p = 0, 1, 2; \quad J \geq 1. \tag{10}$$

In what follows we will restrict the discussion to the most important case of the one-dimensional chains (i.e., we shall discuss only $e_1(t)$). As we mentioned above, $e_p(t)$ will be different for a different choice of the level-set parameter a, as well as for an alternative choice of the reference curvature b. Furthermore, a truncation of the domains $D_{\nu'}(b)$, instead of $D_\nu(b)$ ($\nu' \neq \nu$), would lead to a different, complementary description.

The previous analysis shows that a very detailed shape characterization is available for every transformation path $\mathbf{R}(t)$ in configuration space. In order to focus the attention on some particular effects one may only consider a range of curvatures b for a given a, or *vice versa*. In the next section we illustrate briefly some of the possibilities analyzing some simple conformational changes.

It is clear that along a trajectory $\mathbf{R}(t)$ the more interesting points are those at which some changes in the J-tuple $e_1(t)$ take place, either in the values or in the number of its components. When crossing those points ("transition points"), that we will indicate with $\mathbf{R}(t_1^*)$, $\mathbf{R}(t_2^*)$, and so forth (in order of occurrence), an infinitesimally small change in configuration space is accompanied by a significant change in the MS. This can be expressed in a different way: consider two subsequent "transition points" $\mathbf{R}(t_i^*)$ and $\mathbf{R}(t_{i+1}^*)$; then, for all $\mathbf{R}(t)$ with $t_i^* < t < t_{i+1}^*$ the overall topological features of the shape are invariant under the change in space M. Our method then provides a framework to describe the "transition points" $\{\mathbf{R}(t_i^*)\}$ and their interrelations. In terms of all paths and their transition points the metric space M can be partitioned into domains where the shape is essentially conserved ("shape regions"). As

discussed above, this constancy is expressed in terms of the invariance of the set of Betti numbers. In the next section some examples of different "shape regions" in M are displayed.

There exists an alternative partitioning of configuration space, which may also possess some interesting properties. For a given conformation, the distribution of homology groups can be studied as a function of both parameters a and b [8]. This gives us a detailed description of the shape features of a large family of isodensity surfaces for a molecule in a frozen geometry. The partitioning of the a-b parameter plane into domains, characterized by the set (10), reveals a particular topological structure. In other words, there exists a clear pattern in the number and interrelations of these parameter domains of the a-b plane within which homology groups are preserved. This pattern will remain invariant for certain conformational reorderings. However, there will exist some particular configurations $\mathbf{R}(t)$ at which the topological structure of the a-b parameter plane will change. These configurations will be again very special points in the space M with regard to changes in the shape. An independent new partitioning of configuration space M can be given in terms of the topologically significant changes in the distribution of homology groups within the a-b parameter plane. This possibility will be discussed in a subsequent paper.

It is worth mentioning here that a similar method of "shape" classification in terms of homology (or cohomology) groups can be used to study domains of higher-dimensional reaction globes, defined in terms of the potential energy hypersurface. This appeared to be valuable when analyzing sets of reaction paths and mechanisms on the surface [14–16].

"Shape Regions" in Configuration Space

As a first illustration of the method discussed in the preceding section we have considered a triatomic system AB_2. From the point of view of symmetry group theory, the space M for this system can be partitioned into regions with the following symmetries: C_{2v}, C_s, D_{xh}, C_{xv}, and the spheric group K. As we shall see, our shape characterization based on density contours allows a further, very detailed, subdivision of those regions.

The full analysis of even a small system as AB_2 becomes a nontrivial task, owing to the presence of a large number of parameters. To specify a configurational array we need in this case two bond distances AB (q_1 and q_2) and the bond angle θ. We choose $\theta = 0$ corresponding to the C_{xv} symmetry (linear A-B-B) and $\theta = \pi$ to the linear array B-A-B (D_{xh} symmetry for $q_1 = q_2$). In addition to these three parameters we have the freedom in the choice of the truncation index μ as well as the constants b and a. We will restrict ourselves to the following case:

(1) The level-set constant a for the density is given a unique, representative value. It is chosen to be an intermediate value according to the following criteria: (1) a is *not large enough* to lead to a contour $G(a, \mathbf{R}(t))$ composed from *disjoint pieces* when both q_1 and q_2 are *small*; (2) It is *not small enough* to lead to a contour $G(a, \mathbf{R}(t))$ of a *single piece* when both q_1 and q_2 are *large*.

(2) The reference curvature b is chosen in the neighborhood of the value $b = 0$. That is, the curvature domains to be discussed do not differ too much from the conventional concave, saddle, and convex domains.

(3) The index for the truncated domains will be taken as $\nu(b) = 2$.

We have analyzed the most characteristic conformational rearrangements using the above constraints on a, b and $\nu(b)$. The results obtained are displayed schematically in Figures 1–4. The domains determined are qualitative in the sense that their areas are not necessarily proportional; only their interrelations and connectedness properties are of importance. In other words, the topology of the shape region partitioning is the feature of importance and it is properly represented. Furthermore, when drawing Figures 1–4 we have assumed that the boundaries between shape regions are differentiable functions of the internal coordinates.

To describe the shape of the electronic isodensity surface we considered the general properties of these contours for different nuclear arrays [see, for instance, Ref. 17]. With respect to the local concentration of charge we assume that $Z_A > 2Z_B$. That is, even when $q_1 = q_2$ and $\theta = 0$ there will be more electronic charge about nucleus A than about B (in other words, $G(a, R(t))$ will be somewhat "broad" around A).

It is clear that there will exist certain conformational arrays in which $G(a, R(t))$ will be formed by one or more closed surfaces, where all points on all of them will belong to $D_2(b)$ domains. According to our convention, these pieces must be cut away; as a result, nothing will be left for analysis. This represents a no group situation, that cannot be described by Eq. (10). To take this case into account we opt for the notation $N(i)$ ($i = 1, 2, 3, \ldots$), meaning that no group is presented because i closed surfaces ("pieces" of $G(a, R(t))$) have been eliminated because of being $D_2(b)$ domains. In addition, the notation $e_1(t)[N(i)]$ will be used when, even though there exists a series of domains left, still i closed pieces have been eliminated. This convention allows us to include in a single, local diagram some information that would only be revealed by an analysis for a range of negative b values [8].

Figure 1 shows schematically the typical results for the linear B-A-B array ($\theta = \pi$). The values of q_i are restricted to the range $q_i \leq L$; it is implied that for $q_i > L$ the diagram shows no new features. After truncation, in this example we find no other objects than topological cylinders. A simple analysis of Figure 1 reveals the following important relationships between conformational rearrangements and changes in the shape of the isodensity surface:

1. Quasisymmetrical stretchings: in this case the conformational reordering is given by $q_1 = qt$, $q_2 = q_1 + \delta$, with $\delta \ll 1$ and $t \geq 0$. We notice only two "transitions" (see Method) t_1^* and t_2^* corresponding to the changes $N(1) \rightarrow (1, 1)$ and $(1, 1) \rightarrow N(3)$. These transitions involve a change of two generators for the homology groups. This property may be encountered in those conformational changes where a nontrivial symmetry element is conserved along the path $R(t)$. In rigor, present transformation might strictly correspond to the case $\delta = 0$ (a symmetry plane perpendicular to the internuclear axis is preserved).
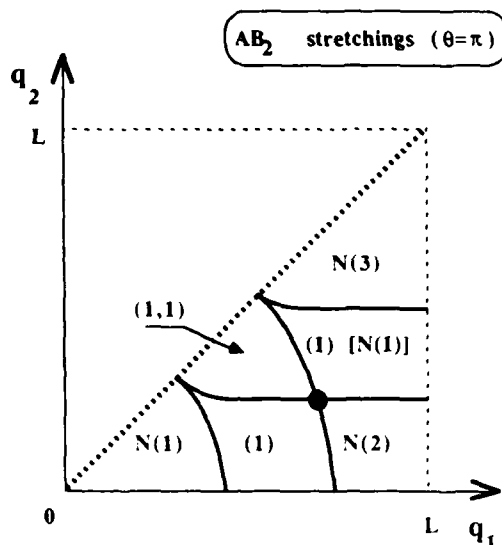
AB_2 stretchings $(\theta = \pi)$

q_2

L

N(3)

(1,1)

(1) [N(1)]

N(1)   (1)   N(2)

0   L q_1

Figure 1.   Schematic representation of isodensity surface distribution of shape groups for the triatomic linear system B-A-B. (Truncation type $\nu(b) = 2$; see text for details about the notation for domains. Black dots indicate points where four shape regions meet.)

2. Asymmetrical stretchings: this transformation corresponds to larger values of δ defined above. The relationships between "shape regions" reveal the likelihood of configuration changes with the transitions: $N(1) \to (1) \to (1,1) \to (1)$ $[N(1)] \to N(3)$ and $(1) \to N(2) \to (1)$ $[N(1)] \to N(3)$, but the low probability of the change in density shape described by: $N(1) \to (1) \to N(2) \to (1)$ $[N(1)] \to N(3)$. These "shape transitions" involve in all cases a change in a single generator after crossing a transition point.

Some other important processes that can be seen in this figure are the "compressions" from the symmetrical case ($q_1 = $ const, for all t, and $q_2 = q_1 t$) and "dilatations" ($q_2 = $ const, for all t, and $q_1 = q_2 t$). As it is displayed in the figure, the boundaries of some "shape regions" are parallel to paths corresponding to such transformations. Consequently, some transitions between shape regions are not allowed, or are very unlikely, along some of these conformational rearrangements. The unlikelihood of certain processes is a consequence of the fact that some boundaries between shape regions are curves turning rapidly into horizontal or vertical straight lines.

In addition, observe that, as a consequence of the assumption of differentiability of these boundaries and the symmetry of the shape region diagram with respect to the straight line $q_1 = q_2$, the curves become perpendicular to this line.

Figures 2 and 3 show the analogous information for $\theta = 0$ and $0 < \theta_i < \pi$. In the latter case, we consider the angle θ_i to be small. We notice that the structure of the partitioning into shape regions is richer in these cases. Furthermore, a new topological object is present in the latter example: $e_1(t) = (2)$, corresponding to a cylinder
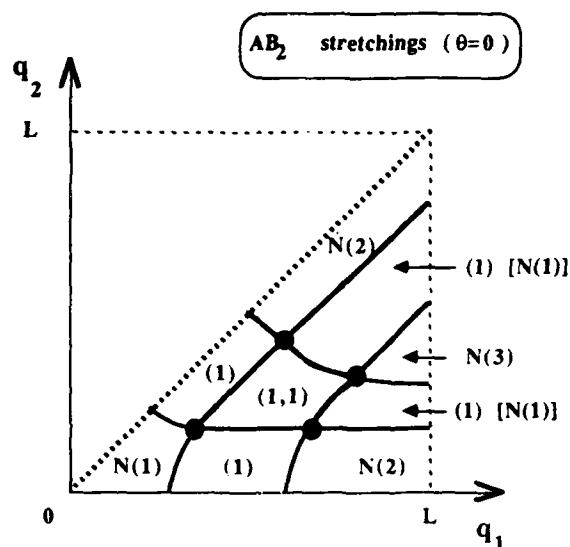
Figure 2.   Schematic representation of isodensity surface distribution of shape groups for the triatomic linear system A-B-B. (Truncation type $\nu(b) = 2$; see text for details. Black dots indicate points where four shape regions meet.)
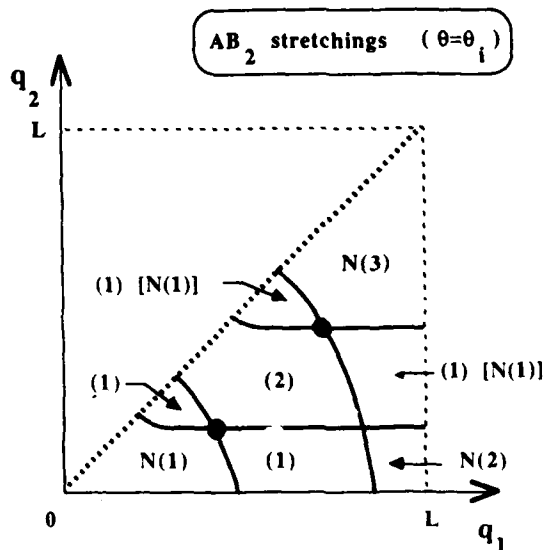

Figure 3.   Schematic representation of isodensity surface distribution of shape groups for the triatomic angular system AB_2 for some intermediate angle θ_i. (Truncation type $\nu(b) = 2$; see text for details. Black dots indicate points where four shape regions meet.)

with an "additional hole." The analysis of the important shape transitions for different conformational changes can be accomplished as shown above. It is worth mentioning that the structure of domains in Figures 2 and 3 being much detailed, these cases present several constraints to possible and unlikely shape transitions. The interrelations among the boundaries of regions present a pattern of regular features whose use might be valuable to analyze more complicated systems. Moreover, some other properties may be useful, for instance, the fact that the patterns in Figures 1–3 should be homotopically interconvertible into one another.

Figure 4 completes the description of the AB_2 system by analyzing the domains for $q_1 = q_2 = qt$, and all the range of values $0 \le \theta \le \pi$.

The study of larger molecules can be performed in a comparable manner. Although the space M will be in general higher dimensional, one may focus the attention only to the changes in the isodensity contours produced by some particular conformational rearrangements of interest. In Figure 5 we study schematically one typical example. We have considered a trisubstituted benzene derivative; two substituents are generic atoms X and the third is the generic group CZ_3 (see Fig. 5). Here we restrict to the stretching between the ring and the group CZ_3, and the internal rotation of the same group, measured by the dihedral angle α. The conventions for the choice of a, b, and $v(b)$ follow those used in Figures 1–4.

We choose here the contour constant a under the assumption that the surface $G(a, \mathbf{R}(t))$ around the benzene ring and the hydrogen atoms show no unexpected curvature features of importance. Moreover, we consider that all noticeable changes occur between the above regions and about the substituents. In particular, the group
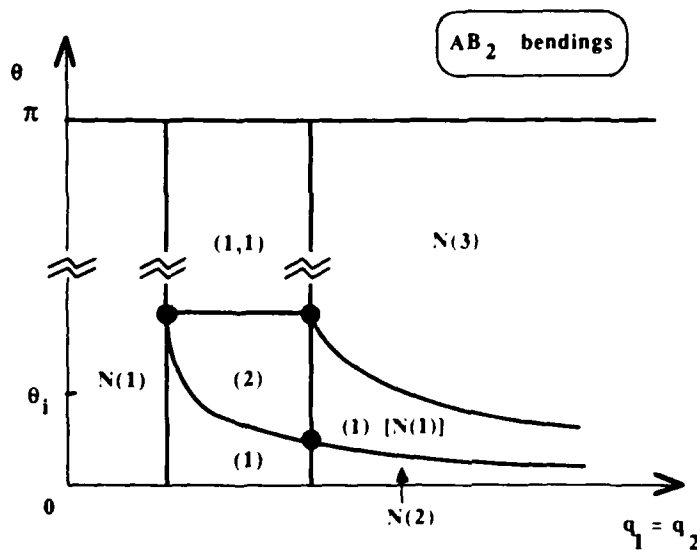

Figure 4. Schematic representation of isodensity surface distribution of shape groups for the triatomic system AB_2. The bending vibrations coupled to symmetric stretchings are displayed. (Truncation type $v(b) = 2$; see text for details. Black dots indicate points where four shape regions meet.)
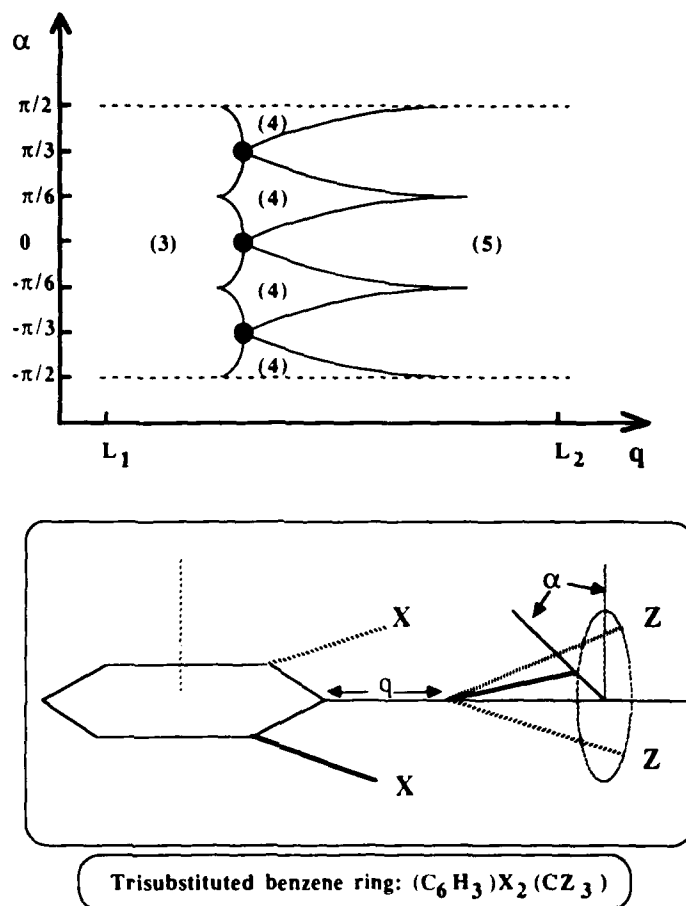
Figure 5. Schematic representation of isodensity surface distribution of shape groups for the trisubstituted benzene ring $(C_6H_3)X_2(CZ_3)$. The configurational rearrangements are restricted to a stretching (coordinate q) and an internal rotation (dihedral angle α). (Truncation type $\nu(b) = 2$; see text for details. Black dots indicate points where four shape regions meet.)

CZ_3 is considered to originate a single D_2 domain. Observe that, in addition, the range of q is bounded to an appropriate interval $[L_1, L_2]$.

Evidently, the periodic nature of angular rotation from $\alpha = \pi/2$ to $\alpha = -\pi/2$ is clearly shown in the diagram of "shape regions." It is worth noting that the transition $(3) \rightarrow (5)$ is critical in the sense that it could only happen at a particular pair of α and q values. This transition involves a change in two generators; notice that, as commented above for the AB_2 system, it is associated to a conformational rearrangement where a symmetry plane is conserved.

Further Comments

The examples given in the preceding section illustrate the possibilities of the method to characterize the shape of molecular conformations. In actual cases of inter-

est, after an appropriate computation of the chosen function $f(\mathbf{r}, \mathbf{R})$ [1–6], the conditions for the selection of parameters a, b, and v can be exploited to provide a very detailed description of the coupling between MS and the nuclear geometry.

The above possibilities in the implementation of the method suggest it for the analysis of the relationships between biochemical activity and the actual changes in the molecular conformations. The procedure can be used to study either shape characteristics, disregarding the molecular size effects, or the characteristics depending simultaneously on size *and* shape. To this end, one may choose the criteria to define the level set $F(a, \mathbf{R}(t))$ from those conditions suggested in the Method section. Furthermore, the method can be easily extended to study other surfaces by using a more detailed partitioning than the one based on curvature. For instance, a molecular van der Waals surface with a superimposed map of ranges of values of electrostatic potential can be treated without introducing significant changes in the formalism.

It is worth reiterating here that above analysis is symmetry independent. Consequently, many features of the relations among "shape regions" in the metric space M should be expected to remain valid, even after symmetry changes. For instance, if one replaces the case AB_2 by a triatomic system ABC, the symmetry-group distribution in M will be totally different. However, many characteristics of the shape regions would be unaltered. This suggests that the homology group analysis would allow using, indirectly, some symmetry information to describe asymmetrical objects. For example, the link can be established by studying systems whose point group contains the group of interest as a subgroup.

Acknowledgment

Bibliography

[1] P. Coppens and M. B. Hall, Eds., *Electron Distributions and the Chemical Bond.* (Plenum, New York and London, 1982).

[2] S. Fliszár, *Charge Distributions and Chemical Effects.* (Springer, New York, 1983).

[3] G. D. Purvis III and C. Culberson, Int. J. Quantum Chem. QBS **13**, 261 (1986).

[4] A. Pullman and B. Pullman, in *Molecular Electrostatic Potentials in Chemistry and Biochemistry,* P. Politzer and D. G. Truhlar, Eds. (Plenum, New York, 1981).

[5] W. Graham Richards, *Quantum Pharmacology,* 2nd. Ed. (Butterworths, London, 1983).

[6] J. C. Culberson, G. D. Purvis III, M. C. Zerner, and B. A. Seiders, Int. J. Quantum Chem. QBS **13**, 267 (1986); J. R. Rabinowitz and S. B. Little, Int. J. Quantum Chem. QBS **13**, 9 (1986).

[7] P. G. Mezey, Int. J. Quantum Chem. QBS **12**, 113 (1986).

[8] P. G. Mezey, J. Comput. Chem. **8**, 462 (1987).

[9] E. H. Spanier, *Algebraic Topology.* (McGraw-Hill, New York, 1966).

[10] W. S. Massey, *Homology and Cohomology Theory.* (Dekker, New York and Basel, 1978).

[11] P. G. Mezey, Int. J. Quantum Chem. QCS **17**, 137 (1983).

[12] P. G. Mezey, Int. J. Quantum Chem. **26**, 983 (1984).

[13] R. F. W. Bader, T.-H. Tang, Y. Tal, and F. W. Biegler-König, J. Am. Chem. Soc. **104**, 940, 946 (1982).

[14] P. G. Mezey, Theor. Chim. Acta **54**, 95 (1980).
[15] P. G. Mezey, Int. J. Quantum Chem. **29**, 333 (1986).
[16] P. G. Mezey, Int. J. Quantum Chem. **QCS 19**, 93 (1986).
[17] A. Streitwieser and P. H. Owens, *Orbital and Electron Density Diagrams: An Application of Computer Graphics*. (Macmillan, New York. 1973).

Molecular Similarity in Aminothiol Radioprotectors: A Randić Graph Approach

D. VASILESCU AND R. VIANI

Laboratoire de Biophysique-Université de Nice 06034 Nice Cedex, France

Abstract

This work is a continuation of our previous experimental and theoretical investigations in the field of aminothiol radioprotectors and anticancer drugs. The best known aminothiols are cysteamine and the natural intracellular radioprotector glutathione (GSH). In this study, we present a tentative discussion of similarity in a class of aminothiols such as: cysteamine, methylated cysteamine, cysteine, AET, WR-1065, WR-2578 S, I-102 S, I-143, penicillamine, and GSH, by using the Randić graph topological method. We have used the Randić graph approach by introducing weighting factors for hetero bonds ($C — O = x$, $C — N = y$, and $C — S = z$) involved in the studied aminothiols. In which case, the paths become polynomials in variables x, y, and z. The similarity is discussed versus (x, y, z) values, using similarity matrices, introduced as a set of euclidean distances between a pair of vectors in the n-dimensional vector space of paths or in the n-dimensional vector space of atomic indexes.
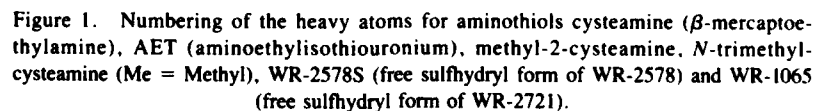
Introduction

In a recent paper, D. H. Rouvray introduces the art to "predicting chemistry from topology" by saying: "At the heart of the new technique is the topology of individual molecules: the pattern of interconnections among each molecule's atoms, which determines the ultimate architecture of the molecule" [1].

Molecular topology is indicated when searching descriptors (atomic or molecular) which are able to give an eventual correlation between them and a particular property (physical, chemical, biological, or pharmaceutical) or which are able to approach the similarity concept in a class of molecules. This way, we have attempted to broach the notion of similarity in the aminothiol molecules. Sulfur-containing molecules, and particularly aminothiols, constitute a very interesting class because of their radioprotector, and in some cases, their anticancer properties [2-7].

Aminothiol radioprotector agents are able to reduce the radiation damage when administered to animals or cellular culture before irradiation with ionizing radiation. They are characterized by their dose reduction factor: DRF = ratio of lethal dose of irradiation for 50% of animals treated with radioprotector to lethal dose of irradiation for 50% of control animals.

Aminothiol anticancer drugs can protect normal tissues and tumors, in a differential manner, in chemotherapy and radiotherapy.

Figures 1 and 2 represent 10 aminothiols which will be studied from the topological point of view. In this class, the naturally occurring molecules are: the aminoacid cysteine (DRF = 1.4) and glutathione (GSH) which is the tripeptide thiol γ-Glu-Cys-

Figure 1.   Numbering of the heavy atoms for aminothiols cysteamine (β-mercaptoe-
thylamine), AET (aminoethylisothiouronium), methyl-2-cysteamine, N-trimethyl-
cysteamine (Me = Methyl), WR-2578S (free sulfhydryl form of WR-2578) and WR-1065
(free sulfhydryl form of WR-2721).

Gly. It is to be noted that GSH is the most abundant thiol inside cells and possesses
multifunctional activities in biology, cancer therapy, and pharmacology [8]. GSH is
also an endogenous intracellular radioprotector [2].

The best known synthetic aminothiol is cysteamine which is directly derived from
cysteine. Cysteamine is a good radioprotector *in vivo* (DRF = 1.6), but introduces an
unacceptable toxicity.

I-102S

I-143

GSH

Cysteine

Figure 2. Numbering of heavy atoms for aminothiols containing a peptide or a pseudopeptide groups: I-102 S (free sulfhydryl form of I-102), I-143, glutathione or GSH (γ-Glu-Cys-Gly) and the aminoacid cysteine.

Other radioprotective drugs synthesized on the basis of cysteamine formula are: methyl-2-cysteamine (DRF = 1.8), N-trimethyl-cysteamine, AET (DRF = 1.6), WR-1065, which is the free sulfhydryl form of WR-2721 (DRF = 2.7), and WR-2578-S (DRF = 2.0) (free sulfhydryl form of WR-2578). WR-2721 is the most effective known radioprotector; it is also an anticancer drug.

Recently, two molecules derived from the γ-Glu-Cys branch of GSH were synthesized by Imbach and co-workers [9, 10]. These new radioprotectors are: I-102 (DRF = 1.4) and I-143 (I-102 S is the free sulfhydryl form of I-102). Generally, the active metabolite of these drugs, is in all cases the free sulfhydryl form, which is able to interact with DNA, and for this reason we have included it in our study.

We have demonstrated previously by using spectrophotometric, dielectric, and nuclear magnetic resonance (NMR) techniques that, *in vitro*, cysteamine, WR-1065, and WR-2578 S interact with DNA. This interaction is essentially electrostatic and involves the cationic sites of radioprotectors and the anionic DNA phosphate sites [11–15]. With the help of quantum chemical computations, we have also demonstrated the possible mechanism of DNA-cysteamine interaction at the molecular level [16] and determined the electrostatic properties (charge distribution and electrostatic potential) of cysteamine, methylated cysteamine, WR-1065, WR-2578 S, and I-102 S [17, 18].

Returning to the Figures 1 and 2, it is obvious that similarities exist between these molecules. Quantitative discussion of molecular similarity, in a class of molecules possessing similar properties, may be conducted using the graph descriptors and indexes introduced by M. Randić [19].

Procedure

The Randić topological descriptors are well described in recent papers [20–23]. Two kinds of topological indexes may be used to qualify each molecule by a graph invariant.

First, the building of the atomic path sequences: $P_O, P_1, \ldots, P_k, \ldots P_n$; in which P_O is the number of heavy atoms (all hydrogen atoms are excluded), P_1 is the number of paths of length 1 (1 bond), P_k is the number of paths of length k (k bonds) [see Tables I and II].

For each atom (i), the atomic index (atomic ID or a_i) is obtained by summing all path numbers on each row (i).

For the molecule we find an invariant signature (total number of paths), at the last row of the molecular graph, by adding the atomic path contributions on each column. Thus the molecule path in the last row is:

$$\Sigma P_O; \qquad \sum_{k \neq 0}^{n} P_k / 2$$

(because each path involves 2 atoms)

Finally, the sum of molecular paths gives the molecular index (molecular ID).

To illustrate this path count, we have represented in Tables I and II the obtained results for WR-1065 and I-102 S. These molecules possess the same molecular index 36, when all bonds are weighted by the factor 1. In this topological description WR-1065 is represented by the signature: 8, 7, 6, 5, 4, 3, 2, 1 and I-102 by the signature: 8, 7, 7, 6, 4, 3, 1.

If we wish to take into account the contributions of heteroatoms such as O, N, and S, it is necessary to associate with each heterobond a weight different from one. We have named the heterobonds: C—$O = x$; C—$N = y$, and C—$S = z$. In this case, the count of path introduces polynomial expressions of variables x, y, and z. Table III shows the obtained molecular graph for I-102 S.

TABLE I   The count of paths in WR-1065 with bond weights: $C—C = 1$; $C—S = 1$; $C—N = 1$.

Atom	Path P_0	P_1	P_2	P_3	P_4	P_5	P_6	P_7	Atomic ID
1	1	1	1	1	1	1	1	1	8
2	1	2	1	1	1	1	1		8
3	1	2	2	1	1	1			8
4	1	2	2	2	1				8
5	1	2	2	2	1				8
6	1	2	2	1	1	1			8
7	1	2	1	1	1	1	1		8
8	1	1	1	1	1	1	1	1	8
	8	7	6	5	4	3	2	1	Molecular ID = 36

TABLE II   The count of paths in I-102 S with bond weights: $C—C = 1$; $C—N = 1$; $C—S = 1$; $C—O = 1$.

Atom	Path P_0	P_1	P_2	P_3	P_4	P_5	P_6	Atomic ID
1	1	1	1	1	1	2	1	8
2	1	2	1	1	2	1		8
3	1	2	2	2	1			8
4	1	2	3	2				8
5	1	3	2	1	1			8
6	1	2	2	1	1	1		8
7	1	1	1	2	1	1	1	8
8	1	1	2	2	1	1		8
	8	7	7	6	4	3	1	Molecular ID = 36

Second, when the sizes of molecules are very different, like GSH and cysteamine for example, it is convenient to introduce a new atomic connectivity index which reduces the contributions of groups far from a studied similar main chain in a lot of molecules.

This index is defined by Randić as a weight $(mn)^{-1/2}$ for each bond with m and n neighbors for the terminal atoms of the bond. Then, a new count of path is built using this $(mn)^{-1/2}$ bond factor.

It is possible also to introduce a supplementary variation in the path count, by adopting different weights for the heterobonds $C—O$, $C—N$, and $C—S$.

Table IV shows the obtained result for the molecule I-102 S, when only the atomic indexes of the six atoms $S_1—C_2—C_3—N_4—C_5—C_6$ of the main chain are taken into account.

TABLE III. The count of paths in 1-102 S with bond weights: $C{-}C = 1$; $C{-}O = x$; $C{-}N = y$; $C{-}S = z$.

Atom \ Path	P_0	P_1	P_2	P_3	P_4	P_5	P_6	Atomic ID
1	1	z	z	yz	y^2z	$xy^2z +$ y^2z y^3	y^3z	$1 + 2z + yz +$ $2y^2z + y^2z + xy^2z$
2	1	$1 + z$	y	y^2	$y^2 + y^2x$			$2 + y + z +$ $2y^2 + y^3 + y^3x$
3	1	$1 + y$	$y^2 + z$	$y^2 + y^2x$	y^3			$2 + y + z + 2y^2$ $+ y^3 + y^3x$
4	1	$2y$	$2y + yx$	$yz + y^2$				$1 + 4y + yx +$ y^2
5	1	$1 + x + y$	$y^2 + y$	y^2	y^2z			$yz + y^2$ $2 + x + 2y +$ $2y^2 + y^2z$
6	1	$1 + y$	$x + y$	y^2	y^2	y^2z		$2 + x + 2y +$ $2y^2 + y^2z$
7	1	y	y	$y^2 + xy$	y^3	y^3	y^3z	$1 + 2y + y^2 +$ $2y^3 + xy + y^3z$
8	8	$2 + x +$ $3y + z$	$x + 3y +$ $z + y^2 +$ xy	$3y^2 + xy$ $+ yz +$ y^2x	$y^2 + y^3 +$ $y^2x + y^2z$	$y^3 + y^2z$ $+ xy^2z$	y^3z	$1 + 2x + 2y$ $+ 2xy^2 + xy^2z$

TABLE IV. Atomic indexes for the 6 atoms in the main chain $S_1—C_2—C_3—N_4—C_5—C_6$ of I-102 S, computed with the $(mn)^{1/2}$ connectivity index and with bond weights: $C—C = 1$; $C—O = x$; $C—N = y$ and $C—S = z$

Atom \ Path	P_0	P_1	P_2	P_3	P_4	P_5	P_6	Atomic ID
1	1	$0.707z$	$0.354z$	$0.177yz$	$0.072y^2z$	$0.042xy^2z +$ $0.029y^2z$	$0.021y^3z$	$1 + 1.061z +$ $0.177yz +$ $0.102y^2z +$ $0.042xy^2z +$ $0.021y^3z$
2	1	$0.5 +$ $0.707z$	$0.25y$	$0.102y^2$	$0.042y^2$ $+ 0.059xy^2$	$0.029y^3$		$1.5 + 0.250y +$ $0.707z + 0.144y^2 +$ $0.059xy^2 +$ $0.029y^3$
3	1	$0.5 + 0.5y$	$0.354z +$ $0.204y^2$	$0.083y^2 +$ $0.118xy^2$	$0.059y^3$			$1.5 + 0.5y +$ $0.354z + 0.287y^2 +$ $0.118xy^2 +$ $0.059y^3$
4	1	$0.908y$	$0.417y +$ $0.236xy$	$0.177yz$ $+ 0.118y^2$				$1 + 1.325y +$ $0.236xy +$ $0.118y^2 +$ $0.177yz$
5	1	$0.408 +$ $0.408y +$ $0.577x$	$0.289y +$ $0.204y^2$	$0.102y^2$	$0.072y^2z$			$1.408 + 0.577x$ $+ 0.697y +$ $0.306y^2 +$ $0.072y^2z$
6	1	$0.408 +$ $0.707y$	$0.236x +$ $0.167y$	$0.083y^2$	$0.042y^2$	$0.029y^2z$		$1.408 + 0.236x$ $+ 0.874y +$ $0.125y^2 +$ $0.029y^2z$

Figure 3.   Euclidean distance $d_{\alpha\beta}$ between two molecules α and β represented by the vectors $\mathbf{V}_\alpha$ and $\mathbf{V}_\beta$. (1) In the n-dimensional vector space of paths $[p_i]$:$p_{i\alpha}$ is the component of $\mathbf{V}_\alpha$ corresponding to a path of length i; $p_{j\beta}$ is the component of $\mathbf{V}_\beta$ corresponding to a path of length j.

$$d_{\alpha\beta} = \{|\mathbf{V}_\alpha - \mathbf{V}_\beta|^2\}^{1/2} = \left\{\sum_{i=0}^{n} |p_{i\alpha} - p_{i\beta}|^2\right\}^{1/2}$$

Observe than if $m < n$, $\mathbf{V}_\beta$ belongs to a subspace. (2) In the n-dimensional vector space of atomic indexes $[\alpha_i]$:$\alpha_{i\alpha}$ is the component of $\mathbf{V}_\alpha$ corresponding to the atomic index of the atom (i) and $\alpha_{i\beta}$ is the atomic index of atom (i) in the molecule β. In this vector space we consider the same number n of atoms in molecules α and β. Euclidean distance is:

$$d_{\alpha\beta} = \left\{\sum_{i=1}^{n} |\alpha_{i\alpha} - \alpha_{i\beta}|^2\right\}^{1/2}$$

Results and Discussion

The quantitative concept of similarity between two molecules α and β, inside a lot of molecules, may be expressed — as indicated by Randić — with the help of similarity matrices. Each element of a similarity matrix represents the Euclidean distance $d_{\alpha\beta}$ between two molecules α and β in a n-dimensional vector space (see Fig. 3). Consequently, two representations are possible: (1) Each molecule is represented by a vector V_α in the n-dimensional vector space of paths $\{P_i\}$. (2) Each molecule α (or a fragment of molecule) is represented by a vector V_α, in the n-dimensional vector space of atomic indexes $\{a_i\}$.

Similarity Matrices in the 8-Dimensional Vector Space of Paths

In a first step, we have chosen to discuss the global approach of similarity between 10 aminothiols of different sizes. The 9 radioprotectors: cysteamine (CYSA). Me-2-cysteamine (ME2C). N-3Me-cysteamine (3MEC). AET, cysteine (CYSE), WR-1065 (1065), WR-2578 S (2578), I-102 S (1102), I-143, and a tenth molecule which is not a radioprotector: the penicillamine (PENI). The shortened names of molecules which will be used in the following tables are in parentheses. Penicillamine ($\beta\beta$ methyl cysteine, see Fig. 4) is obtained by degradation of penicillin. It is a drug of great biological interest and presents different properties with regard to its D and L chiral forms. D-penicillamine is well known as a chelating copper agent and also to decrease the synthesis of collagen; L-penicillamine is a very toxic drug [2].
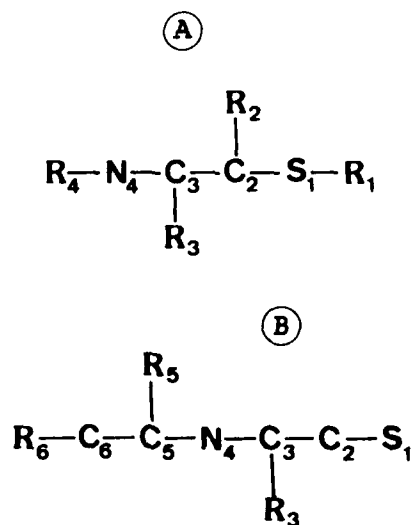

Figure 4. (A) Five molecules based on the cysteamine main chain: $AET:R_1 = -CN_2$; Methyl-2 cysteamine:$R_2 = -C$; N-trimethylcysteamine:$R_4 = -C_1$; cysteine:$R_1 = -CO_2$; and penicillamine:$R_2 = -C_2$ and $R_1 = -CO_2$. (B) Five molecules based on the WR-2578 S main chain:WR-2578 S:$R_6 = -N$; WR-1065:$R_6 = -CN$; I-102 S:$R_5 = -O$ and $R_6 = -N$; I-143:$R_5 = -O$ and $R_6 = -CS$; GSH·$R_1 = -(CO)-NC-CO_2$, $R_5 = -O$ and $R_6 = -C-CN-CO_2$.

We have excluded glutathione from this computation because it contains too great a number of paths. In this space of paths $[P_i]$, WR-1065 and I-143 possess the longest path P_7; the other molecules are represented in subspaces P_3, P_4, P_5, and P_6.

Table V shows the obtained results, in the space of paths, for the 10 molecules studied. Each molecular path is the last row of the count of paths for each amino-thiol, with heterobond weights.

From Table V it is now possible to build similarity matrices, in the 8-dimensional vector space of paths, with empirical values for the set (x, y, z). Starting from the set $(0.4, 0.4, 0.4)$ and arriving to the set $(1.2, 1.2, 1.2)$ we have varied the (x, y, z) values by using a 0.2 step. So, we have obtained 125 similarity matrices with various possibilities for the heterobond weights. As an example, we show in Table VI the obtained similarity matrix for the set $(0.4, 0.8, 0.6)$.

Then, we examined the 10 Euclidean distances $d_{\alpha\beta}$ according to ascending order, starting from the minimal one, in each of the 125 similarity matrices. Table VII shows a sample of obtained results for 4 different set values (x, y, z).

The principal results of this analysis are:

1. The remarkable similarity of the molecule pair WR-1065 and I-102 S. This pair is the only one that appears 125 times in the best 6 $d_{\alpha\beta}$: 46 times first, 40 times second, 18 times third, 14 times fourth, 6 times fifth, and 1 time sixth.

2. The WR-1065 and WR-2578 S couple (which are the most potent radioprotectors) appears 106 times in the best 10 $d_{\alpha\beta}$, but never in first or second rank: 6 times third, 4 times fourth, 24 times fifth, 21 times sixth, 24 times seventh, 13 times eighth, 9 times ninth, and 5 times tenth.

3. The WR-2578 S–AET couple appears 96 times as: 26 times first, 13 times second, 19 times third, 7 times fourth, 8 times fifth, 9 times sixth, 1 time seventh, 8 times eighth, 3 times ninth, and 2 times tenth.

4. The WR-2578 S–I-102 pair placed 98 times in the best 10 $d_{\alpha\beta}$, with only 2 times in first rank.

5. The WR-2578 S–N-trimethyl cysteamine couple appears 82 times with 10 times in first rank.

6. The cysteamine-Me-2-cysteamine pair appears 81 times with 3 times in first rank.

7. The WR-1065–I-143 couple appears 67 times with 8 times in first rank for the following (x, y, z) values: $(0.4, 0.8, 0.4)$; $(0.4, 1, 0.4)$; $(0.4, 1, 0.6)$; $(0.4, 1, 0.8)$; $(0.4, 1.2, 0.6)$; $(0.4, 1.2, 0.8)$; $(0.4, 1.2, 1)$ and $(0.6, 1.2, 0.8)$.

This last result shows that the similarity between I-143 and WR-1065 is reinforced when the aminothiol part is more weighted than the carboxyl one $(y + z)/2 > x$, in all cases).

8. The I-102 S–I-143 couple appears 59 times with 15 times in the first rank for the set values: $(0.4, 1.2, 0.4)$; $(0.6, 1.2, 0.4)$; $(0.6, 1.2, 0.6)$; $(0.8, 1, 0.4)$; $(0.8, 1.2, 0.4)$; $(0.8, 1.2, 0.6)$; $(0.8, 1.2, 0.8)$; $(1, 1, 0.4)$; $(1, 1.2, 0.4)$; $(1, 1.2, 0.6)$; $(1, 1.2, 0.8)$; $(1.2, 1, 0.4)$; $(1.2, 1, 0.6)$; $(1.2, 1.2, 0.4)$; and $(1.2, 1.2, 0.6)$.

TABLE V. The count of paths in studied aminothiols with bond weights: $C-C = 1$; $C-O = x$; $C-N = y$; $C-S = z$.

Path / Molecule	P_0	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Cysteamine	4	$1 + y + z$	$y + z$	yz				
Me-2-Cys	5	$2 + y + z$	$1 + y + 2z$	$y + yz$				
Cysteine	7	$2 + 2x + y + z$	$1 + 2x + 2y + z + x^2$	$2xy + z + 2xy + yz$	$2xz$			
N-3Me-Cys	7	$1 + 4y + z$	$y + z + 6y^2$	$3y^2 + yz$	$3y^2z$			
AET	7	$1 + 3y + 2z$	$y + z + y^2 + z^2 + 2yz$	$z^2 + yz + 2yz^2$	$3yz^2$	$2y^2z^2$		
WR-1065	8	$3 + 3y + z$	$1 + 3y + z + y^2$	$2y + yz + 2y^2$	$3y^2 + y^2z$	$y^2 + y^3 + y^2z$	$y^2z + y^3$	y^3z
WR-2578S	7	$2 + 3y + z$	$3y + z + y^2$	$3y^2 + yz$	$y^2 + y^3 + y^2z$	$y^3 + y^2z$	y^3z	
I-102 S	8	$2 + x + 3y + z$	$x + 3y + z + y^2 + xy$	$3y^2 + xy + yz + y^2x$	$y^2 + y^3 + y^2x + y^2z$	$y^3 + y^2z + xy^2z$	y^3z	
I-143	9	$3 + x + 2y + 2z$	$1 + x + 2y + 2z + y^2 + xy$	$x + y + z + 2y^2 + yz + xy^2 + xy^2$	$2y^2 + yz + xz + xy^2 + y^2z$	$y^2 + 2y^2z + xy^2z$	$2y^2z$	y^2z^2
Penicillamine	9	$4 + 2x + y + z$	$4 + 2x + 2y + 3z + x^2$	$2 + 2x + 2y + z + 2xy + yz$	$4x + 2xz$			

VASILESCU AND VIANI

TABLE VI. Similarity matrix in the 8-dimensional vector space of paths with (x, y, z) = (0.4, 0.8, 0.6).

Molecule	1 CYSA	2 ME2C	3 CYSE	4 3MEC	5 AET	6 1065	7 2578	8 1102	9 1143	10 PENI	
1 CYSA	.000	2.280	4.924	5.875	4.418	7.504	5.265	6.663	8.646	11.136	1
2 ME2C	.000	.000	2.785	3.682	2.559	5.598	3.388	4.705	6.707	8.962	2
3 CYSE	.000	.000	.000	1.412	1.542	3.429	1.714	2.368	4.293	6.416	3
4 3MEC	.000	.000	.000	.000	2.200	2.882	1.909	2.018	3.642	5.513	4
5 AET	.000	.000	.000	.000	.000	3.536	1.394	2.525	4.535	7.442	5
6 1065	.000	.000	.000	.000	.000	.000	2.316	1.235	1.286	5.080	6
7 2578	.000	.000	.000	.000	.000	.000	.000	1.449	3.418	6.538	7
8 1102	.000	.000	.000	.000	.000	.000	.000	.000	2.057	5.367	8
9 1143	.000	.000	.000	.000	.000	.000	.000	.000	.000	4.352	9
10 PENI	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	10

TABLE VII. The ten best euclidean distances in four similarity matrices with different (x, y, z) values.

				(x, y, z) = (0.40, 0.80, 0.60)						
$d_{\alpha\beta}$	1.24	1.29	1.39	1.41	1.45	1.54	1.71	1.91	2.02	2.06
α	1065	1065	AET	CYSE	2578	CYSE	CYSE	3MEC	3MEC	1102
β	1102	1143	2578	3MEC	1102	AET	2578	2578	1102	1143
					(x, y, z) = (0.60, 1.00, 1.00)					
$d_{\alpha\beta}$	1.41	1.60	2.14	2.23	2.46	2.65	2.83	2.87	3.06	3.09
α	AET	1065	AET	2578	1065	CYSA	1065	CYSE	CYSE	1102
β	2578	1102	1102	1102	1143	ME2C	2578	AET	2578	1143
					(x, y, z) = (0.60, 1.20, 0.80)					
$d_{\alpha\beta}$	2.08	2.33	2.38	2.58	2.61	2.80	3.39	3.42	3.90	4.10
α	1065	1065	1102	CYSA	2578	CYSE	1065	AET	ME2C	3MEC
β	1143	1102	1143	ME2C	1102	AET	2578	2578	CYSE	PENI
					(x, y, z) = (0.60, 1.20, 0.40)					
$d_{\alpha\beta}$	1.27	1.99	2.24	2.32	2.55	2.87	3.09	3.23	3.45	3.82
α	1102	1065	1065	CYSA	2578	CYSE	ME2C	1065	2578	ME2C
β	1143	1143	1102	ME2C	1102	AET	AET	2578	1143	CYSE

In this case, the carboxyl and the amino groups are more weighted than the thiol one $(y + x)/2 > z$, in all cases).

9. The N-trimethyl cysteamine–AET couple placed 63 times with 9 times in the first rank.

10. Another noteworthy fact is that the nonradioprotective molecule penicillamine is almost excluded from the 10 best $d\alpha\beta$. This drug appears only 5 times, coupled with the N-3Me-cysteamine, in tenth rank.

Similarity Matrices in the Vector Space of Atomic Indexes

In a second step, we introduced two lots of molecules containing the same basic fragment.

Figure 4 shows a first class of 6 aminothiols derived from the cysteamine main chain: cysteamine, methyl-2-cysteamine, N-trimethyl-cysteamine, AET, cysteine, and penicillamine.

A second class of aminothiols is based on the main chain fragment of WR-2578 S. This lot of molecules contains the most potent radioprotector WR-1065, the new synthesized drugs I-102 S and I-143, and the most important endogenous thiol GSH.

The discussion of similarity between molecules inside these two classes will be conducted by using the $(mn)^{-1/2}$ connectivity index with weighted heterobonds and similarity matrices containing elements computed in the atomic index vector space.

Tables VIII and IX show the obtained results for atomic indexes in the 4-dimensional vector space for the first lot of drugs and in the 6-dimensional vector space for the second one.

Then, we have computed: the 125 similarity matrices for the 6 drugs derived from cysteamine (each matrix contains 15 $d_{\alpha\beta}$) and the 125 similarity matrices for the 5 molecules based on the WR-2578 S fragment (each matrix contains 10 $d_{\alpha\beta}$), with the same sets (x, y, z) described above.

TABLE VIII. Atomic indexes of the 4 atoms in the main chain $S_1 — C_2 — C_3 — N_4$, computed with the $(mn)^{-1/2}$ connectivity index and with bond weights: $C — C = 1$: $C — O = x$; $C — N = y$; $C — S = z$.

Cysteamine	$a_1 = 1 + 1.061z + 0.250yz$
	$a_2 = 1.5 + 0.354y + 0.707z$
	$a_3 = 1.5 + 0.707y + 0.354z$
	$a_4 = 1 + 1.061y + 0.250yz$
Methyl-2-cysteamine	$a_1 = 1 + 1.146z + 0.167yz$
	$a_2 = 1.986 + 0.289y + 0.577z$
	$a_3 = 1.644 + 0.707y + 0.236z$
	$a_4 = 1 + 1.163y + 0.167yz$
N-trimethyl-cysteamine	$a_1 = 1 + 1.061z + 0.125yz + 0.187y^2z$
	$a_2 = 1.5 + 0.177y + 0.707z + 0.265y^2$
	$a_3 = 1.5 + 0.354y + 0.354z + 0.530y^2$
	$a_4 = 1 + 2.030y + 0.125yz$
AET	$a_1 = 1 + 1.158z + 0.648yz$
	$a_2 = 1.5 + 0.354y + 0.5z + 0.204z^2 + 0.236yz^2$
	$a_3 = 1.5 + 0.707y + 0.250z + 0.102z^2 + 0.118yz^2$
	$a_4 = 1 + 1.061y + 0.177yz + 0.072yz^2 + 0.083y^2z^2$
Cysteine	$a_1 = 1 + 1.092z + 0.111xz + 0.167yz$
	$a_2 = 1.544 + 0.157x + 0.236y + 0.707z$
	$a_3 = 1.742 + 0.385x + 0.577y + 0.289z$
	$a_4 = 1 + 1.006y + 0.222xy + 0.167yz$
Penicillamine	$a_1 = 1 + 1.192z + 0.056xz + 0.083yz$
	$a_2 = 2.385 + 0.111x + 0.167y + 0.5z$
	$a_3 = 1.911 + 0.385x + 0.577y + 0.144z$
	$a_4 = 1 + 1.103y + 0.222xy + 0.083yz$

TABLE IX. Atomic Indexes of the 6 atoms in the main chain S_1—C_2—C_3—N_4—C_5—C_6, computed with the $(mn)^{-1/2}$ connectivity index and with bond weights: C—$C = 1$; C—$O = x$; C—$N = y$; C—$S = z$.

WR-2578 S	$a_1 = 1 + 1.061z + 0.177yz + 0.132y^2z + 0.031y^3z$
	$a_2 = 1.5 + 0.250y + 0.707z + 0.187y^2 + 0.044y^3$
	$a_3 = 1.5 + 0.5y + 0.354z + 0.375y^2 + 0.088y^3$
	$a_4 = 1 + 1.5y + 0.177yz + 0.177y^2$
	$a_5 = 1.5 + 0.854y + 0.375y^2 + 0.088y^2z$
	$a_6 = 1.5 + 0.957y + 0.187y^2 + 0.044y^2z$
WR-1065	$a_1 = 1 + 1.061z + 0.177yz + 0.154y^2z + 0.016y^3z$
	$a_2 = 1.5 + 0.250y + 0.707z + 0.218y^2 + 0.022y^3$
	$a_3 = 1.5 + 0.5y + 0.354z + 0.437y^2 + 0.044y^3$
	$a_4 = 1 + 1.625y + 0.177yz + 0.088y^2$
	$a_5 = 1.75 + 0.677y + 0.375y^2 + 0.088y^2z$
	$a_6 = 2 + 0.604y + 0.187y^2 + 0.044y^2z$
I-102 S	$a_1 = 1 + 1.061z + 0.177yz + 0.102y^2z^2 + 0.042xy^2z + 0.021y^3z$
	$a_2 = 1.5 + 0.250y + 0.707z + 0.144y^2 + 0.059xy^2 + 0.029y^3$
	$a_3 = 1.5 + 0.5y + 0.354z + 0.287y^2 + 0.118xy^2 + 0.059y^3$
	$a_4 = 1 + 1.325y + 0.236xy + 0.118y^2 + 0.177yz$
	$a_5 = 1.408 + 0.577x + 0.697y + 0.306y^2 + 0.072y^2z$
	$a_6 = 1.408 + 0.236x + 0.873y + 0.125y^2 + 0.029y^2z$
I-143	$a_1 = 1 + 1.061z + 0.177yz + 0.116y^2z + 0.010y^2z^2 + 0.042xy^2z$
	$a_2 = 1.5 + 0.250y + 0.707z + 0.165y^2 + 0.059xy^2 + 0.015y^2z$
	$a_3 = 1.5 + 0.5y + 0.354z + 0.329y^2 + 0.118xy^2 + 0.029y^2z$
	$a_4 = 1 + 1.408y + 0.236xy + 0.236yz$
	$a_5 = 1.612 + 0.577x + 0.408y + 0.144z + 0.306y^2 + 0.072y^2z$
	$a_6 = 1.908 + 0.236x + 0.167y + 0.354z + 0.125y^2 + 0.029y^2z$
GSH	$a_1 = 1 + 1.092z + 0.056xz + 0.157yz + 0.110y^2z + 0.038xy^2z + 0.002y^3z$
	$a_2 = 1.544 + 0.079x + 0.223y + 0.707z + 0.157y^2 + 0.054xy^2 + 0.003y^3$
	$a_3 = 1.742 + 0.192x + 0.544y + 0.289z + 0.384y^2 + 0.133xy^2 + 0.008y^3$
	$a_4 = 1 + 1.413y + 0.328xy + 0.118yz + 0.076y^2 + 0.039y^3 + 0.013xy^3$
	$a_5 = 1.723 + 0.609x + 0.456y + 0.291y^2 + 0.032xy^2 + 0.023y^3 + 0.048y^2z$
	$\quad + 0.016y^4 + 0.005xy^4$
	$a_6 = 2.180 + 0.314x + 0.285y + 0.118y^2 + 0.013xy^2 + 0.009y^3$
	$\quad + 0.020y^2z + 0.007y^4 + 0.002xy^4$

These obtained results are classified using the best 10 Euclidean distances test (in the case of the second lot of 5 molecules, all the $d_{\alpha\beta}$ are contained in the list). In Table X we have reported the computed classification which contains: the total frequencies of appearance of a pair of molecules the advent of n times at the first, second, ... tenth rank.

This classification enables us to put the discussion in the framework of $(mn)^{-1/2}$ topological descriptor applied to atomic indexes, which is more sensible to proximal substituents R_i than to more distant ones (see Fig. 4).

TABLE X. Frequency of Appearance of pair molecules derived from cysteamine in the 10 best $d_{\alpha\beta}$ in 125 similarity matrices (T is the total frequencies of appearance for the ten ranks).

Pair of molecules \ Rank	T	1	2	3	4	5	6	7	8	9	10
CYSA-ME2C	125	41	46	20	11	6	0	1	0	0	0
CYSA-3MEC	71	0	12	4	13	1	4	6	9	17	5
CYSA-AET	124	57	12	9	8	7	4	5	7	8	7
CYSA-CYSE	125	23	35	10	9	20	5	9	4	6	4
CYSA-PENI	48	0	0	0	0	0	9	1	2	19	17
ME2C-3MEC	47	0	0	0	0	1	5	3	4	10	24
ME2C-AET	125	4	14	33	7	22	19	16	8	2	0
ME2C-CYSE	125	0	0	20	48	23	17	13	1	3	0
ME2C-PENI	107	0	2	10	16	8	7	28	20	9	7
3MEC-AET	43	0	0	4	3	7	3	3	9	9	5
3MEC-CYSE	23	0	0	0	0	0	0	0	0	8	15
3MEC-PENI	0	0	0	0	0	0	0	0	0	0	0
AET-CYSE	125	0	0	14	6	13	43	22	13	9	5
AET-PENI	59	0	0	0	0	0	2	8	5	18	26
CYSE-PENI	103	0	4	1	4	17	7	10	43	7	10

For the first family of molecules derived from cysteamine, it is easy to verify that the most similar couples are: cysteamine–Me-2-cysteamine, cysteamine–AET, and cysteamine–cysteine. Other pairs having a 125 frequency of appearance are Me-2-cysteamine–AET, Me-2-cysteamine–cysteine, and AET–cysteine. This is a very logical result since cysteamine derives directly from cysteine by decarboxylation of this last pair and Me-2-cysteamine is the more isotopologous neighbor of cysteamine.

Another important fact is that Me-2-cysteamine is a good radioprotector (DRF = 1.8 > DRF of cysteamine) with the advantage of reducing the toxic effects of cysteamine.

It is also very interesting to note the exclusion of penicillamine–N-3Me-cysteamine couple.

When observing the obtained classification with regard to the second family of molecules derived from the main chain of WR-2578 S, it is interesting to observe that the more similar couples are: WR-2578 S–WR-1065, WR-2578 S–I-102 S, WR-1065–I-143, I-102 S–I-143 (the best frequency of appearance in first rank), and WR-1065–I-102 S. (Table XI).

These similarities are reasonable with regard to the connectivity index $(mn)^{-1/2}$ properties. Concerning the pairs containing GSH, it is also logical to find a poor similarity in the first ranks of Table X, because of the great size of glutathione.

Nevertheless, it is interesting to note that the GSH–I-143 couple is the more similar with a frequency of appearance of 36 in the first five ranks; this frequency is only 17 for the WR-1065–GSH pair, 6 for the WR-2578 S–GSH pair, and zero for the I-102–GSH pair.

TABLE XI. Frequency of appearance of pair molecules derived from WR-2578 S in 125 similarity matrices.

Pair of molecules / Rank	T	1	2	3	4	5	6	7	8	9	10
2578-1065	125	23	44	13	19	23	2	1	0	0	0
2578-1102	125	25	27	20	22	27	4	0	0	0	0
2578-1143	125	5	4	16	15	9	24	19	17	16	0
2578-GSH	125	0	0	4	1	1	5	6	14	21	73
1065-1102	125	9	15	27	20	12	24	5	5	6	2
1065-1143	125	17	19	16	20	24	18	2	4	2	3
1065-GSH	125	0	0	2	6	9	17	42	31	18	0
1102-1143	125	46	15	11	13	10	13	9	6	1	1
1102-GSH	125	0	0	0	0	0	0	17	28	45	35
1143-GSH	125	0	1	16	9	10	18	24	20	16	11

Conclusion

On the whole of obtained results, we can emphasize the following assertions:

(1) The mathematical topological description introduced by Randić gives a good framework for discussing the similarity in the studied aminothiol class of molecules.

(2) In the global approach of similarity in the vector space of paths, the most important fact is the selection of the WR-1065-I-102 S couple, which are both radioprotective and anticancer drugs. Another interesting point is the exclusion of penicillamine from the lot as a nonradioprotective molecule.

(3) Analysis of the similarity, when using the $(mn)^{-1/2}$ atomic index, leads to logical couples of molecules in the two studied families.

The new synthesized drugs I-102 S and I-143 are associated with the best frequency of appearance in the first rank. It is also notable to observe that for the best similar pair involving glutathione (GSH-I-143) we obtain an association between two molecules with a weak radioprotective effect.

This study shows that when introducing the ideas and the vocabulary of Dubois [24], we can say: If we start with a substructure (cysteamine), we can build a structure (WR-2578 S fragment), then an hyperstructure GSH.

All molecules of these families are isotopologous and isochromatic (same atoms at the nodes) with regard to the main chains described in Figure 4.

Finally, this work suggests new lines of future research associating topology and quantum properties with the aim of new perspectives in drug design.

Acknowledgments

Bibliography

[1] D. H. Rouvray, Sci. Am. **254**, 40 (1986).

[2] P. C. Jocelyn, *Biochemistry of the SH group*. (Academic Press, New York 1972).

[3] W. O. Foye, Int. J. Sulfur Chem. **8**, 161 (1973).

[4] Z. M. Bacq, "Sulfur Containing Radioprotective Agents", *International Encyclopedia of Pharmacology and Therapeutics*. (Pergamon, New York 1975).

[5] J. M. Yuhas, *Radiation-Drug Interactions in the Treatment of Cancer*, G. H. Sokol and R. P. Maickel, Eds. (Wiley, New York, 1980) p. 113.

[6] M. Fatome, Radioprotection **16**, 113 (1981).

[7] O. F. Nygaard and M. G. Simic, Eds., *Radioprotectors and Anticarcinogens*. (Academic Press, New York, 1983).

[8] A. Meister, Science **220**, 472 (1983).

[9] F. Lespinasse, J. Oiry, M. Fatome, P. Ardouin, J. Imbach, E. P. Malaise, and M. Guichard, Int. J. Rad. Oncol. Biol. Phys. **11**, 1035 (1985)

[10] J. Imbach, personnal communication

[11] M. A. Rix-Montel, D. Vasilescu, and H. Sentenac, Studia Biophys. **69**, 209 (1978).

[12] D. Vasilescu and M. A. Rix-Montel, Physiol. Chem. Phys. **12**, 51 (1980).

[13] D. Vasilescu, M. A. Rix-Montel, H. Kranck, H. Broch, and E. Savant-Ros, in *The Molecular Basis of Cancer*, R. Rein Ed. (A. R. Liss, New York, 1985) p. 525.

[14] D. Vasilescu, and G. Mallet, Biopolymers **24**, 1845 (1985).

[15] D. Vasilescu, H. Broch, and M. A. Rix-Montel, J. Mol. Structure (Theochem) **134**, 367 (1986).

[16] H. Broch, D. Cabrol, and D. Vasilescu, Int. J. Quantum Chem., Quantum Biology Symposium **7**, 283 (1980).

[17] H. Broch, D. Cabrol, and D. Vasilescu, Int. J. Quantum Chem., Quantum Biology Symposium **9**, 111 (1982).

[18] H. Broch and D. Vasilescu, Int. J. Quantum Chem., Quantum Biology Symposium **13**, 81 (1986).

[19] M. Randić, J. Am. Chem. Soc. **95**, 6609 (1975).

[20] M. Randić and C. L. Wilkins, Int. J. Quantum Chem., Quantum Biology Symposium **6**, 55 (1979).

[21] M. Randić, Int. J. Quantum Chem., Quantum Biology Symposium **11**, 137 (1984).

[22] M. Randić, in *The Molecular Basis of Cancer*, R. Rein, Ed. (A. R. Liss, New York, 1985) p. 309.

[23] S. C. Grossman, B. Jerman Blažić Džonova and M. Randić, Int. J. Quantum Chem., Quantum Biology Symposium **12**, 123 (1986).

[24] J. E. Dubois, in *Chemical Applications of Graph Theory*, A. T. Balaban, Ed. (Academic Press, New York, 1976) p. 333.

The Water-Membrane Interface as a Substrate for H^+-H^+ Superflow

MICHAEL CONRAD

Departments of Computer Science and Biological Sciences, Wayne State University, Detroit, Michigan 48202 U.S.A.

Abstract

Mobile protons in water bound at the interface with the membrane are analogous in a number of respects to the conduction electrons in a thin layer of metal spread on the surface of a dielectric. The hypothesis analyzed in this paper is that such mobile hydrogen bonds may be paired by propagating electronic oscillations in polar side groups of the membrane. The model is dual to the BCS theory of electronic superconductivity in that (1) the mobile H^+ bonds play the role of conduction electrons and (2) the pairing interaction has its origin in electronic excitations rather than lattice vibrations. The pairing mechanism is similar to that in Little's proposed room temperature organic superconductor, except that it involves transient alterations in ground state energies of polarizable electrons in a colorless membrane rather than polarization of dye-like side groups. Numerical estimates based on formulae applicable to metallic superconductors show that condensation of proton pairs would be feasible in small connected domains on the membrane surface if the water structure is closely packed and the effective mass of the protons sufficiently reduced (though significant disanalogies between the bound water and metal cases make the quantitative applicability of these formulae unlikely). Some of the factors that favor a superfluid transition are high salt concentration, polarizable macromolecules on the membrane that increase the three-dimensionality of the interaction, and the presence of proton donors and acceptors other than oxygen. The dynamic order inherent in a proton superflow could provide the basis for a wide variety of highly ordered motions in biological systems.

Introduction

The manifest spatial and temporal coherence of biological organisms led early investigators in the fields of quantum theory and superconductivity to suggest that low temperature order might play a role in living matter [1, 2]. However attractive these speculations, the difficulty remains that no substance has been found that convincingly exhibits any of the known forms of superfluidity at physiological temperatures. This is connected with the fact that superfluid behavior is based on a Bose condensation of paired particles which individually obey the exclusion principle. All known mechanisms of pairing yield pairing energies too weak relative to kT for condensation to occur.

The pertinent example is the BCS theory of superconductivity. The superfluidity of the electrons is based on an exchange of virtual phonons which derives from the interaction with nuclei in the metal lattice [3–6]. As an electron moves through the lattice it attracts the positively charged nuclei, which in turn attract another electron. This indirect attraction increases as the shift in charge density associated with the nuclear coordinates increases, hence increases as the nuclei are more displaced from

their equilibrium positions. As a consequence, the transition to superfluidity occurs at a temperature proportional to $M^{-1/2}$, where M is the isotopic mass of the lattice nuclei that mediate the pairing interaction. As long as lattice nuclei are the source of the pairing interaction, M is never small enough to accommodate a transition to the superconductive state at physiological temperatures.

Little [7] has proposed that it might be possible to synthesize an organic molecule with superconducting properties at room temperature by replacing the nuclear oscillators in the BCS theory with virtual electronic oscillators. In Little's model, the free electrons are replaced by delocalized electrons that occupy the spine of an organic polymer and the lattice nuclei are replaced by electrons in dye-like side groups. To the extent that the analogy is correct, the isotope effect suggests that the transition temperature could be as much as $(M/m_{el})^{1/2}$ times larger than for a metallic superconductor, where m_{el} is the mass of the electron. This is on the order of 300 times larger, well above room temperature. Nevertheless, despite intense efforts, no room temperature organic superconductor has been synthesized. One problem is that organic polymers are effectively one-dimensional systems (in the absence of stacking), but a superfluid phase transition is only possible in systems with dimensionality greater than two, thus, at minimum in a surface of finite thickness [8-10]. A second problem may be that the dynamics of any two groups of electrons in a single organic molecule are too sensitively dependent on one another for one to induce a pairing interaction in the other.

The purpose of this article is to suggest a new mechanism of superflow behavior in biological systems that is in a sense the dual of the BCS theory. As in the Little model, electronic oscillations provide the coupling mechanism. However, the particles paired are protons in an adjacent aqueous film rather than electrons belonging to the same molecule. The boundary conditions required for this type of interaction are met at water-lipid interfaces, such as occur with all biological membranes. *The ypothesis is that mobile hydrogen bonds in the layer of bound water are paired by propagating electronic oscillations in the polar side groups of the membrane.*

umerical estimates based on this analogy are necessarily highly approximative, especially as the analogy is strained at a number of potentially critical points. At this stage we could probably present the model in a variety of different quantitative regimes. However, as a detailed model is more likely to be clear and useful than a vague one we will adopt the policy of making specific choices among alternative assumptions wherever possible. We note that the oscillations will involve shifts in the ground state energy of a colorless membrane rather than polarization of dye-like side groups.

Bound Water and Metal Compared

The structure of the membrane and the proposed interaction are schematically illustrated in Figures 1 and 2. In this section we briefly review some pertinent facts about the membrane and consider the extent to which hydrogen bonds near the membrane can be treated as a system of weakly interacting fermions analogous to the weakly interacting electrons of a metallic superconductor.

In electron micrographs, biological membranes appear as two high electron density layers spanning approximately 75 Å and sandwiching a layer of low electron density,
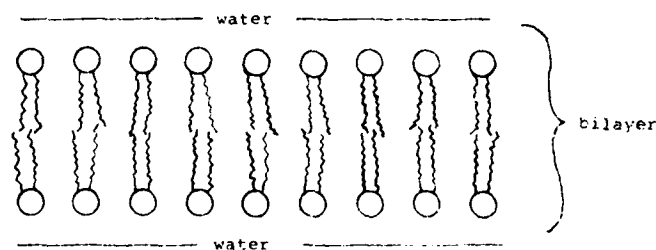
Figure 1.   Lipid bilayer. The ionic and polar groups of phospholipids (represented by circles) are in contact with the layer of bound water. Wavy lines represent fatty acid chains. Surface and integral proteins are not shown [see Ref. 11].
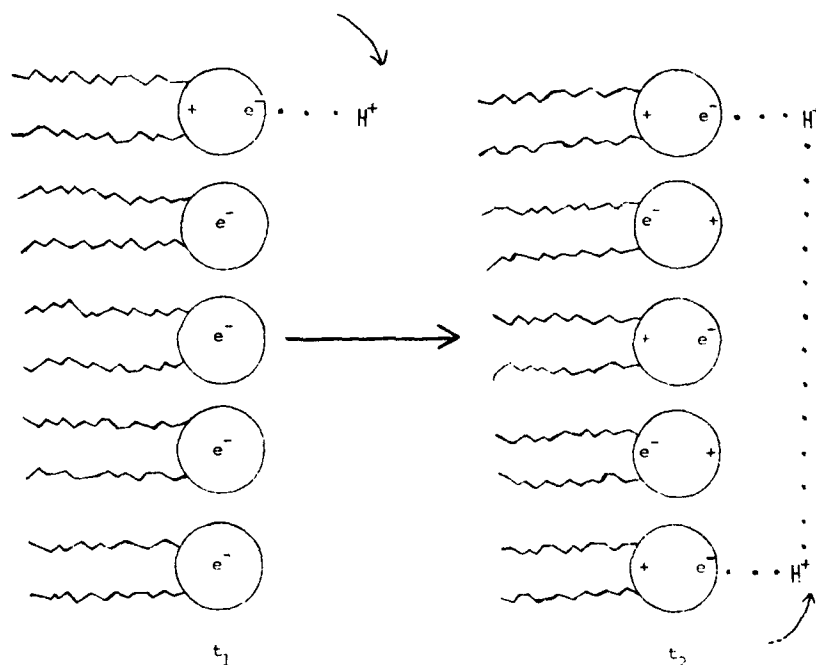

Figure 2.   Schematic diagram of membrane mediated proton-proton pairing interaction. Circles represent polar head groups and wavy lines fatty acid chains belonging to one of the two surfaces of a lipid bilayer. The hopping of a proton in the water layer at time t_1 attracts an electron in a nearby polar group (top left). This initiates an oscillatory wave of electronic polarization in the membrane and at some later time, t_2, a polarized side group attracts a second hopping proton (bottom right). The two protons are attracted to each other by virtue of being attracted to the same mobile charge center.

The model generally accepted has as its essential component a lipid bilayer, with nonpolar groups pointing inward and arrays of polar groups pointing outward into the aqueous medium [11, 12]. A large fraction of this surface is coated with proteins, glycoproteins, and mucopolysaccharides that also have outward-pointing polar groups. Other proteins penetrate into the bilayer or extend through it. The pertinent feature of water are the protons, or mobile hydrogen bonds, that hop between oxygen

atoms. Molecular dynamics studies indicate that, due to cooperative effects of proton acceptance on proton donation, water structures itself into networks of hydrogen bonds with uninterrupted path vays running in all directions from any macroscopic region of the liquid [13]. The structure of bound water is still conjectural. However, nuclear magnetic resonance (NMR) studies indicate that structuring occurs in layers, with layers closest to the membrane having greater organization and a larger fraction of mobile protons [14, 15]. The inner layer probably spans 3 to 5 molecular diameters (ca. 8.4 to ca. 14.0 Å), but how far the influence of the membrane extends altogether is extremely uncertain [16]. The innermost hydrogen bonds extend into the interstices between phospholipid groups, and proton channels consisting of chains of hydrogen bonds are believed to extend across the membrane. These channels, which have given rise to the "proton wire" concept, are probably associated with proteins [17–19], though some authors have viewed them as comprising purely aqueous hydrogen bonds. In either case it is likely that aqueous and nonaqueous hydrogen bonds become mixed at the water–membrane and water–protein interface.

In the hypothesis to be analyzed, free protons hopping through the water lattice at the interface with the membrane attract electrons associated with the polar or ionic groups of phospholipid or with bilayer-associated proteins. When the proton hops again, the residual displacement of the electron exerts an attractive effect on a second proton. Following Ginzburg [20] we shall borrow the term "exciton" for this electronic excitation and tentatively picture it as a transverse wave of polarization propagating through the lattice of polar side groups on the surface of the bilayer. As in the BCS model, the intuitive picture is that particles attracted to the same charge center act as if they are attracted to each other, except that in this case the vibrations of the lattice nuclei are replaced by propagating disturbances in the electronic structure of the membrane and the conduction electrons are replaced by mobile protons in the layer of bound water. When a proton hops, the residual-induced membrane charge should make it more favorable for a neighbor in the chain of hydrogen bonds to hop. Thus it is likely that the interaction between membrane electrons and water protons yields a facilitating effect on the flow of protons that is more local than the pairing interaction, but which is desirable from the standpoint of pairing due to the fact that it enhances the proton wire aspect of the hydrogen bonds near the membrane.

As in Little's discussion of a linear organic molecule, the consideration underlying the "superwater" concept is that, all else constant, the isotope effect could in principle accommodate an increase in the coupling interaction by a factor of $(M/m_{e_1})^{1/2}$, or about 300 times over that in a metallic superconductor. However, since the water-membrane interface is a surface of finite thickness, a basic requirement for the occurrence of a Bose condensation to a coherent state is met. Furthermore, the absence of covalent interactions between the lipid-protein phase and the water phase allows the electronic oscillations in the former to provide the excitonic interaction that couples the protons in the latter without compromising the integrity of normal chemistry, as would likely be the case if all the interactions occurred in a single molecule.

As a consequence of this chemical distinctness, the protons in bound water are in some measure analogous to the conduction electrons in a thin layer of metal spread on the surface of a dielectric [see Refs. 21, 22]. The difference is that the protons in nor-

mal water are localized, their free motion consisting in a diffusive hopping which corresponds to very tight binding and which is probably more analogous to localized electrons that diffuse than to electrons moving relatively freely in a periodic potential. The attraction between H^+ and an oxygen atom is ca. .2 eV, almost one-twentieth of the covalent OH bond. Whether protons behave as pairs, however, depends not on the strength of these bonds, which can be viewed as a modification of the potential in which they move, but on whether the exciton mediated H^+-H^+ bond is sufficiently strong to overcome the Coulomb repulsion and thermal disruption. A second difference is that the ground state of conduction protons in water is not supported by an aufbau structure of lower energy protons as are the conduction electrons at the Fermi level in a metal. The hopping of protons in normal water is a fluctuation phenomenon, rather than a minimum kinetic energy compatible with the exclusion principle. This is connected with the fact that mobile protons are too heavy and occur with too low a concentration in water for Bose–Einstein statistics to be operative at body temperature. In order for the protons to degenerate — that is, be sufficiently delocalized for the wave function to apply to the whole proton system — it would be necessary for their effective mass to be substantially reduced at the water-membrane interface. For these and other reasons, the surrounding interactions that determine whether a coherent state can emerge are very different in water, regardless of its detailed structure in the bound state. than in metals.

The BCS model ignores surrounding interactions, treating a metal as a system of weakly interacting fermions in a square well [23]. The water–membrane interface probably does not lend itself to such abstraction; nevertheless, we will proceed on the basis of the working assumption that major factors determining whether a coherent state can occur are represented in the BCS framework and that it is therefore instructive to consider whether they are more or less favorably met by conduction protons in water or conduction electrons in metal. In making this comparison we will first consider the feasibility of H^+-H^+ bonds occurring at the interface assuming a normal water structure, and then consider how the picture has to be modified in order to satisfy the requirements for Bose–Einstein statistics to be operative. That the bilayer is dressed with a forbidding variety of proteins and carbohydrates will be viewed as a simplifying rather than as a complicating factor, since it is entirely justified to postulate that these macromolecules could be selected in the course of evolution to tailor the interactions at the water-membrane interface to favor a superfluid transition.

Feasibility of Proton Pairing

The matrix element of the interaction energy between two protons can be written as $V = V_{Coul} + V_{ex}$, where V_{Coul} is the Coulomb repulsion and V_{ex} is the exciton-mediated proton-proton attraction. Whether proton pairs will form depends on the strength of the Coulomb repulsion, the strength of the excitonic proton-proton attraction, the relaxation time of proton-proton correlations, the effectiveness with which excitons are created and absorbed due to the interaction of mobile H^+ bonds with membrane electrons, and the constraints imposed on exciton energy by the molecular structure of the membrane.

Coulomb Repulsion

The Debye Hückel formula for the proton pair potential in an aqueous solution may be approximated by

$$V_{Coul} = \frac{Q^2}{Dr} e^{-Kr} \tag{1}$$

where Q is the proton charge, D is the dielectric constant, and r is assumed large in comparison to the proton collision parameter (see, e.g., Refs. 24, 25). The inverse length, K, is related to the ionic strength $1/2 \sum n_i Z_i^2$ by

$$K^2 = \frac{4\pi Q^2}{Dk_bT} \sum n_i Z_i^2 \tag{2}$$

where Z_i is the valence of ion i, n_i the average number per unit volume of this ion, and k_b is Boltzmann's constant. This corresponds to the screening parameter, $\lambda^2 = 4\pi e^2 n_o/k_bT$, for the electron–electron interaction in a metal, except that $\sum n_i Z_i$ is replaced by n_o, the local average density of charge carriers [26].

The Coulomb interaction is attenuated by the high dielectric constant of water ($D = 80$). The main factor, however, is screening. In a good metallic superconductor n_o is typically ca. 10^{23} cm^{-3}, leading to an effective screening distance of ca. 5 to ca. 1 Å (see Ref. 27). Sea water, with a composition very close to that of blood, has a concentration about 5×10^{-1} mol NaCl, giving an effective screening distance of less than 10 Å due to this salt alone. Ionic concentrations inside cells can differ substantially due to ionic pumps, as in the replacement of Na$^+$ by K$^+$ as the chief intracellular anion at the resting potential. Screening is also influenced by Mg^{2+}, substantial and highly variable reservoirs of Ca^{2+}, by H$_3$O$^+$ and OH$^-$ (only 2×10^{-9} mol/L in normal water, but higher in bound water), and by organic anions. The latter include soluble polyelectrolytes, such as soluble proteins, which have a major screening effect due to their high Z value. These must be distinguished from the charged mucoproteins and glycoproteins on the membrane, which represent fixed charges, and which are, therefore, more likely to contribute to mediating the excitonic interaction between proteins than to the screening of the Coulomb repulsion. Finally, the membrane appears to concentrate salts at a level higher than their average intracellular concentration, suggesting that screening would be especially strong at the membrane–water interface (see Ref. 28).

The extent of screening evidently depends sensitively on a number of salts and polyelectrolytes whose concentrations are influenced by intracellular control mechanisms. However, no unusual assumptions are necessary in order to obtain levels of screening sufficient to prevent the Coulomb repulsion from overriding a proton-proton pairing interaction.

Inorganic ions and polyelectrolytes have a second pertinent effect, distinct from screening. Their presence significantly alters the network of hydrogen bonds, modifying the flow of protons and thereby potentially facilitating the pairing interaction.

Proton-proton Interaction

In analogy to the electron-phonon case and following Little [7, 29] and Ginzburg [20, 21] we write the matrix element of the proton-exciton interaction as

$$V_{ex} = -\frac{1}{\hbar}\left\{\frac{2|Mq|^2\omega_{ex}(\mathbf{q})}{\omega_{ex}^2(\mathbf{q}) - \omega_{pr}^2}\right\} \qquad (3)$$

here $\hbar\omega_{ex}(\mathbf{q}) = E_{e1}' - E_{e1}$ is the energy of an exciton of momentum $\hbar\mathbf{q}$, associated with a charge density fluctuation of a membrane electron between energy levels E_{e1}' and E_{e1}; $\hbar\omega_{pr} = E_{pr}' - E_{pr}$ is the shift in proton energy associated with a charge density fluctuation in the proton system from a state of momentum $(\mathbf{p}_1, \mathbf{p}_2)$ to a state of momentum $(\mathbf{p}_1', \mathbf{p}_2')$; and M_q is the matrix element for emission and absorption of excitons that connects the excited states of the membrane to the protons of water. If $\omega_{ex} > \omega_{pr}$ overscreening occurs and the interaction is attractive, the strength of the interaction increasing as the resonant frequency is more closely approached and as M_q increases. A proton-phonon interaction analogous to the conventional electron-phonon interaction should also operate at the membrane. This would be repulsive, but can be ignored since the massiveness of membrane molecular structures means a small ω_{ph}. In this section we consider plausible constraints on each of the terms in Eq. (3).

Proton Motions in Normal Water. A liter of pure liquid water (ca. 55 mol) at 298 K contains about 6×10^{16} proton charge carriers at any given time, each with a mean lifetime of about 10^{-12} s in the nonhopping state [17, 30]. These protons may be thought of as undergoing harmonic oscillations in local wells of a periodic potential, but occasionally hopping (e.g., tunneling) to a neighboring well [31]. If a proton hops into a well, another one hops out, though not necessarily in the same direction. To get a rough estimate of the significance of replacing conduction electrons with protons, we will take the rather naive view that the proton is traveling along a chain of hydrogen bonds and that a local average velocity can be assigned to it; that is, we for the moment ignore the fact that the translational symmetry of the momentum is badly broken by the interaction between H^+ and O^- and that the momentum of the proton is largely transmitted to the whole water-membrane lattice after each hop. This picture of a "proton wire" becomes more useful near the membrane, where the number of proton charge carriers should be significantly increased, and as mentioned earlier has been used to picture the migration of protons across the membrane in the presence of a potential difference.

With this in mind, we write $E_{pr}' - E_{pr} = (p'^2 - p^2)/2m_{pr}$. Substituting $p' = p + \hbar q$ gives $E_{pr}' = (2p\hbar q + \hbar^2 q^2)/2m_{pr}$. Since $m_{pr} \sim 10^3 m_{e1}$, the energy change in the proton system would be reduced by a factor of 10^3 even if p could have the same value as in the electron case. Setting $p = m_{pr}v_{pr}$ gives $E_{pr}' - E_{pr} = v_{pr}\hbar q + \hbar^2 q^2/2m_{pr}$. Since the water molecular diameter is 2.8 Å and since the mobile proton remains associated with a water molecule for only ca. 10^{-12} s, we can take $v_{pr} \sim 2.8 \times 10^4$ cm/s as the effective velocity (recognizing that in reality the proton motion consists of constituent tunneling, rotational, and translational motions and that the hops are not necessarily in the same direction). Alternatively, taking σx in the uncertainty relation $\sigma p_x \sim \hbar/\sigma x$ as the water molecular diameter gives a value of $\sigma v_{pr} \sim 2.2 \times 10^4$ cm/s, which is sufficiently similar to $v_{pr} \sim 2.8 \times 10^4$ cm/s to indicate that tunneling is the enabling motional process. If we take $\sigma x \sim 78$ Å (the average minimum distance that H^+ must traverse when it jumps from one oxygen

atom to another) $\sigma v_{pr} \sim 7.9 \times 10^4$ cm/s, indicating that the maximum velocity is not too different from the average velocity. The difference is due to the fact that the rotation of the water molecule is the rate-limiting step. The NMR-measured value of the rate of proton transfer in the reaction $H_2O + H_3O^+ \rightarrow H_3O^+ + H_2O$ is 10.6×10^9 liter mol^{-1}s^{-1} [30, 32]. Expressed in H$^+$ transfers per second, this corresponds to a velocity $v_{pr} \sim 1.7 \times 10^4$ cm/s, the same order of magnitude as our calculated proton velocities. Since the typical velocity of conduction electrons at the Fermi surface at 0 K is ca. 10^8 cm/s, $v_{el}/v_{pr} \sim 10^4$, implying that $\hbar\omega_{pr}$ is between 10^3 and 10^4 times smaller than $\hbar\omega_{el}$. Even if the proton velocity were substantially enhanced in H-bonded chains at the membrane interface the bracketed term in Eq. (3) is certain to be positive, satisfying the *sine qua non* requirement for pairing to occur. This equation may thus be approximated by $V_{ex} \sim -|M_q^2|/\omega_{ex}(\mathbf{q})$, which emphasizes the fact that V_{ex} will be big enough to overcome the Coulomb repulsion only if the electronic excitation energy, $\hbar\omega_{ex}(\mathbf{q})$, is not too large, otherwise "over" overscreening will occur.

The relaxation time between collisions that alter the direction of an electron in a metal is on the order of 10^{-14} s when defined through conductivity. The time of flight of a mobile H$^+$ between neighboring oxygen atoms is at least on the order of 10^{-13} seconds. Thus there is sufficient time for an exchange of excitons to establish a pairing interaction even though the proton velocity is not well defined over distances greater than one hop in normal water.

Proton-Proton Correlations Between Hopping Events. In order for the proton-proton interaction to be significant the protons must remain paired during the 10^{-12} interval in the nonhopping state. Water is especially well suited to this due to the range of hydrogen bond distortions that are possible. Protons will remain paired during the nonhopping interval if the exchange of excitons during the hopping interval serves to correlate these distortional displacements. Thus when a hopping proton imparts energy to a membrane electron and this attracts another hopping proton, a correlation is established that will persist as a distortional correlation for a short amount of time after the hopping is completed. In effect, the initial distortional state of the H$^+$ bonds are correlated in the nonhopping state by virtue of the exciton exchange during the immediately preceding hopping state. When the protons hop away from the oxygen atoms to which they have become associated there is an increased likelihood that they will do so with some correlation. This will be the case as long as the relaxation time of the distortional correlation exceeds 10^{-12} s. This condition would be less stringent at the water–membrane interface to the extent that the presence of highly specific molecular structures and the mixing of aqueous and non-aqueous hydrogen bonds reduce the time between hops and increase the distortional correlation time.

The mean free path of an electron in a metal is typically 10^3 Å or more. This is obviously very much larger than the mean free path of a proton in water if this is identified with the distance of a hop. However, the mean free path should be identified with the distance over which a correlation can be maintained. In normal water correlations created during the hopping interval would have to be glued together by residual distortional correlations during the nonhopping interval. Later we will see that the proton degeneracy required for condensation would substantially reduce the

irregular character of hopping and thereby simplify the problem of maintaining correlations. If condensation occurs correlations would be frozen in by the requirement that most pairs have zero center of mass momentum.

Proton-Exciton Coupling. The coupling of excitons and protons, symbolized by the matrix element M_q, is favored by proximity since the effect of an induced charge falls off rapidly with distance. This condition is met by the water-membrane interface since the protruding polar side groups are immediately adjacent to and interleaved by the inner network of hydrogen bonds, the electron density reaching its maximum at most 9.0 Å from the exterior of the membrane [see Ref. 28]. Due to its nonpolar character the interior of the membrane is not a good conductor of electricity. Generally, poor conductors make superior superconductors since the particle-hole interaction is stronger. The dielectric character of the interior suggests that membrane electrons attracted toward the aqueous layer by passing protons will be subject to a strong restorative force, yielding a significant particle-hole interaction. Hydration layers more than one or two molecular diameters from the surface of the membrane lose these advantages; however, protruding proteins and carbohydrates could serve to carry the interaction beyond these layers, conferring the important element of three dimensionality.

For the exciton–proton interaction to occur beyond the innermost layers it would have to be transmitted by the water structure. It might be supposed that oxygen valence electrons serve as a substrate for tightly bound excitons (similar to Frenkel excitons). However, this would blur the chemical distinctness of the proton and exciton systems. For the purposes of the discussion that follows we will therefore make the more conservative assumption that the exciton interaction between protons is mediated completely by membrane components and by organic moieties attached to the membrane.

Spectroscopic and Bond Energy Constraints. We are now in a position to consider constraints imposed on the exciton energy by membrane spectroscopy and membrane molecular organization. The following considerations suggest that the "exciton" is a transient alteration in the ground state energy of polarizable membrane electrons with an upper limit of about .2 eV. Such transient ground state shifts do not correspond to optical excitations, but they should be observable as a blurring of the UV spectral lines.

1. *Infeasibility of optical excitations*. In the absence of pigments such as chlorophyll or rhodopsin, membranes appear colorless. Oxidation produces some color and samples consisting of a large number of membrane layers are said to have a yellowish color, possibly due to carotene. The lipid components of the membrane absorb at less than 200 μm, corresponding to electronic excitations of about 6 eV. Unpigmented proteins and carbohydrates on the membrane increase this to 400 μm at most, decreasing the excitation energy to about 4 eV. These values correspond to covalent bond strengths (ca. 2 to ca. 6 eV) and are therefore too high to mediate a pairing interaction, both from the standpoint of over-over screening and from the more critical standpoint of the integrity of the membrane-water lattice. Membrane components and water absorb in the infrared, the latter being opaque to the UV. These excitations must also be excluded since they correspond to vibrational rather than electronic modes.

Even if lateral propagation of excitations corresponding to spectroscopic levels could occur it would do so with a low velocity. The time required for such excitations is ca. 10^{-8} s. Assuming that neighboring phospholipid side groups are separated on the average by the molecular diameter of water (since H^+ can extend into the interstices of these groups) we get a propagation velocity of at most 3 cm/s, as compared to a typical phonon velocity of about 5×10^5 cm/s. Yet a high propagation velocity is especially important for water-membrane superfluidity since the H^+ bonds move so much more slowly than metallic electrons and since screening of the Coulomb interaction is not quite as strong. This argument does not apply to photons whose energy does not equal a difference in the energy levels of the material; such photons are reradiated with a change in phase in a time on the order of 10^{-15} s, giving an exciton velocity that could be as high as 10^7 cm/s.

2. *Ground state energy shifts*. The above arguments exclude spectroscopically observable excitations from a fixed ground state as carriers of the pairing energy. However, these arguments do not exclude transient shifts in the ground state energy of membrane electrons associated with a time dependent potential. The time dependence is initially due to the hopping of an H^+ bond into the vicinity of a membrane electron. This can be viewed as transiently altering the potential in which the electron moves. This initial electronic motion alters the potential of electrons in neighboring membrane moieties, leading to the propagation of the exciton along the membrane. Such shifts in ground state energy occur when salt is added to the bathing solution [33]. Typically the UV spectrum of a molecule shifts by about 5 μm in the presence of salt. This is an average effect. The shifts due to an interaction with an individual H^+ would be much larger, but as fluctuations would only be observable as a slight blurring of the UV spectral lines.

3. *Bond energy and thermal constraints*. It is reasonable to demand that the energy shift be less than covalent bond energies and less than weak bond energies significant for the integrity of the membrane-water lattice. Hydrogen bond energies range from ca. .09 to ca. .4 eV, with the crucial aqueous H^+ bonds being ca. .2 eV. These bonds will re-form if disrupted by an exciton, but too much disruption would burn the overall proton wire structure that we are assuming. Thus it is reasonable to make the tentative assumption that the excitation energy should be less than ca. .2 eV. The hydrogen bond energies and larger electrostatic energies dominate in the polar region of the membrane, whereas van der Waal's forces (.04 eV/particle for CH_3 and .08 for CH_2) and hydrophobic interactions dominate in the nonpolar interior [34]. Van der Waal's interactions and other weak interactions are also critically important for the positioning of proteins in and on the membrane. But it is not necessary to cut the exciton energy off at the energy of van der Waal's interactions, first because the energy here is highly geometry dependent (falling off as r^{-6}), second because large numbers of weak interactions can add up to an interaction significant relative to kT, and third because the geometry-based specificity of these interactions allows the membrane to self-assemble into its original form subsequent to any disruption that would be caused by the exciton.

Life processes generally break down at $313°$ K, three degrees above body temperature. However, some bacteria live in hot springs at $353°$ K, only $20°$ K below the

boiling point of water. The corresponding thermal energy is .03 eV (as compared to .027 eV at body temperature). Exciton energies usually have a lower limit of ca. .1 eV [21], which stands well above the thermal energy at the maximum temperature of life.

Feasibility of a Bose Condensation

Whether proton pairs will condense in momentum space depends on the effective mass, the density of states that can contribute to the condensation process, the stability of the water-membrane lattice to the coupling of protons and excitons, and the operation of the exclusion principle in the condensation process.

Effective Mass and Proton Degeneration

The Bose–Einstein distribution passes over into the Maxwell–Boltzmann distribution if the mean de Broglie wavelength of the particles is small compared to the mean distance between them. The condition for strong degeneracy is met if

$$n \sim \frac{(2\pi m k_b T)^{3/2}}{h^3} \qquad (4)$$

where n is particle density, m is particle mass, and the expression on the right-hand side may be identified with the thermal de Broglie wavelength of particles in an ideal gas [35, 36]. Taking m as the proton rest mass ($m_{pr} = 1.67 \times 10^{-24}$ g) and T as body temperature ($310°$ K) gives a value of $n \sim 10^{24}$ protons per cubic centimeter, which means that the distance between protons would have to be about 1 Å (assuming each proton to be at the center of a small cube). Normal water contains about $.066 \times 10^{24}$ H atoms/cm^3, corresponding to a mean separation of about 2.5 Å. The shortest hydrogen atom separation in water structure is about 1.4 Å on the average, assuming the tetrahedral coordination of water molecules in ice. The situation is comparable in proposed proton wires, such as hydrogen-bonded chains formed from the side chains of proteins and bound water [19]. Hydrogen atom separations would be too large for strong degeneration to occur if hydrogen bond lengths in the range of 2.5–3.5 Å are assumed.

The de Broglie wavelength of a proton moving in a complicated potential of the type that would occur at the water-membrane interface will in general be different than the free proton wavelength. To take this into account let us replace m in Eq. (4) by an effective mass $m^* = (d^2E/dp^2)^{-1}$ of the type that occurs in the simplest band theory of metals. The effective mass of the proton would be increased if it were tightly bound to an oxygen atom since in that case an excitation would be associated with a smaller than expected change in momentum (or, alternatively, a force exerted on it would produce a smaller than expected acceleration). If the proton is only loosely bound to the oxygen atom, the effective mass would be smaller since a small applied force would have a larger effect on the momentum than would be expected.

Let us consider what m_p^* would have to be in order to satisfy Eq. (4), it being recognized that in reality the effective mass would be a tensor quantity and that its use here is purely conceptual given the fact that the acceleration would probably be a

very nonlinear function of the applied force. For a hydrogen atom separation of $d = 2.5$ Å, m_{pr}^* must be about $m_p/7$. For $d = 2.8$ Å, $m_{pr}^* \sim m_{pr}/8$. These estimates assume that each H^+ is at the center of a small cube, as would be the case in a mixed water-ice structure in which water molecules of average diameter 2.8 Å are loosely packed as spheres with six nearest neighbors on the average. Pure ice is loosely packed, with each oxygen packed as a 2.76 Å diameter sphere with four nearest neighbors. The diameter increases slightly in water. But melting of the bonds holding the ice crystal together allows for closer packing, yielding the well known fact that water has a higher density than ice [24]. Extreme close packing, if it could occur, would approximately double the density, reducing the mean separation to ca. 2 Å, thus reducing the required value of m_{pr}^* to about $m_{pr}/4$.

The hypothesis on which the remainder of the paper will be based is that the interaction of water molecules with the membrane surface distorts it enough so as to degenerate the H^+ atoms in small domains on this surface. The concentration of H^+ should be increased by distortion, since weakening of bonds would allow for closer packing. The concentration should also be increased on the inside surface of the cell membrane by fixed negative charges on proteins and by the accumulation of negative ions on the membrane. Distortion of the water structure resulting from interleaving with the polar groups of the membrane surface could reduce the activation barriers to H^+ migration, thereby decreasing the effective mass. Any structure on the membrane that catalyzes the flow of protons would serve to decrease effective mass since it would allow for small fluctuations in energy to have a greater effect on the proton momentum than would be the case for a free proton. The effective mass would thus be further reduced by the integration of channel and wire-like proteins into the already facilitated hydrogen-bonded chains at the water-membrane interface. We will picture the membrane as a network of degenerated patches that comprise a small fraction of its surface but which are tied together by connecting channels and tunnelling processes. This is consistent with the borderline character of proton degeneration and with the fact that degeneration of all the protons on the membrane would interfere with normal chemistry. The time between hops should be shorter for protons with a lower effective mass, and as a consequence the pairing interaction should become stronger under conditions that favor degeneration. But more to the point, the delocalization connected with degeneration is incompatible with a simple hopping picture.

Transition Temperature, Lattice Stability, and Density of States

If these patches and connecting channels were strictly analogous to a weak coupling superconductor we could write

$$T_c = \frac{1.14\hbar\omega_{ex}}{k_b} e^{-1/N(\mu)V} \tag{5}$$

where T_c is the critical temperature and $N(\mu)$ will be identified with the density of states available for contributing to condensation of protons of a given spin. Weak coupling means that $N(\mu)V = N(\mu) (V_{ex} + V_{Coul}) \ll 1$, where $V = |V|$. In the metal case $N(\mu)$ is the density of available states as a function of kinetic energy of electron pairs measured from the Fermi level, but approximated by the density of single spin states at the Fermi level at $T = 0°K$. In the proton case we will take $N(\mu)$ to be the

density of available states as a function of the kinetic energy of the proton pairs, but approximated by the density of single spin states at the chemical potential of the lowest energy unpaired state of mobile hydrogen bonds in the membrane-water system under consideration. As $N(\mu)$ increases so does the number of transitions between momentum states $(\mathbf{p}_1, \mathbf{p}_2)$ and $(\mathbf{p}_1', \mathbf{p}_2')$ of two protons that conserve total momentum and that are therefore compatible with a phase transition. As in the metal case, the maximum occurs when the members of a pair have equal but opposite momenta.

According to Eq. (3) the proton energy $\hbar\omega_{pr}$ as measured from the chemical potential μ (actually the proton energies before and after scattering) must be less than $\hbar\omega_{ex}$. The energy $\hbar\omega_{ex}$ thus enters Eq. (5) as a cutoff of the range of pairing energies under the simplifying assumption that the pairing energy is essentially constant whenever this cutoff is satisfied. The term $\hbar\omega_{ex}$ can be thought of as the width of a potential well in which the protons are interacting, whereas $N(\mu)V$ can be thought of as its depth.

Some models of metallic superconductivity set the upper limit of $N(\mu)V$ as $1/2$ [see for example Ref. 21]. The existence of an upper limit is due to the fact that too strong a coupling between electrons and phonons destabilizes the lattice, thereby leading to lower phonon frequencies. The situation is different in the membrane-water case in that the coupling between protons and excitons occurs in hydrogen bond networks that are distinct from the membrane structures in which the electronic oscillations occur. Furthermore the membrane structures are comparatively massive and the electrons that mediate the excitonic interaction are bound with covalent bond strengths. As a consequence there is no reason to suspect that $N(\mu)V$ is subject to restrictions more stringent or even as stringent as in a metal.

For the moment accepting Eq. (5) as indicative of the constraints on ω_{ex} and $N(\mu)V$ and taking T_c as body temperature, we obtain the numbers in Table 1. If $\hbar\omega_{ex}$ is set at 2 eV, the maximum compatible with covalent bonding, we can get by with a fairly small value of $N(\mu)V$; but this case is incompatible with our assumption that $\hbar\omega_{ex}$ should not exceed hydrogen bond strengths. If $N(\mu)V$ is set equal to 1 (violating the weak coupling condition) $\hbar\omega_{ex} \sim .06$ eV, which is lower than necessary in comparison to hydrogen bond strengths and likely to be masked by vibratory activity at the border of the far infrared. If $N(\mu)V$ is set equal to $1/2$, a possible value in metallic models, we get values of $\hbar\omega_{ex}$ that are slightly less than aqueous hydrogen bond energies, corresponding to a transient shift in the ground state energy of membrane electrons of ca. 1.4×10^3 cm^{-1} to ca. 1.5×10^3 cm^{-1}. These are physically reason-

TABLE I.   Exciton energies that would correspond to different coupling parameters if Eq. (5) were valid for the water-membrane interface (parenthesized value is at 350° K).

$N(\mu)V$	$\hbar\omega_{ex}$ (ev)	λ (Å)
.23	2.00	ca. 6×10^3
.50	.17 (.19)	ca. 7×10^4
.68	.10	ca. 1×10^5
1.00	.06	ca. 2×10^5

able exciton energies in that they satisfy the bond energy and thermal constraints discussed earlier. It is logically possible that the membrane-water lattice would melt at a temperature that would otherwise be below the critical temperature. However, to the extent that any higher value of T_c would increase $\hbar\omega_{ex}$ we can conclude that $N(\mu)V$ cannot be much less than 1/2 if Eq. (4) is, in fact, applicable.

Let us now compare the values of $N(\mu)V$ for conduction electrons to those that might be achieved by H^+ bonds. We will concentrate on $N(\mu)$ since there is no reason to suspect unfavorable limitations on V given the fact that membrane proteins can be tailored for highly specific interactions. The conduction electrons in a metal move in a periodic potential even though they may be treated in a free electron model as an ideal Fermion gas with the energy of each particle given by $E = p^2/2m$. The mobile hydrogen bonds at the water-membrane interface also move in a spatially varying potential, determined by the positions of the oxygen atoms. The difference is that the translational symmetry of the H^+ momentum is badly broken by the ca. .2 eV interaction with oxygen and there are some mutual interactions among the hydrogen bonds that facilitate their motion. For the purposes of a rough comparison we will nevertheless begin with the overidealization that the H^+ bonds can be treated as free Fermi particles, later modifying the picture to ascertain the qualitative effect of the $H^+ - O^-$ and the local H^+, H^+ interactions.

The density of single Fermion states of one spin orientation per unit volume is given by

$$N(E)\,dE = \frac{1}{4\pi^2}\left(\frac{2m}{\alpha\hbar^2}\right)^{3/2} E^{1/2}\,dE \qquad (6)$$

where the energy of the particle ranges from E to $E+dE$ and $E = \alpha p^2/2m$ [37]. For free particles $\alpha = 1$ and for particles subject to a potential α may be defined as $md^2E/dp^2 = m/m^*$. The energy of interest in metals is the chemical potential of the electrons (i.e., the Fermi energy, E_f). Since this is high relative to room temperature (ca. 10 eV for a typical metallic superconductor), $N(E_f)$ is essentially equal to its zero temperature value and is essentially constant over the region of interest. If for the moment we assume that the chemical potential of the mobile H^+ is not too different if measured from the highest or lowest energy nonhopping state, we can identify it with the translational kinetic energy used in the proton wire argument presented earlier. This gives a value of $E_{pr} \sim 4 \times 10^{-4}$ eV, corresponding to the velocity $v_{pr} \sim 2.8 \times 10^4$ cm/s. Setting $\alpha = 1$ gives the ratio $N(\mu_{pr})/N(\mu_{el}) \sim (m_{pr}/m_{el})^{3/2} \times (E_{pr}/E_f)^{1/2} \sim 500$. Idealizing the protons as free particles evidently leads to an increase in the density of states as compared to free electrons, since the increase in mass dominates the decrease in kinetic energy. If $\alpha = 4$, corresponding to an effective mass that could allow for degeneration, and v_{pr} increased accordingly (as discussed in the next section), the density of proton states would still have an approximately 350-fold advantage.

The assumption that E_{pr} is equal to the kinetic energy is equivalent to burying the highest and lowest energy of the unbroken hydrogen bond in the potential in which H^+ is traveling. This corresponds to the fact that the kinetic energy of an H^+ in the field of an oxygen is reduced as compared to that of a free proton of the same total energy. This would appear to increase m^*_{pr} as compared to m_{pr}, which would mean an

increase in the density of states at the expense of the required degeneracy of the protons. This would be the case in normal water. In terms of the naive analogy to the band picture the attraction to O^- can be thought of as compressing the H^+ states into a narrower band, so that in normal water a comparatively long residence time is naturally associated with an increase in the density of states. However, our assumption is that the activation barrier for proton hopping is reduced at the water-membrane interface due to distortion of the water structure and mixing of aqueous and nonaqueous hydrogen bonds. This would expand the H^+ states into a wider band, leading to the decrease in the density of states. The fluctuation energies required to put the proton into this band—hence to put it in a position to be captured by a neighboring oxygen—should be small enough to effectively delocalize the proton. The local H^+, H^+ interaction, insofar as it facilitates the jump to a new O^-, would also have a decreasing effect on the effective mass and on the density of states.

Three further points bear on the applicability of Eq. (5). The first is that this equation results from an integration that assumes the constancy of $N(\mu)$. This assumption cannot be correct for the membrane-water lattice due to the high specific heat of water. Thus the configurational component of the thermal energy of pure liquid water increases at a constant slope of ca. $.3 \times 10^{-2}$ eV/H^+ per degree Kelvin over the whole range from the freezing to the boiling point of water [see Fig. 4.12 in Ref. 30]. This alters the form of Eq. (4), but not the inverse exponential character of the dependence of T_c on $N(\mu)V$. Furthermore, the specific heat would probably be decreased in degenerated regions. The second objection is that we have not considered the potential distortional correlations of H^+ bonds. These should be of minor importance in the case of degenerated protons, but in any case would increase the density of states in the energy range of these distortions. The third point is connected with the concentration of mobile protons. The concentration of conduction electrons in a superconducting metal is typically 10^{22} cm^{-3}, not much larger than the density of states at the Fermi surface (given by $N(\mu_{el}) \sim 10^{21}$ eV^{-1}cm^{-3}). The concentration of protons mobile at any given time in normal water is about 10^9-fold smaller. The proper comparison, however, should be to the concentration of degenerated protons in local patches and channels. This is about 10^{23} H^+/cm^3, an order of magnitude larger than the conduction electron concentration.

If we drop the weak coupling assumption and still wish to adhere to the metal analogy, Eq. (5) can be replaced by phenomenologically more flexible formulae of the McMillan type [22, 38]. The main alteration is that $\lambda = N(\mu)V_{er}$ would be replaced by the renormalized interaction parameter $\lambda/(1+\lambda)$ (with λ redefined). The Coulomb repulsion is reduced in the case of the phonon-electron interaction, with the degree of reduction decreasing with increasing phonon frequency. This effect should be smaller in the case of the exciton-proton interaction due to the larger mass of the proton. The important point, for the present purposes, is that the qualitative dependence of T_c on $N(\mu)V$ remains the same. However, the differences between metal and water are undoubtedly sufficient to disturb the corresponding state aspect of critical temperature equations that have been developed for metals.

Role of the Exclusion Principle

The thermal velocity of a classical gas with particles of proton mass is $v_{pr}^c = (3k_bT/m_{pr})^{1/2} \sim 2.8 \times 10^5$ cm/s. This is an order of magnitude larger than the velocities estimated in the discussion of Proton Motions in Normal Water and, therefore,

consistent with the expected velocity-decreasing effect of the interaction with O^-. The velocity of protons treated as an ideal Fermi gas at $0°K$ would be $(\hbar/m_{pr})$ $(3\pi^2 n)^{1/3} \sim .76 \times 10^2$ cm/s if n is taken as the concentration of mobile protons in pure water at liquid phase temperatures. Thus if mobile H^+ bonds could be viewed as an ideal Fermi gas we would have an inversion of the metal case, the thermal velocity of free electrons in a metal being an order of magnitude less than the zero temperature Fermi velocity. This is why the hopping motion of protons in normal water must be a fluctuation phenomenon rather than a consequence of their being forced into a high minimum energy state by the Pauli exclusion principle, as in the case for free electrons in a metal. The high specific heat of water has its origin in this difference, that is, in the fact that it is the distortion and breakage of hydrogen bonds that allows the configurational contribution to the internal energy of water to be somewhat more than double the vibrational contribution [30].

Degeneration of the protons would alter this picture substantially. Suppose, again as a very rough approximation, that the protons in degenerated patches on the membrane could be treated as an ideal Fermi gas so far as their velocity is concerned. The zero temperature Fermi velocity would then be ca. 3.2×10^5 cm/s, where n is taken as the total concentration of H atoms under the assumption of tight packing and m_{pr} is replaced by $m_{pr}^* = m_{pr}/4$. This is the same order of magnitude as the classical thermal velocity (which would increase to ca. 5.6×10^5 cm/s if we use the effective mass). Though these estimates are unreliable, they do indicate that the flow of protons would be much more influenced by the exclusion principle under our degenerate patch assumption than would be the case in normal water. The flow should become much less irregular, reducing the problem of maintaining proton-proton correlations between hopping events. This is why the specific heat of "superwater" patches and channels should be decreased as compared to normal water. The distinction between mobile and nonmobile proton pools would also become blurred. Since protons are identical particles it is, of course, operationally impossible to determine which pool a particular proton belongs to in any case.

The exclusion principle plays a related role in the water–membrane model which is important and analogous to its role in the metal case. Each electron in the metal must be in a distinct single electron level state. This is true both before and after pairing. However, when pairs form it also becomes necessary to occupy the pair levels. Since ergy of electrons due to phonon exchange is less than the decrease in potential energy due to pairing. The Fermi level is replaced by the chemical potential of a single pool of movable hydrogen bonds in the water–membrane case and it is this level that is unstable to pairing. This pool should be equivalent to the pool of exclusion principle supported conduction electrons so far as the stability of its ground energy level to a pairing interaction is concerned, as long as the increase in kinetic energy of the protons due to exciton exchange is less than the decrease in potential energy due to pairing.

The exclusion principle plays a related role in the water–membrane model which is important and analogous to its role in the metal case. Each electron in the metal must be in a distinct single electron level state. This is true both before and after pairing. However, when pairs form it also becomes necessary to occupy the pair levels. Since

separated by a water molecular diameter. We recall, from our discussion in stereo-
the pairs have integral spin, multiple occupancy is possible and the pairs can con-
dense into the lowest energy pair level. Similarly, if proton pairs form they must oc-
cupy pair levels distinct from the levels occupied by the individual protons. Since the
pairs are bosons they would tend to condense into the lowest lying pair state. As pairs
they have to respect conflicts arising from the Coulomb interaction, but not conflicts
arising from the exclusion principle.

As with electronic superconductors, proton superflows can be thought of in terms
of a two-fluid model in which a condensate of pairs is the true ground state and the
normal or unpaired protons are excitations of this ground state. Above the critical
temperature all the mobile H^+ bonds would exist in the excited states, if in fact a
critical temperature exists below the melting point of the water–membrane lattice.

The singlet state is favored in the case of metal electrons due to the stabilizing ef-
fect of exchange interactions. Such exchange effects would occur in the bound water
case, but in much smaller degree due to the larger mass of the proton. Thus the occur-
rence of triplet-state proton pairs cannot be excluded. The coexistence of singlet and
doublet pairs would yield a condensed state with enriched hydrodynamic behavior.

Pair Separation

The disanalogies between bound water and metal suggest that the condensed state
could differ substantially in a number of respects. The key feature is the spatial range
of the pair wave function. This is large in a metal, typically ca. 10^3 Å. The primary
controlling factor is the electron velocity, this being ca. 10^3 times larger than the
phonon velocity. The electron velocity is restricted to a narrow range, due to the
reasonable requirement that the e^--e^- wavefunction be a superposition of single e^-
wavefunctions within an energy range of the Fermi energy on the order of the energy
gap (see [39]). As a consequence of this restriction the spatial range and energy gap
are roughly connected by the uncertainty formula $(\xi_{el}/v_{el})\Delta_{el}\hbar$, where Δ_{el} is the
energy gap, ξ_{el} is the spatial range, v_{el} is the Fermi velocity, and $(\Delta_{el}/v_{el}) \sim \tau_{el}$ can
be interpreted as the approximate time required for the two electrons to interact with
a phonon.

A similar restriction on proton velocities is not implausible for water due to the
high density of states close to the lowest energy unpaired H^+ state. However, the ex-
citon velocity can be very much higher than the phonon velocity and very much
higher than the effective proton velocity. As a consequence ξ_{pr} is primarily controlled
by v_{ex} rather than v_{pr}, destroying the simple relation to Δ_{pr}. But, to the extent that v_{ex}
is the dominant factor we can take the approximate time for the two protons to inter-
act with the exciton as given by $\tau_{pr} \sim \xi_{pr}/v_{pr}$, subject to the uncertainty principle
constraint $\tau_{ex} \sim \hbar/\omega_{ex}$. Thus, $\xi_{pr} \sim \hbar v_{ex}/\omega_{ex} \sim 35 \times 10^{-16}v_{ex}$. If we set $\xi_{pr} \sim \xi_{el}$
then v_{ex} must be ca. 3×10^9 cm/s, about $1/10$ of the speed of light. This is quite
high, especially as it would require the period of the electronic oscillation to be ca.
$10^{-17}s^{-1}$. If ξ_{pr} is reduced to a value between 10 and 100 Å we obtain the more plausi-
ble oscillation period in the range 10^{-14}–10^{-15} s^{-1}. Since values of ξ_{pr} much less than
10 Å are unlikely due to the Coulomb repulsion, we can set a lower bound of ca.
3×10^6 cm/s on the exciton velocity, assuming side groups on the membrane are

separated by a water molecular diameter. We recall, from our discussion in spectroscopic and Bond Energy Constraints, that 10^{-15}s^{-1} is a typical electronic rearrangement time, giving a maximum velocity of 3×10^7 cm/s.

Though very approximative, this argument does strongly suggest that the spatial extent of the pair wavefunction is at least an order of magnitude smaller in the bound water than in the metal case. If the exciton velocities could become extremely large the pair separation would still be limited by damping. As the salt concentration decreases, the pair separation should increase due to the decreased effect of screening.

Zero Voltage Supercurrents

Biological systems are probably too complicated for observations involving magnetic fields, microwaves, or exotic hydrodynamic behavior to provide more than suggestive evidence for a proton superfluid. However, it might be possible to perform a critical experiment due to the fact that phase coherence of the type that is responsible for the DC Josephson effect is an unambiguous manifestation of the superfluid state. We assume, in analogy to the argument made by Josephson [40], that the proton superfluid is characterized by an order parameter $\psi(r)$ with a definite phase that may be different on the two sides of the insulating membrane. If a large number of degenerated channels (presumably associated with integral proteins) cross the membrane it should be possible for the long-range order of the phase to cross as well. According to the Ginzburg-Landau equation for the free energy density of a superconductor, the free energy of the channels should then increase the free energy of the system as a result of the variation in the phase by $1/2m|-i\hbar\nabla\psi|^2$ (omitting, however, the dependence on the magnetic vector potential). If the voltage is clamped at a zero potential difference, current should flow to reduce this free energy and would continue to flow if the number of protons on each side were maintained constant by being circuited with the help of a battery.

This effect, though definitive, could be extremely small due to the fact that the total number of degenerated H$^+$ atoms on the membrane must be small and due to the fact that the number density of paired charge carriers, $n_s(r) = |\psi(r)|^2$, might comprise a small fraction of these. The effect would also be smaller for protons than electrons since the free energy of the channel current density decreases with increasing mass. Simple tunneling could not be the correct picture in normal water since protons hop across the membrane through a sequence of tunneling events facilitated by the proton channel in this case. When degeneration sets in, however, the analogy to electrons tunneling across a thin barrier should be more applicable.

A small AC Josephson effect could also occur. A constant potential across a Josephson junction induces a supercurrent with frequency $\omega_j = 2QV/\hbar$, provided no other mechanism of dissipation is available. V is typically on the order of millivolts, corresponding to emitted radiation in the short wavelength microwave range [see Ref. 41]. The resting potential of most cells is in the range -10 to -100 mV, with some cells going up to -200 mV [28]. It is possible that an AC type Josephson effect is responsible for the high voltage sensitivity of channel proteins in the membrane that serve as Na$^+$ and K$^+$ gates [see Ref. 42]. For nerve and muscle cells the resting potential is typically -60 to -70 mV. The threshold for firing is 15–20 mV less negative than this, with Na$^+$ channel proteins beginning to respond when the depo-

larization exceeds 7 mV. It is therefore conceivable that the voltage sensor presumed to control Na$^+$ and K$^+$ gates responds to ω, rather than to the field itself, it being recognized that the current is statistically less well defined than in the metal case. Alternatively, in the presence of a suitably adjusted dissipative process the depolarization could induce a DC proton current that reorients the gates to an open position. In the absence of such an alternative mode of dissipation the current oscillations would produce radiation in the short wavelength microwave range, the wavelengths increasing as other modes of dissipation are introduced. This may be connected to the disorienting effects of microwave radiation and a functional role for such membrane potential governed current oscillations should not be excluded. These types of effects, in which phase order crosses a membrane, would also confer a high sensitivity to magnetic fields.

Biological Implications

The primary constituent of biological tissues — from 60 to 90% — is water and a significant fraction of this is in bound form. The lipid bilayer is particularly important because of its ubiquity as a cellular and intracellular structure. However, the basic elements of the proposed interaction would be present wherever bound water occurs, that is, wherever a nonpolar moiety is attached to a polar moiety, which in turn attracts a film of water. This is the case with free proteins, which bind water in the form of a sphere of hydration, and in networks of protein structures (microtubules and microfilaments) that form the membrane and cellular cytoskeletons. Hydrogen bonds contribute to the conformation of nearly all macromolecules in the cell and contribute to the mechanism of action in many cases. The pairing interaction extends with only minor modifications to these nonaqueous hydrogen bonds. Many of these undoubtedly exchange with mobile protons of the aqueous hydrogen bond network. Yet a small superproton fraction in this elaborate network could easily be hidden by the fact that the tools traditionally used for establishing the existence of a superfluid state have complex functional effects in biological organisms.

Standing above this ambiguity, however, is the manifest coherence of biological materials. The question is whether this coherence is based in part on a physically coherent substrate, such as a proton superfluid. Except for zero-voltage supercurrents, which are likely to be very small, it is probably impossible to separate the existence of a proton superfluid from its complex functional roles, and it is therefore pertinent to consider how a superflow could interface with known mechanisms of biochemistry to contribute in a distinctive manner to biological structure and function. We consider phenomena at three levels of organization.

1. *Coordination of intracellular events*. According to Mitchell's chemiosmotic hypothesis proton flow across the membrane is a key contributor to energy transduction processes such as photosynthesis and oxidative phosphorylation [43]. Proton superflow could serve to provide a dynamic skeleton that coordinates proton movements and associated electron movements in energy transfer processes even if most of the protons involved were unpaired most of the time. Such a skeleton would also serve to order the motion of other ions and could contribute to persistent and sometimes coordinated motions exhibited by both large and small molecules. A partial list of such

phenomena includes persistent conformational motions of membrane proteins, translational motions of proteins associated with membrane fluidity, motions of nucleic acids and protein enzymes between sites of action in protein biosynthesis, relative motions of enzyme and substrate in the recognition process, cytoplasmic streaming, and axoplasmic transport. If such motions are controlled in part by the dynamic order of an underlying proton superflow they could be significantly modified by the production of proteins that tailor the pairing interaction differently, allowing for very different dynamic organizations in different types of cells or in different phases of the cell cycle.

2. *Cellular pattern recognition*. Dynamic order may be required to fully explain the ability of cells to respond appropriately to conditions and objects in their environment. This capability reaches highly specialized forms in CNS neurons, which must respond to patterns of transmitter and mediator input, and in immune system cells, which must produce antibody or take other actions in response to the pattern of antigen and hormones impinging on their external membrane. The surfaces of all cells are dotted with receptors that respond to specific molecular shapes, largely through a "lock-key" type mechanism involving r^{-6} van der Waal's interactions. They may also respond to local physical perturbations, such as stretching. Some of these receptors produce internal messenger molecules, in particular cyclic nucleotides and calcium, which in turn can trigger events that modify proteins on the membrane and in the cytoskeleton [44]. The recognition of patterns distributed over significant regions of space or time probably involves elaborate transmission and processing of signals from the different receptors. The problem of pooling all this information would be greatly reduced and the recognition power greatly increased by a superfluid network of mobile H^+ bonds. The rigid aspect of such a network would allow it to respond to patterns of membrane perturbation and receptor activity over a much larger region of space than could any single macromolecule, in effect extending the applicability of the lock-key metaphor to macroscopic dimensions.

The phase coherence of proton superflow could serve as an initial screening mechanism for distinguishing self from not self if the not-self entity (say a virus or another cell) is covered by a layer of bound water. Proton flows accompanying the disappearance of the phase difference between the two superflows subsequent to contact could serve to alert the specific molecular and cellular mechanisms already known to play a role in rejecting undesirable foreign objects.

3. *Multicellular pattern formation*. All cells in a multicellular organism must have positional information in order to differentiate and move in a manner that leads to the development and maintenance of proper three-dimensional form. Cellular pattern recognition contributes to this insofar as it enables cells to ascertain their location in the organism from local chemical influences impinging on them. In addition, close junctions between membranes of different cells (in particular desmosomes) could serve to extend proton superflows into the intercellular networks, thereby lending stability, integrity, and unity to the organism.

It is not difficult to extend the list of interface mechanisms mentioned above. Our discussion is intended to illustrate how proton superflow can be used to construct "order-from-order" models of biological phenomena. The principle of order from

order was first proposed by Schrödinger [1] in his discussion of the analogy between life processes and low temperature phenomena. It is possible to construct models that exhibit all of the phenomena mentioned in some degree without invoking an underlying superflow. The dynamic order in such models is based on the dissipation of energy in the presence of suitable nonlinearities. It is undoubtedly the case that such order through dissipation is of profound importance in biology [see Ref. 45]. In the proton superflow model energy dissipation is required to create the macromolecules that comprise the membrane. These molecules self-assemble into a structurally ordered membrane in the presence of water through a free energy minimization that is dominated by potential energy, in the fashion of crystalline order. The importance of such potential energy-based order in biological materials is due to the fact that macromolecules are big enough to specifically stick to one another and small enough to interact by means of diffusion. If the pairing mechanism proposed in this paper is operative, the dynamic order of proton superflow is inherent in the structural order of the water-membrane interface. No extra energy cost is required. As would be expected, models in which some of the observable order arises from the order in an underlying superflow would thus have enhanced capabilities as compared to models in which all of the dynamic order requires an energy expenditure over and above that required to produce the structural order.

Acknowledgment

Bibliography

[1] E. Schrödinger, *What is Life?* (Macmillan, New York, 1945).

[2] F. London, *Superfluids*, vol. 1, (John Wiley and Sons, New York, 1950).

[3] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, (5), 1175 (1957).

[4] A. D. C. Grassie, *The Superconducting State*, (Sussex University Press, Sussex England, 1975).

[5] E. A. Lynton, *Superconductivity*, (Methuen and Co., London, 1964).

[6] J. R. Schrieffer, *Theory of Superconductivity*, (Benjamin/Cummings, Reading, MA, 1983).

[7] W. A. Little, Phys. Rev. **134**, (6A), A1416 (1964).

[8] P. C. Hohenberg, Phys. Rev. **158**, 383 (1967).

[9] N. D. Mermin and H. Wagner, Phys. Rev. Letters **17**, 1133 (1966).

[10] J. S. Langer, J. Polymer Sci., Part C., **29**, 87 (1970).

[11] S. J. Singer and G. L. Nicolson, Science **175** (4023), 720 (1980).

[12] A. G. Lee, Endeavor, **34**, 67 (1975).

[13] F. H. Stillinger, Science **209** (4455), 451 (1980).

[14] F. W. Cope, Biophys. J. **9**, 303 (1969).

[15] R. Mathur-De Vre, Prog. Biophys. Molec. Biol. **35**, 103 (1979).

[16] J. S. Clegg, in *Cell Associated Water*, W. Drost-Hansen and J. S. Clegg, Eds., (Academic Press, New York, 1979).

[17] M. Eigen and L. De Maeyer, Proc. Roy. Soc. (London) A**247**, 505 (1958).

[18] M. S. Chen, L. Onsager, J. Bonner, and J. Nagle, J. Chem. Phys. **60**, 405 (1974).

[19] J. F. Nagle and Tristram Nagle, J. Membrane Biol. **74**, 1 (1983).

[20] V. L. Ginzburg, J. Polymer Sci.: Part C, **29**, 3 (1970).

[21] V. L. Ginzburg, Contemp. Phys. **9** (4), 355 (1968).

[22] V. L. Ginzburg and D. A. Kirzhnits, Eds., *High Temperature Superconductivity*, (Consultants Bureau, New York and London, 1982).

[23] L. N. Cooper, in *Theory of Superconductivity*, J. R. Schrieffer, Ed. (Benjamin/Cummings Publ. Co., Reading, MA, 1964), pp. 279–299.

[24] J. T. Edsall and J. Wyman, *Biophysical Chemistry*, vol. 1, (Academic Press, New York, 1958).

[25] R. C. Setlow and E. C. Pollard, *Molecular Biophysics*, (Addison-Wesley, Reading, MA, 1962).

[26] J. M. Ziman, *Principles of the Theory of Solids*, (Cambridge University Press, Cambridge, England, 1964).

[27] C. Kittel, *Introduction to Solid State Physics*, (John Wiley and Sons, New York, 1963).

[28] A. C. Giese, *Cell Physiology*, 5th Ed. (W. B. Saunders Co., Philadelphia, 1979).

[29] W. A. Little, J. Polymer Sci.: Part C, **29**, 17 (1970).

[30] D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water*, (Oxford University Press, New York and Oxford, 1969).

[31] J. Brickmann and H. Zimmermann, J. Chem. Phys. **50**, 1608 (1979).

[32] S. Meiboom, J. Chem. Phys. **34**, 375 (1961).

[33] D. Hadzi, in *Spectroscopy of Biological Molecules*, C. Sandorfy and T. Theophanides, Eds. (Reidel Publ. Co., Dordrecht, The Netherlands, 1984), pp. 39–60.

[34] H. Hauser, in *Water: A Comprehensive Treatise*, Vol. 4: *Aqueous Solutions of Amphiphiles and Macromolecules*, F. Franks, Ed. (Plenum Press, New York, 1975), pp. 209–303

[35] R. H. Fowler and E. A. Guggenheim, *Statistical Thermodynamics*, (Cambridge University Press, Cambridge, England, 1939).

[36] E. S. R. Gopal, *Statistical Mechanics and Properties of Matter*, (Ellis Horwood Limited, Chichester, England, 1974).

[37] N. F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys*, (Dover Publications, Inc., New York, 1958).

[38] W. L. McMillan, Phys. Rev. **167**, 331 (1968).

[39] N. W. Ashcroft, and N. D. Mermin, *Solid State Physics*, (Saunders College, Philadelphia, 1976).

[40] B. D. Josephson, Advances in Physics (Philosophical Magazine Suppl.) **14** (56), 419 (1965).

[41] A. C. Rose-Innes and E. H. Rhoderick, *Introduction to Superconductivity*, (Pergamon Press, London, 1969).

[42] B. Hille, BioSystems **8**, 195 (1977).

[43] P. Mitchell, Glynn Res. Lab., Bodmin, Cornwall, England (1966).

[44] M. Conrad, BioSystems **16** (3–4), 345 (1984).

[45] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems*. (Wiley-Interscience, New York, 1977).

Studies on Proton Transfers in Water Clusters and DNA Base Pairs

YOUNG SHIK KONG AND MU SHIK JHON

*Department of Chemistry, Korea Advanced Institute of Science and Technology, P.O. Box 150,
Cheong Ryang Ri, Seoul, Korea*

PER-OLOV LÖWDIN

*Quantum Theory Project, Department of Chemistry and Physics, University of Florida, Gainesville,
Florida 32611 U.S.A.*

Abstract

Studies on the proton transfers in water clusters $(H_2O)_n$ $(n = 2, 3, 4, 5, 6)$, and in DNA base pairs, adenine-thymine base pair, and guanine-cytosine base pair, have been done by using the potential functions of polarization model for water and *ab initio* SCF method with STO-3G basis set for base pairs, respectively.

Introduction

There are many studies on proton transfers in a wide array of chemical and biological processes [1] due to the prevalence of hydrogen bonding [2] in chemical and biological systems. Some studies on proton transfer between water molecules have been done by *ab initio* [3–13] or semiempirical methods [14–17]. Studies on multiple hydrogen bonding systems as well as single hydrogen bonding systems are also important to explain a number of biologically important phenomena, such as the base pair interactions in DNA which are essential to understand the character of the genetic code [18, 19]. Löwdin [20–22] has suggested that hydrogen-bonded protons can be transferred from one base of DNA to its complement and that proton tunneling may play a key role in mutagenesis. Namely, in the mechanism of cell duplication, the hydrogen bonds in the double helix are released, two DNA strands become at least partly free, and each one produces by means of enzymes and available material its own complement so that two complete DNA molecules with the same base sequence as the original one are produced. The complementary nature of the DNA molecule explains the stability of genetic message and how it is propagated at a cell duplication. According to Löwdin's work, there is a possible error mechanism in the genetic code replication as a consequence of tautomer formation. The proton-electron pair codes of the normal and rare tautomeric nucleotide bases are shown in Figure 1. Therefore, potential energy profiles for the proton transfers in DNA base pairs are important to understand the possibility that mutational lesions in DNA can be caused.

In this study, we have studied proton transfers in water clusters with optimum geometries, $(H_2O)_n$ $(n = 2, 3, 4, 5, 6)$, using the potential functions of polarization model for water [23–24] because studies which have been done until now have been

CCC 0360-8832/87/010189-21$04.00

$$A \begin{vmatrix} :H & : \\ : & H: \\ & : \end{vmatrix} T \qquad\qquad G \begin{vmatrix} : & H: \\ :H & : \\ :H & : \end{vmatrix} C$$

(a)

$$A^\star \begin{vmatrix} : & H: \\ :H & : \\ & : \end{vmatrix} T^\star \qquad\qquad G^\star \begin{vmatrix} :H & : \\ : & H: \\ :H & : \end{vmatrix} C^\star$$

(b)

$$A^\star \begin{vmatrix} : & H: \\ :H & : \\ & : \end{vmatrix} C \quad G \begin{vmatrix} : & H: \\ :H & : \\ :H & : \end{vmatrix} T^\star \quad G^\star \begin{vmatrix} :H & : \\ : & H: \\ :H & : \end{vmatrix} T \quad A \begin{vmatrix} :H & : \\ : & H: \\ & : \end{vmatrix} C^\star$$

(c)
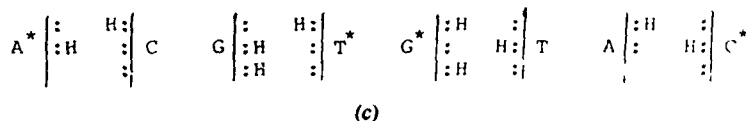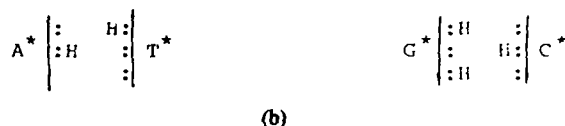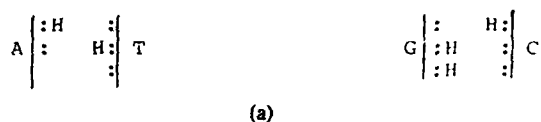
Figure 1.   The proton-electron pair code of (a) normal, (b) rare tautomeric, (c) normal and
rare tautomeric nucleotide bases.

concerned principally with water dimer and linear chain of water polymer. We have
also calculated the potential energies by *ab initio* SCF method with STO-3G basis set
and obtained the potential energy profiles for proton transfers in DNA base pairs,
adenine-thymine base pair, and guanine-cytosine base pair. In addition, the possibil-
ity of Löwdin's error mechanism in genetic code replications has been discussed.

Methods of Calculations

Water Clusters

The polarization model is used to calculate potential energies for proton transfer re-
actions in water clusters, $(H_2O)_n$ $(n = 2, 3, 4, 5, 6)$. The interaction potential for the
polarization model consists of two parts

$$\Phi = \Phi_I + \Phi_{II} \tag{1}$$

In Eq. (1), the first term is composed of a sum of potentials for each pair of particles
in the system

$$\Phi_I = \sum_{i<j=1}^{N_H} \phi_{HH}(r_{ij}) + \sum_{i=1}^{N_H} \sum_{j=1}^{N_O} \phi_{OH}(r_{ij}) + \sum_{i<j=1}^{N_O} \phi_{OO}(r_{ij}) \tag{2}$$

The second term is a nonadditive potential, whose form is suggested by classical
electrostatics for polarizable particles. Upon using Å as the distance unit and kcal/
mol as the energy unit, three distinct atom pair functions which are comprised in Φ_I
have the following forms

$$\phi_{HH}(r) = 332.1669/r$$

$$\phi_{OH}(r) = (332.1669/r)[10 \exp(-3.69939282r) - 2]$$
$$+ [-184.6966743(r - r_e) + 123.9762188(r - r_e)^2]$$
$$\times \exp[-16(r - r_e)^2]$$

$$\phi_{OO}(r) = 1328.6676/r + 8.255 \exp[-18.665(r - 2.45)]$$
$$+ 84.293/(1 + \exp[2.778(r - 2.56)])$$
$$- 12.299/(1 + \exp[4.817(r - 3.10)]) \tag{3}$$

Here, r_e stands for the equilibrium bond length in water (0.9584 Å). The polarization interaction requires self-consistent calculation of induced dipole moments on each of the oxygen particles. The moment on particle 1 is determined by ($\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_l$)

$$\boldsymbol{\mu}_l = -\alpha \sum_{j(\neq l)} (\mathbf{r}_{ij} q_j/r_{ij}^3)[1 - K_0(r_{ij})]$$
$$-\alpha \sum_{m(\neq l)} (\mathbf{T}_{lm} \cdot \boldsymbol{\mu}_m/r_{lm}^3)[1 - K_0(r_{lm})] \tag{4}$$

Here, $\alpha = 1.444$ A^3 is the oxygen polarizability. The first sum covers all charges q_j (excluding that on l) while the second sum includes all other dipoles $\boldsymbol{\mu}_m$. Also, dyadic tensor $\mathbf{T}_{lm}$ is defined by

$$\mathbf{T}_{lm} = 1 - 3\mathbf{r}_{lm}\mathbf{r}_{lm}/r_{lm}^2 \tag{5}$$

The dimensionless factors $1 - K_0$ account for spatial extension of the polarizable electron cloud about each oxygen. Specifically, we have

$$1 - K_0(r) = r^3/[r^3 + F(r)]$$
$$F(r) = 1.855785223(r - r_e)^2 \exp[-8(r - r_e)^2]$$
$$+ 16.95145727 \exp(-2.702563425r) \tag{6}$$

Once the induced moments have been obtained, the polarization energy can be computed as a sum of modified charge–dipole interactions

$$\Phi_{11} = (1/2) \sum_{\substack{j,l \\ (j \neq l)}} ((\boldsymbol{\mu}_l \cdot \mathbf{r}_{ij})q_j/r_{ij}^3)[1 - L_0(r_{ij})] \tag{7}$$

The dimensionless factors $1 - L_0$ also could account for electron extension with

$$1 - L_0(r) = 1 - \exp(-3.169888166r)[1 + 3.169888166r + 5.024095492r^2$$
$$- 17.99599078r^3 + 23.92285r^4]$$

The details about the polarization model are given in the literature [23, 24].

Potential energy profiles for proton transfer reaction in water clusters, $(H_2O)_n$ ($n = 2, 3, 4, 5, 6$), are obtained from the following procedures. The optimum

geometries of water clusters are determined using the potential functions of polarization model and the minimization method [25]. Dimer is linear and trimer is cyclic-like. Tetramer, pentamer, and hexamer are cyclic. These structures of water clusters are consistent with other results [26]. Among several ROHs which are the bond lengths between hydrogen-donor oxygen atom and transferred hydrogen atom, one ROH is determined as a reaction coordinate (of course, in dimer, there is only one ROH). ROH is increased or decreased by a certain magnitude (usually 0.1 Å) and other atoms, except one oxygen atom and one hydrogen atom which are fixed as a reaction coordinate, are optimized to obtain the minimum energy geometries. In other studies, the positions of protons in simultaneous proton transfer are determined artificially, while in our study, the positions of protons move freely except one hydrogen atom which is fixed as a reaction coordinate.

DNA Base Pairs

Ab initio SCF calculations with STO-3G basis set have been performed to obtain the potential energy profiles for the proton transfers in DNA base pairs, adenine-thymine base pair, and guanine-cytosine base pair, respectively. All calculations in this study have been done on an IBM 3083 computer using a version of the Gaussian 82 system of programs. The structures of DNA base pairs, adenine-thymine base pair, and guanine-cytosine base pair, are shown in Figure 2. Bond lengths, angles, and intermolecular distances between DNA bases are taken from Arnott et al. [27]. In adenine-thymine base pair, the bond length between N_{11} and H_{13} and that between N_{16} and H_{22} were elongated in steps of 0.1 Å along the lines N_{11}-O_{23} and N_{16}-N_1, respectively, and energies for double proton transfer were calculated [see Fig. 2(a)]. In guanine-cytosine base pair, two cases of the transferred protons are possible. One is the pair of H_{10} and H_{24} and the other is that of H_{12} and H_{24} [see Fig. 2(b)]. In the first case, the bond length between N_1 and H_{10} and that between N_{23} and H_{24} were elongated in steps of 0.1 Å along the lines, N_1-N_{17} and N_{23}-O_{14}, respectively, and energies for double proton transfer were calculated. Similarly, in the second case the bond length between N_{11} and H_{12} and that between N_{23} and H_{24} were elongated in steps of 0.1 Å along the lines N_{11}-O_{29} and N_{23}-O_{14}, respectively, and energies for double proton transfer were calculated.
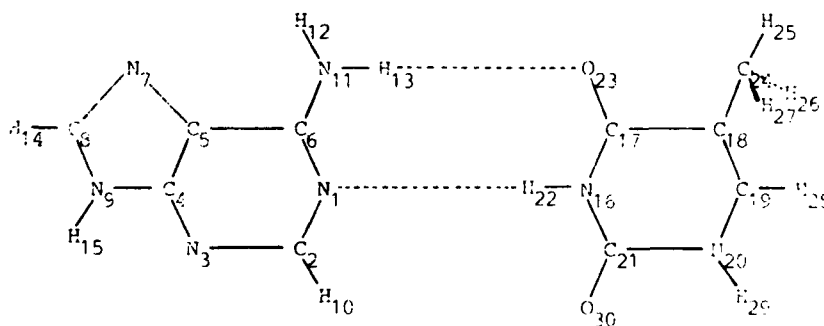
Figure 2(a).   Structures of DNA base pairs. Adenine-thymine base pair.

Results and Discussions

Water Clusters

We obtained potential energies and their profiles for proton transfer reactions in water clusters, $(H_2O)_n$ (n = 2, 3, 4, 5, 6). These results are summarized in Tables I to V and Figures 3 to 7. Proton transfer reaction in dimer shows single-well potential energy profile, while proton transfer reactions in trimer, tetramer, pentamer, and hexamer show double-well potential energy profiles. We also obtained three-dimensional pictures for proton transfer reactions in water clusters. Figure 8 shows the three-dimensional pictures for proton transfer reaction in dimer as an example of single

TABLE I. Potential energies for proton transfer reaction in dimer.

RC[a]	Potential energy (kcal/mol)
1	−2066.17
2	−2072.66
3	−2066.11
4	−2053.73
5	−2040.94
6	−2029.90
7	−2019.80
8	−2008.99

[a]When RC = 1, RO_1H_1 = 0.894 Å. As RC is increased by 1, RO_1H_1 is increased by 0.1 Å. RO_1H_1 is the distance between oxygen (1) and hydrogen (1) in dimer.

TABLE II. Potential energies for proton transfer reaction in trimer.

RC[a]	Potential energy (kcal/mol)
1	−3107.58
2	−3114.11
3	−3108.38
4	−3097.60
5	−3102.81
6	−3108.91
7	−3111.66
8	−3110.91

[a]When RC = 1, RO_1H_1 = 0.900 Å. As RC is increased by 1, RO_1H_1 is increased by 0.1 Å. RO_1H_1 is the distance between oxygen (1) and hydrogen (1) in trimer.

TABLE III. Potential energies for proton transfer reaction in tetramer.

RC[a]	Potential energy (kcal/mol)
1	−4157.56
2	−4163.46
3	−4162.29
4	−4159.41
5	−4162.50
6	−4162.71
7	−4161.14

[a]When RC = 1, RO_1H_1 = 0.927 Å. As RC is increased by 1, RO_1H_1 is increased by 0.1 Å. RO_1H_1 is the distance between oxygen (1) and hydrogen (1) in tetramer.

TABLE IV Potential energies for proton transfer reaction in pentamer.

RC[a]	Potential energy (kcal/mol)
1	−5218.19
2	−5221.57
3	−5220.23
4	−5221.95
5	−5222.42
6	−5220.13
7	−5215.94

[a]When RC = 1, RO_1H_1 = 1.007 Å. As RC is increased by 1, RO_1H_1 is increased by 0.1 Å. RO_1H_1 is the distance between oxygen (1) and hydrogen (1) in pentamer

TABLE V.   Potential energies for proton
transfer reaction in hexamer.

RC[a]	Potential energy (kcal/mol)
1	−6270.92
2	−6273.30
2.5	−6273.20
3	−6274.41
4	−6275.85
5	−6275.59
6	−6272.36

[a]When RC = 1, RO_1H_1 = 1.029 Å. As
RC is increased by 1, RO_1H_1 is increased by
0.1 Å. RO_1H_1 is the distance between oxy-
gen (1) and hydrogen (1) in hexamer.
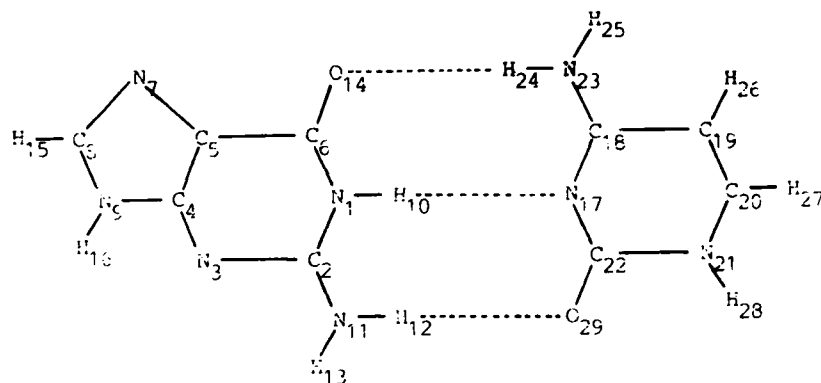
Figure 2(b).   Structures of DNA base pairs, Guanine-cytosine base pairs.

proton transfer, while Figure 9 shows the three-dimensional pictures for proton trans-
fer reaction in tetramer as an example of simultaneous proton transfer. In Figure 8,
hydrogen (1) is transferred from oxygen (1) to oxygen (2). As reaction coordinate
(RC) is increased from 4 to 5, hydrogen (1), which is close to oxygen (1) when
RC=4, becomes close to oxygen (2). In Figure 9, hydrogen (3), hydrogen (5), and
hydrogen (7) are transferred simultaneously with hydrogen (1). Similarly, as RC is
increased from 4 to 5, hydrogen (1), hydrogen (3), hydrogen (5), and hydrogen (7),
which are close to oxygen (1), oxygen (2), oxygen (3), and oxygen (4), respectively,
when RC=4, become close to oxygen (4), oxygen (1), oxygen (2), and oxygen (3),
respectively. As a result, proton transfer reaction in dimer shows single proton trans-
fer and single-well potential energy profile, while proton transfer reactions in trimer,
tetramer, pentamer, and hexamer show simultaneous proton transfers and double-well
potential energy profiles. The facts that single proton transfer results in a single-well
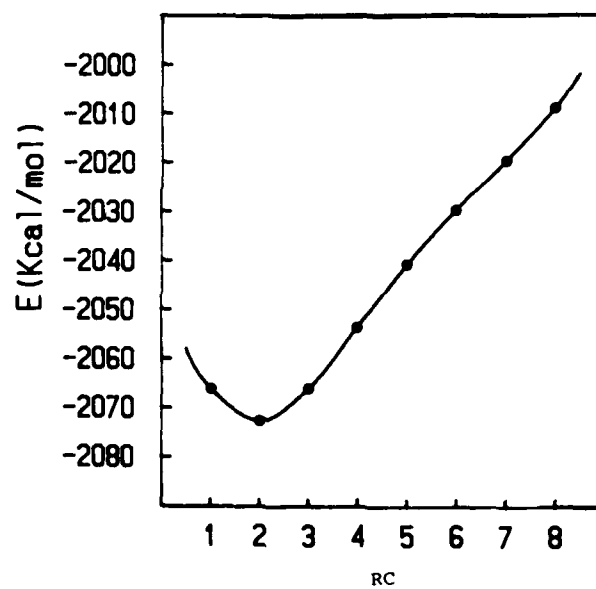
Figure 3. Potential energy profile for proton transfer reaction in dimer.
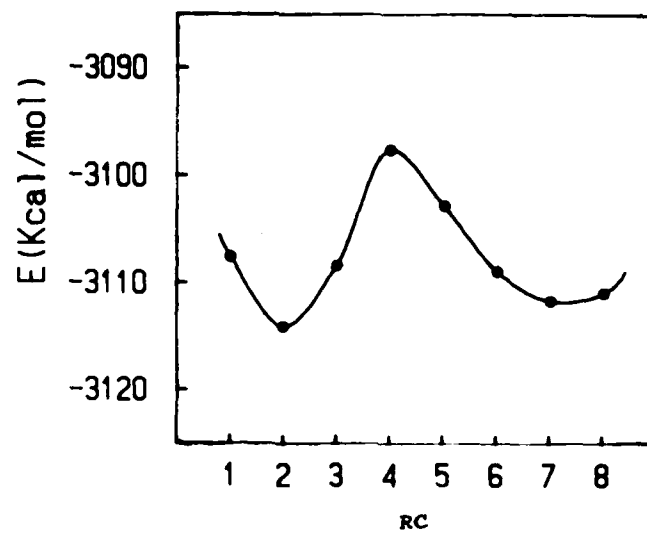
Figure 4. Potential energy profile for proton transfer reaction in trimer.

KONG, JHON, AND LÖWDIN
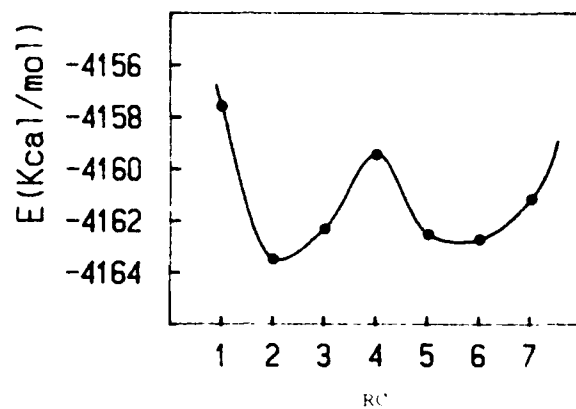
Figure 5. Potential energy profile for proton transfer reaction in tetramer.
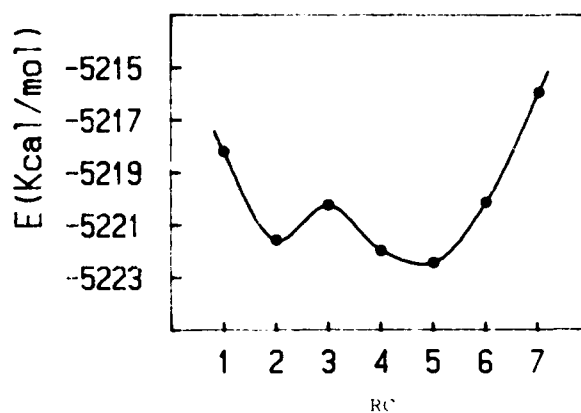
Figure 6. Potential energy profile for proton transfer reaction in pentamer.
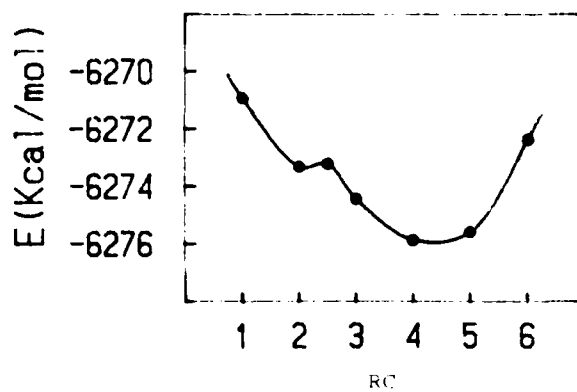
Figure 7. Three-dimensional picture for proton transfer reaction in hexamer.
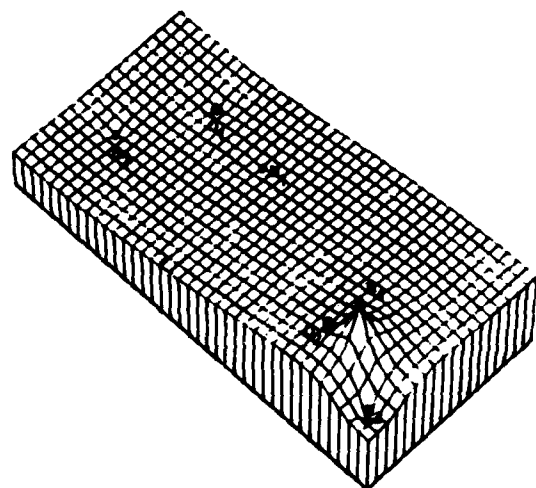
NW=2   RC=1
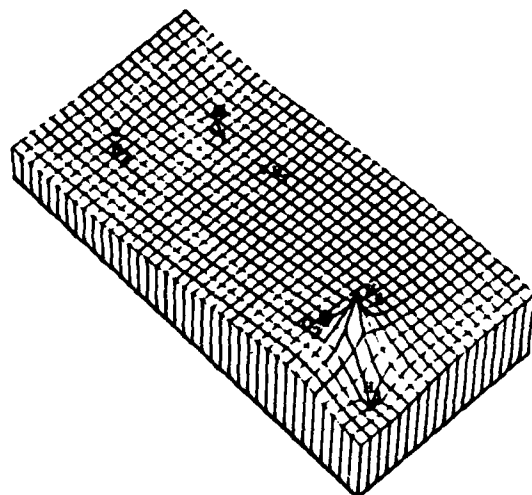
NW=2   RC=2

Figure 8.   Three-dimensional picture for proton transfer reaction in tetramer.
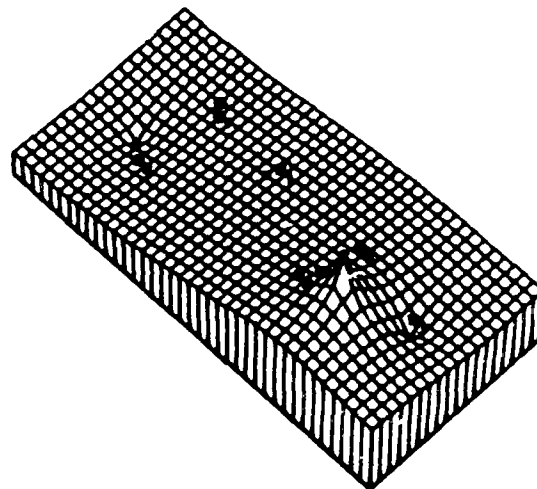
SW = 2   RC = 3
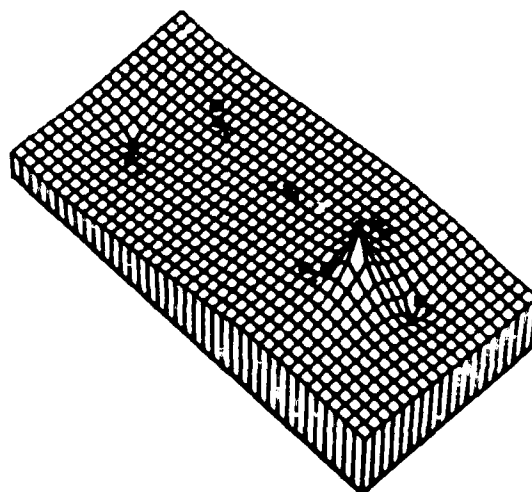
SW = 2   RC = 4

Figure 8   (Continued)
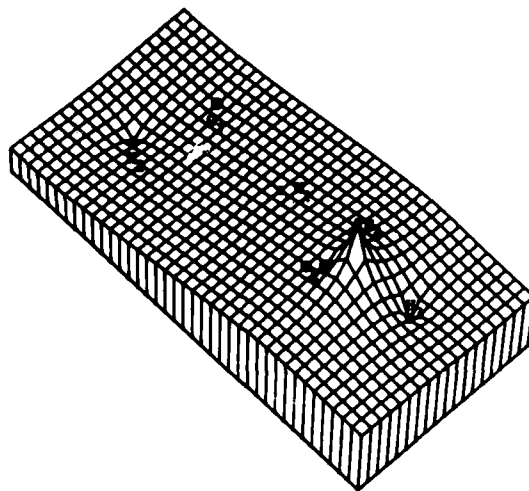
NW=2   RC=5
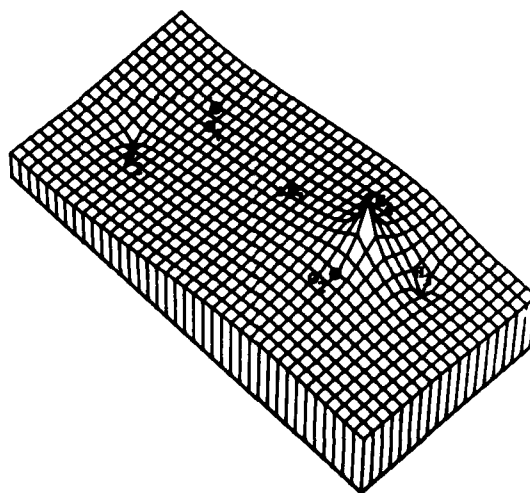
NW=2   RC=6

Figure 8.   *(Continued.)*

KONG, JHON, AND LÖWDIN
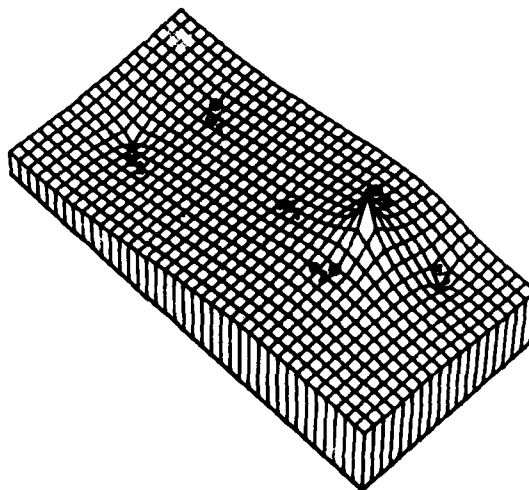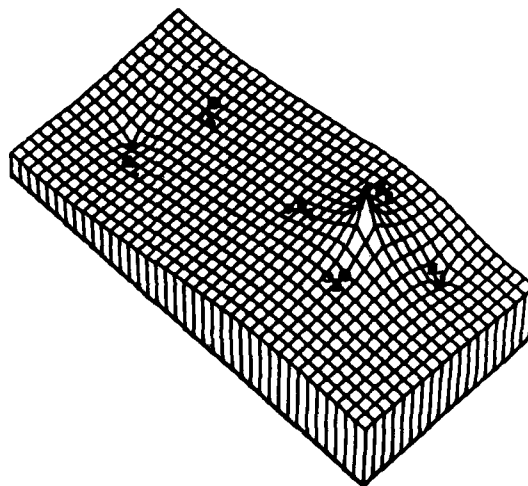
NW=2   RC=7

NW=2   RC=8

Figure 8.   *(Continued.)*

potential and simultaneous proton transfers result in double-well potentials are consistent with other studies [28, 29]. Namely, in dimer, the potential energy increases monotonically and results in a single-well potential, because single proton transfer leads to a pair of charged species. But in trimer, tetramer, pentamer, and hexamer, simultaneous proton transfers generally preserve electroneutrality in the system and the potential energies result in double-well potentials. In cyclic geometries especially, simultaneous proton transfers occur easily, since the barriers of simultaneous proton transfer reactions are relatively small.

DNA Base Pairs

The calculated energies for the double proton transfers in DNA base pairs are listed in Tables VI to VIII. Figures 10 to 12 also show the energy profiles for the double proton transfers in DNA base pairs.

Adenine-thymine base pair; The double proton transfer in adenine-thymine base pair gives double-well potential energy profile. From Table VI and Figure 10, one sees two minima (when RC=2 and RC=10). The barrier height for the double proton transfer is 70.24 kcal/mol. Guanine-cytosine base pair (H_{10} and H_{24}); The double proton transfer in guanine-cytosine base pair (the pair of H_{10} and H_{24}) gives a double-well potential energy profile. From Table VII and Figure 11, one sees two minima (when RC=2 and RC=10). The barrier height for the double proton transfer is 48.85 kcal/mol.

Guanine-cytosine base pair (H_{12} and H_{24}); The double proton transfer in guanine-cytosine base pair (the pair of H_{12} and H_{24}) gives double-well potential energy profile. From Table VIII and Figure 12, one also sees two minima (when RC=2 and RC=9). The barrier height for the double proton transfer is 64.02 kcal/mol.

TABLE VI. Energies for double proton transfer in adenine-thymine DNA base pair (H_{13} and H_{22}).

RC	$d(N_1-H_{13})$ (Å)	$d(N_{16}-H_{22})$ (Å)	E (a.u.)	ΔE^a (kcal/mol)
1	0.91	0.92	−904.26040	20.45
2	1.01	1.02	−904.29299	0.00
3	1.11	1.12	−904.28411	5.57
4	1.21	1.22	−904.25518	23.73
5	1.31	1.32	−904.22124	45.03
6	1.41	1.42	−904.19383	62.22
7	1.51	1.52	−904.18105	70.24
8	1.61	1.62	−904.18493	67.81
9	1.71	1.72	−904.19933	58.77
10	1.81	1.82	−904.21099	51.46
11	1.91	1.92	−904.20056	58.00

^{a}Energy difference when compared with the energy (RC = 2).

TABLE VII. Energies for double proton transfer in guanine-cytosine DNA base pair
(H_{10} and H_{24}).

RC	$d(N_1\text{-}H_{10})$ (Å)	$d(N_{23}\text{-}H_{24})$ (Å)	E (a.u.)	ΔE^a (kcal/mol)
1	0.92	0.91	−919.99521	21.60
2	1.02	1.01	−920.02963	0.00
3	1.12	1.11	−920.02368	3.73
4	1.22	1.21	−919.99939	18.98
5	1.32	1.31	−919.97244	35.89
6	1.42	1.41	−919.95443	47.19
7	1.52	1.51	−919.95178	48.85
8	1.62	1.61	−919.96311	41.74
9	1.72	1.71	−919.97897	31.79
10	1.82	1.81	−919.98336	29.04
11	1.92	1.91	−919.95326	47.92

^{a}Energy difference when compared with the energy (RC = 2).

TABLE VIII. Energies for double proton transfer in guanine-cytosine DNA base pair
(H_{12} and H_{24}).

RC	$d(N_{11}\text{-}H_{12})$ (Å)	$d(N_{23}\text{-}H_{24})$ (Å)	E (a.u.)	ΔE^a (kcal/mol)
1	0.91	0.91	−919.99412	22.28
2	1.01	1.01	−920.02963	0.00
3	1.11	1.11	−920.02338	3.92
4	1.21	1.21	−919.99727	20.31
5	1.31	1.31	−919.96628	39.75
6	1.41	1.41	−919.94111	55.55
7	1.51	1.51	−919.92813	63.69
8	1.61	1.61	−919.92761	64.02
9	1.71	1.71	−919.93193	61.31
10	1.81	1.81	−919.92514	65.57
11	1.91	1.91	−919.88186	92.73

^{a}Energy difference when compared with the energy (RC = 2).
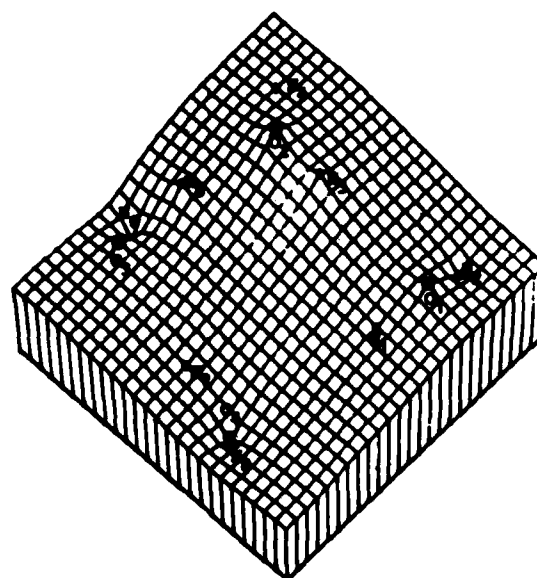
NW=4 RC=1
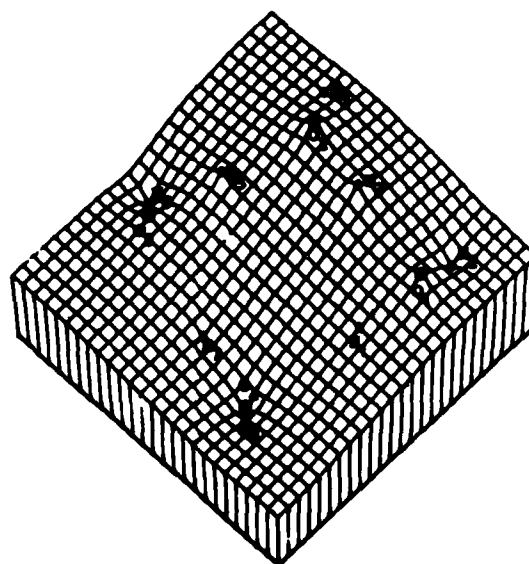
NW=4 RC=2

Figure 9. Energy profile for double proton transfer in adenine-thymine DNA base pair $(H_{13}$ and $H_{22})$.
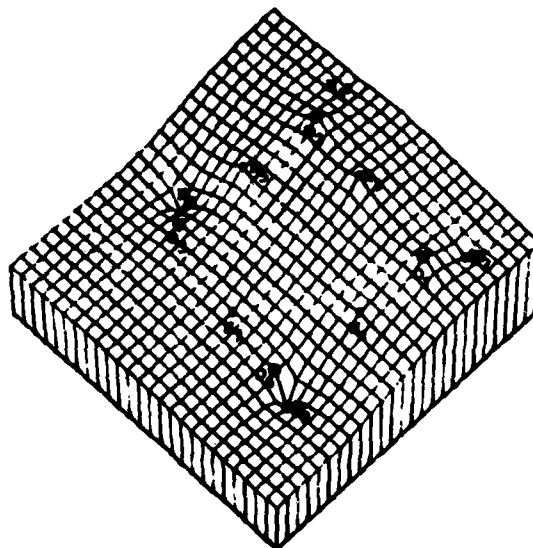
NW=4   RC=3
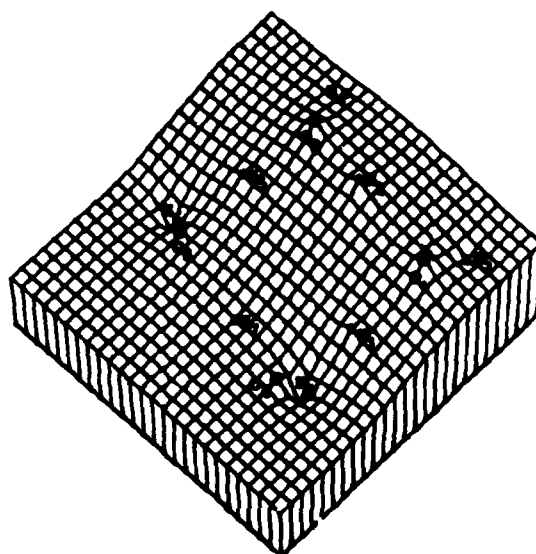
NW=4   RC=4

Figure 9.   *(Continued.)*
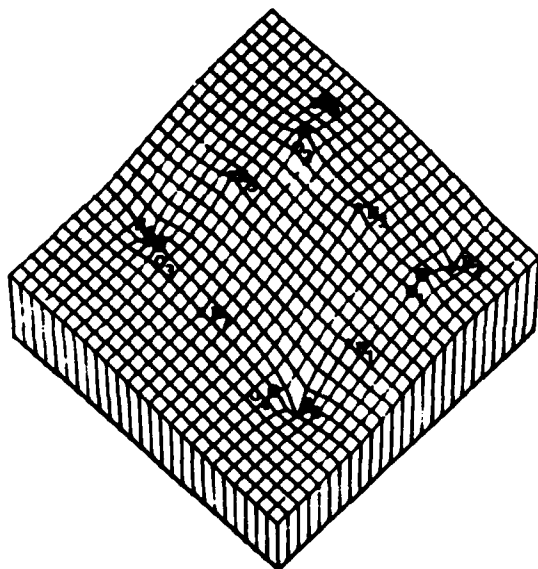
NW=4   RC=5
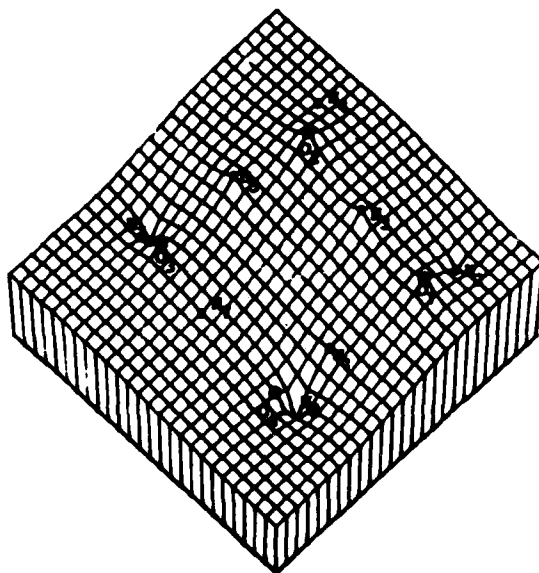
NW=4   RC=6

Figure 9.   *(Continued.)*
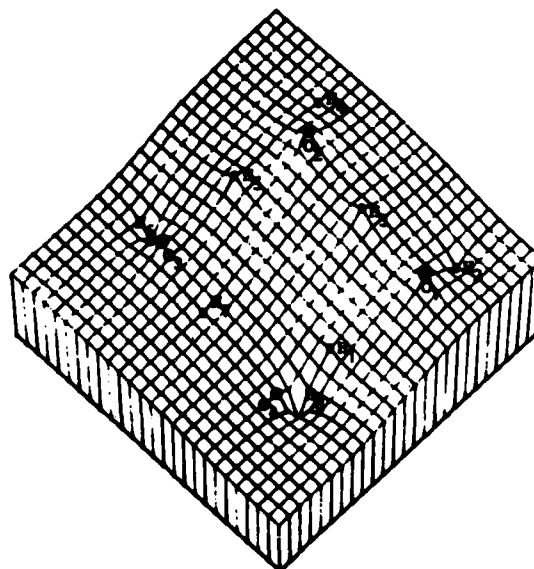
NW=4   RC=7

Figure 9.   (Continued.)
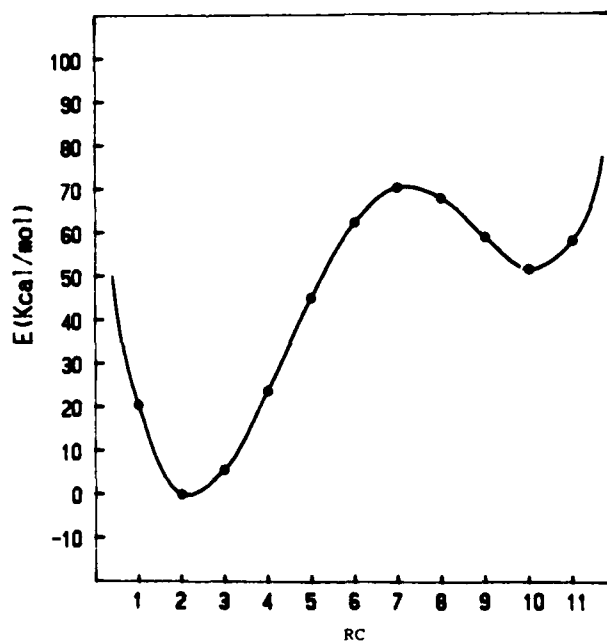
Figure 10.   Energy profile for double proton transfer in guanine-cytosine DNA base pair (H_{10} and H_{24}).
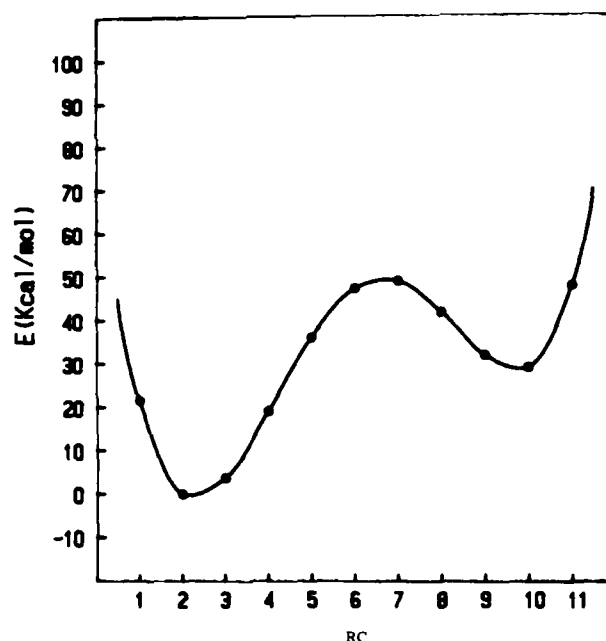
Figure 11. Energy profile for double proton transfer in guanine-cytosine DNA base pair
$(H_{12}$ and $H_{24})$.

From the above results, one can conclude that the double proton transfers in DNA base pairs give double-well energy profiles and Löwdin's proton tunneling model for mutagenesis [22] can be adopted. Namely, an error mechanism in the genetic code replication as a consequence of tautomer formation can be considered as a possible error mechanism, because the double proton transfers in DNA base pairs give double-well energy profiles. But an error mechanism in the genetic code replication as a consequence of ion-pair formation which is incurred by single proton transfer may be considered an unreasonable error mechanism, as Clementi et al. [28] pointed out that a single proton transfer in DNA base pair, which formed an ion-pair, gave a single-well energy profile characterized by a monotonically increasing energy function.

In conclusion, if we adopt Löwdin's error mechanism in the genetic code replication as a consequence of tautomer formation, our calculated results provide information that the possibility of error in the genetic code replication in guanine-cytosine base pair occurs more easily than that in adenine-thymine base pair because the barrier height of the double proton transfers in guanine-cytosine base pair (48.85 kcal/mol and 64.02 kcal/mol) is lower than that in adenine-thymine base pair (80.24 kcal/mol). Also, in the guanine-cytosine base pair, the possibility of error in the genetic code replication in the pair of H_{10} and H_{24} occurs more easily than that in the pair of H_{12} and H_{24} because the barrier height of the double proton transfer in the pair of H_{10} and H_{24} (48.85 kcal/mol) is lower than that in the pair of H_{12} and H_{24} (64.02 kcal/mol). However, we may conclude that, in order to obtain the more reliable quantitative results for the proton transfers in DNA base pairs, we have to make fur-
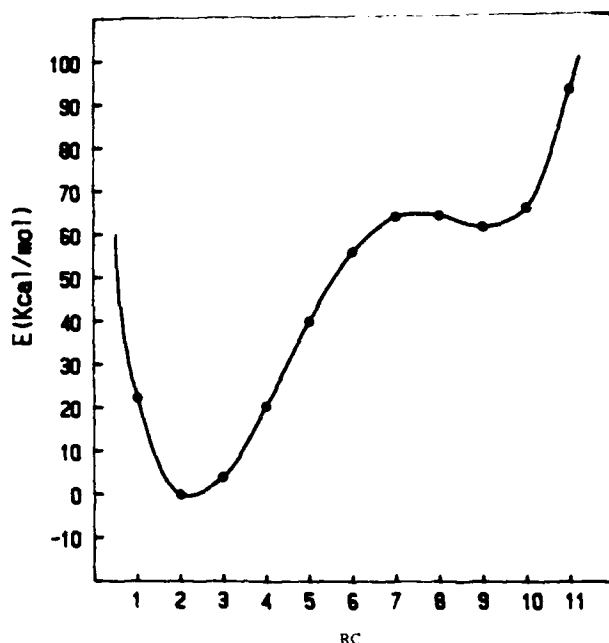
Figure 12.   Energy profile for double proton transfer in guanine-cytosine DNA base pair
$(H_{12}$ and $H_{24})$.

ther studies to obtain potential energies with a larger basis set and optimize the geometries of DNA base pairs which may require a lot of computing time. In the near future, we will carry out these extended studies and predict tunneling rates using the improved results (e.g., more reliable potential barrier heights and band widths).

Bibliography

[1] E. Caldin and V. Gold, Eds., *Proton Transfer Reactions*, (Wiley, New York, 1975).

[2] P. Schuster, G. Zundel, and C. Sandorfy, Eds., *The Hydrogen Bond-Recent Developments in Theory and Experiments*, (North-Holland Publishing Co., Amsterdam, 1976).

[3] M. D. Newton, J. Chem. Phys. **67**, 5535 (1978).

[4] M. D. Newton and S. Ehrensen, J. Am. Chem. Soc. **93**, 4971 (1971).

[5] R. Janoschek, E. G. Weidemann, and G. Zundel, J. Chem. Soc., Faraday Trans. **2**, 69, 505 (1973).

[6] R. Janoschek, E. G. Weidemann, H. Pfeiffer, and G. Zundel, J. Am. Chem. Soc. **94**, 2387 (1972).

[7] P. A. Kollman and L. C. Allen, J. Am. Chem. Soc. **92**, 6101 (1970).

[8] W. P. Kraemer and G. H. F. Direcksen, Chem. Phys. Lett. **5**, 463 (1970).

[9] W. Meyer, W. Jakubetz, and P. Schuster, Chem. Phys. Lett. **21**, 97 (1973).

[10] G. Alagona, R. Cimiraglia, and V. Lamanna, Theor. Chim. Acta **29**, 93 (1973).

[11] J. J. Delpuech, G. Serratrice, A. Strich, and A. Veillard, Mol. Phys. **29**, 849 (1975).

[12] P. Schuster, W. Jakubetz, G. Beider, W. Meyer, and R. M. Rode, in *Chemical and Biochemica Reactivity*, E. D. Bergmann and B. Pullman, Eds. (Academic Press, Jerusalem, 1974), p. 257.

[13] S. Scheiner, J. Am. Chem. Soc. **103**, 315 (1981).
[14] L. L. Ingraham, Biochim. Biophys. Acta **279**, 8 (1972).
[15] R. D. Gandour, G. M. Maggiora, and R. L. Schowen, J. Am. Chem. Soc. **96**, 6967 (1974).
[16] H. Chojnacki, Int. J. Quantum Chem. **16**, 299 (1979).
[17] R. Contreras and J. S. Gomez-Jeria, J. Phys. Chem. **88**, 1905 (1984).
[18] J. D. Watson and F. H. C. Crick, Nature (London) **171**, 737, 964 (1953); Cold Spring Harbor Symp. Quant. Biol. **18**, 123 (1953).
[19] J. S. Kwiatkowski and B. Pullman, Adv. Heterocycl. Chem. **18**, 199 (1975).
[20] P. O. Löwdin, Rev. Mod. Phys. **35**, 724 (1963).
[21] P. O. Löwdin, Biopolymers Symp. **1**, 161, 293 (1964).
[22] P. O. Löwdin, Adv. Quantum Chem. **2**, 212 (1965).
[23] F. H. Stillinger and C. W. David, J. Chem. Phys. **69**, 1473 (1978).
[24] F. H. Stillinger and C. W. David, J. Chem. Phys. **73**, 3384 (1980).
[25] R. Fletcher, *Fortran Subroutine for Minimization by Quasi-Newton Method*, A.E.R.E. Report, R7125 (1972).
[26] Y. J. Park, Y. K. Kang, B. J. Yoon, and M. S. Jhon, Bull. Korean Chem. Soc. **3**, 50 (1982).
[27] S. Arnott, P. J. C. Smith, and R. Chandrasekaran, in *CRC Handbook of Biochemistry and Molecular Biology*, G. D. Fasman, Ed. (CRC, Cleveland, 1976), 3rd Ed., Vol. 2, p. 411.
[28] E. Clementi, J. Mehl, and W. Niessen, J. Chem. Phys. **54**, 508 (1971).
[29] H. Grinberg, A. L. Caparelli, A. Spina, J. Maranon, and O. M. Sorarrain, J. Phys. Chem. **85**, 2751 (1981).

Practical Considerations in Calculations of the Proton Transfer in a Model Active Site of Papain

JOHANNES P. DIJKMAN,* ROMAN OSMAN, and HAREL WEINSTEIN

Department of Physiology and Biophysics, Mount Sinai School of Medicine of the City University of New York, New York, New York 10029, U.S.A.

Abstract

Practical aspects of the calculation of the proton transfer process in a model of the active site of the thiol protease papain are explored with basis sets of different sizes. Results from *ab initio* calculations with the STO-3G, 4-31G, 6-31G, 6-31G* basis set, and a 6-31G basis set augmented with polarization functions on the sulfur atom are compared for their performance in describing the proton transfer energy. The nature of the convergence of the calculated properties of the potential curve for proton transfer with the increase in basis set indicates the need for a split-valence basis set and for polarization functions on the sulfur in order to achieve an appropriate description of this system. Correlation corrections to the calculated energies are shown to contribute significantly to the characteristics of the proton transfer energy curve.

Introduction

Proton transfer is a common element in many mechanisms proposed for enzymatic reactions [1, 2]. The role of the proton transfer in these schemes is to generate an acid-base pair which can be chemically reactive. For example, in serine-proteases the nucleophile Ser⁻ is formed by a transfer of the Ser proton to histidine forming a Ser⁻ ... HisH⁺ pair in the active site [3]; in cysteine proteases the proton from a Cys is transferred to a His to form a strong nucleophile S⁻ [for a review see Refs. 1–3]. Because the proton transfer has a central role in such enzymatic mechanisms, much effort is spent investigating the details of the process. Many theoretical studies have focused on the proton transfer process in a variety of small model systems [4–9, 19, 20]. Important details of this process have been revealed by such studies, and a common finding has been that the calculated potential energy curve is strongly basis set dependent. This paper evaluates some practical considerations in the calculations of the proton transfer mechanism proposed for the enzymatic reaction of papain by Drenth et al. [10, 11]. The first step in this mechanism is a transfer of the proton from the sulfur of Cys-25 to the Ndl of the imidazole of His-159 to form a charge separation in the active site, referred to as the "zwitterion" state.

Based on the conclusions reached from theoretical studies on other model systems (see above), a first concern in the investigation of this mechanism is the identification

---

* On leave from the Laboratory of Chemical Physics, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands.

of a suitable basis set for the calculation of the potential energy curve which describes the proton transfer. The characteristics of the proton transfer energy curves obtained from calculations with a variety of basis sets are compared here in order to identify the minimal computational level at which the results agree with those from the most extensive calculations that were practicable for the system.

Computational Details

The coordinates of the model active site were taken from the x-ray structure of papain refined to 1.65 Å resolution [12]. The residues of the active site that participate in the proton transfer process were modeled by methanethiol for Cys-25, and by imidazole for His-159 (Fig. 1). Throughout the proton transfer process the S—Ndl distance was kept fixed at the value of 3.38 Å observed in the crystal.

The *ab initio* molecular orbital calculations were performed with the GAUSSIAN82 [13] and the HONDO [14] program packages. The basis sets used in this comparison were STO-3G [15], 4-31G [16], 6-31G [17], 6-31G* [18], and a 6-31G^+ which is a 6-31G basis set augmented with a set of 3d-functions (six primitives) on the sulfur atom, with an exponent of 0.65. Correlation effects were calculated with the Moller-Plesset perturbation expansion to second (MP2) and third (MP3) order. Frozen core MP2 and MP3 calculations were performed only with the 6-31G and the 6-31G^+ basis set.
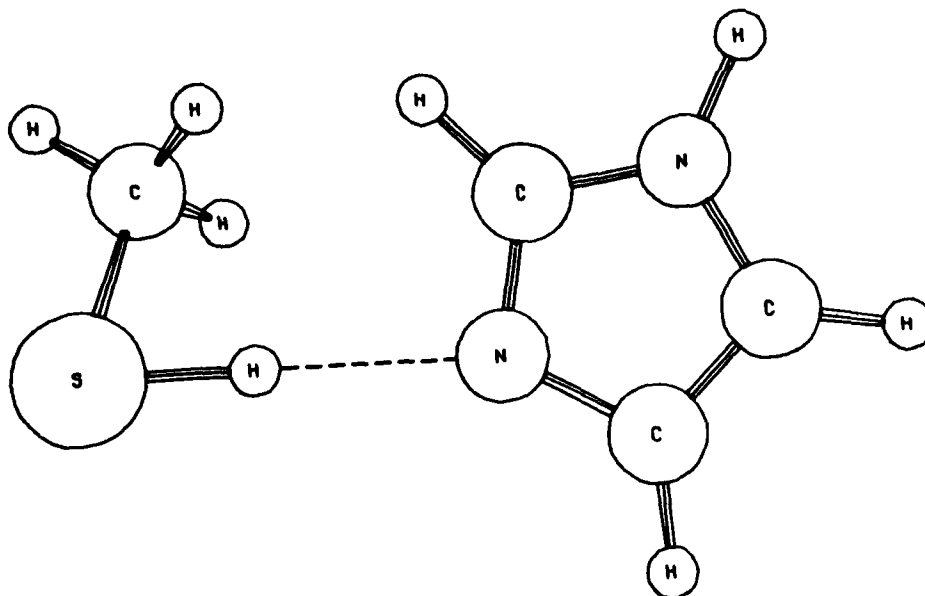
Figure 1.   Structure of the model for the residues Cys-25 and His-159 in the active site of papain. The coordinates were taken from the x-ray structure [12] and the moving proton (shown near the sulfur atom) was placed on the S-Ndl axis.

A total of 16 points was used to construct the potential energy curves. The proton was displaced along the line between the sulfur and the nitrogen (Fig. 1). The energy was calculated at equidistant displacements of the proton from the sulfur, separated by 0.1 Å. The extrema of the curve, MinI, MinII, and E_{max} were calculated from a fit to a polynomial of degree 15. MinI is the minimum with the proton near the sulfur, MinII is the minimum with the proton near the Nd1, and E_{max} is the maximum in the potential energy curve. The parameters chosen to characterize the potential energy curve are E^{stab}, which is defined as (MinII–MinI), and E^{act} defined as $(E_{max}$–MinI).

The correlation corrections to the energy were calculated for the three points on the curve representing the two minima and the transition state.

Results and Discussion

Figure 2 shows the proton transfer energy curve calculated with the STO-3G basis set. The curve has a minimum for the proton near the sulfur atom, but no minimum when it approaches the Nd1 of the imidazole. A change in the slope of the potential energy curve occurs when the proton approaches the Nd1 atom. The single minimum, as well as the other characteristics of the curve are similar to those reported for the same system [19, 20] from a calculation with the minimal basis set developed by Mehler and Paul [21].
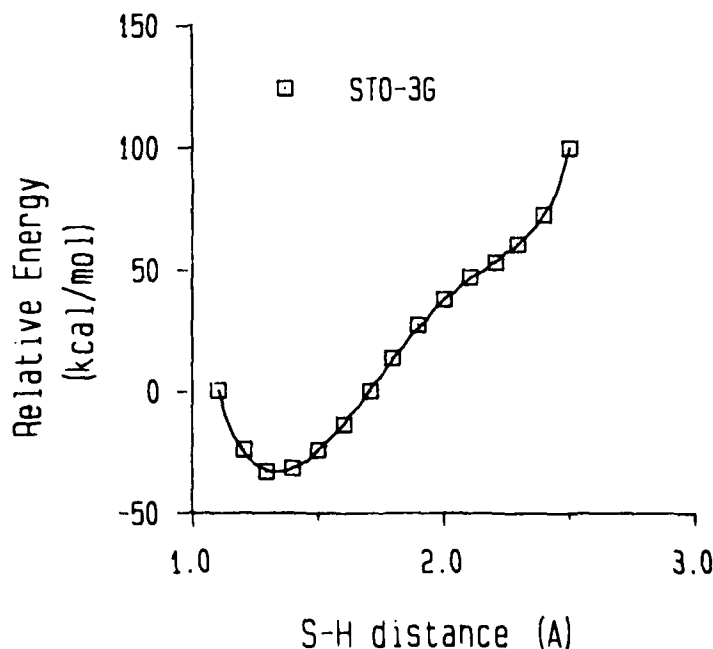
Figure 2.   Energies of the model active site of papain (see Fig. 1) as a function of the position of the proton. The calculations were done with the STO-3G basis set and the results are shown relative to a zero value defined as the energy of the system with the proton near the sulfur atom at an S-H distance of 1.1 Å.

The results of calculations with the split valence basis set 4-31G are represented in Table I and in Figure 3. With this basis set, two minima appear in the proton transfer curve. The energies calculated with the STO-3G and the 4-31G basis sets are compared in Table II and Figure 4(a). The results show that the calculation with the larger basis set affects the zwitterion more than the other two states of the system. Thus, the energy of the zwitterion is lowered far more than the energy of the neutral form or the transition state. This indicates the importance of a split-valence basis set for the description of a state in which a charge separation has occurred.

TABLE I.   Basis set dependence of the characteristic parameters of the proton transfer curve in the model active site of papain.[a]

	STO-3G	4-31G	6-31G	6-31G^+
R(MinI)*	1.332	1.374	1.370	1.339
R(Max)	—	2.009	1.968	2.000
E^{act}	—	21.1	21.1	31.8
R(MinII)	—	2.297	2.317	2.312
E^{stab}	—	16.6	13.3	25.6

[a] R is the distance of the proton to the sulfur, in Å. The extrema on the curve, (MinI, Max, MinII), the activation energy (E^{act}) and the stabilization energy (E^{stab}) are calculated as described in the text. Energies are expressed in kcal/mol.
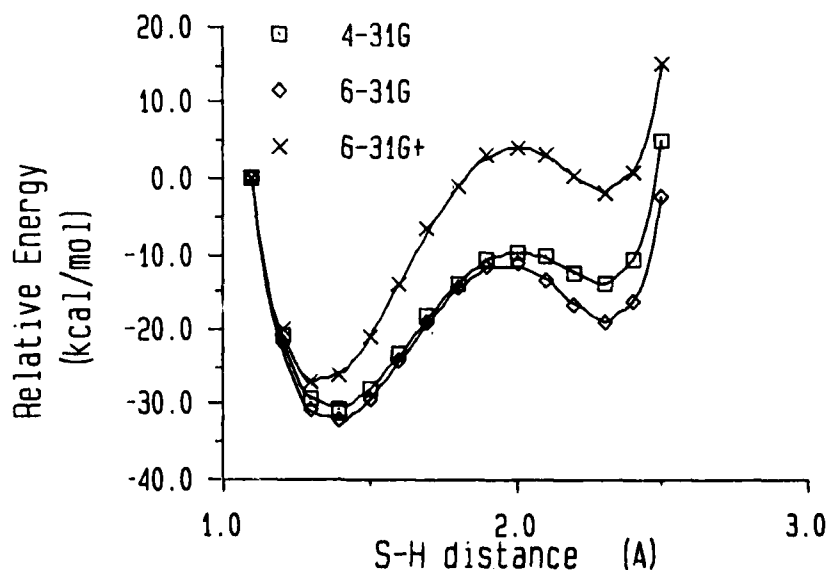
Figure 3.   Energies of the model active site of papain (see Fig. 1) as a function of the position of the proton. The calculations were done with the 4-31G, 6-31G, and 6-31G^+ basis sets, as indicated. The results are shown relative to a zero value defined as the energy of the system with the proton at an S-H distance of 1.1 Å, calculated with the corresponding basis set.

TABLE II. Basis set dependence of the energies of the active site model of papain as a function of the position of the proton.[a]

R(SH)	E1	E2	E3	E4
1.1	−4253.8	−4693.0	−439.2	−29.7
1.2	−4250.0	−4690.4	−440.4	−27.9
1.3	−4249.6	−4690.4	−440.9	−25.8
1.4	−4252.1	−4693.0	−440.8	−23.6
1.5	−4257.2	−4697.7	−440.5	−21.3
1.6	−4264.1	−4704.2	−440.1	−19.2
1.7	−4272.3	−4712.2	−439.8	−17.4
1.8	−4281.4	−4721.4	−439.9	−16.0
1.9	−4291.2	−4731.5	−440.3	−14.8
2.0	−4301.0	−4742.1	−441.0	−13.9
2.1	−4309.7	−4751.7	−442.0	−13.2
2.2	−4320.1	−4763.2	−443.1	−12.8
2.3	−4328.8	−4773.0	−444.2	−12.4
2.4	−4337.7	−4782.6	−445.0	−12.2
2.5	−4349.0	−4795.6	−446.6	−12.0

[a] R is the distance of the proton to the sulfur, in Å. Values in the columns are obtained from differences in energies calculated with different basis sets, defined as follows: E1 = E(4-31G)-E(STO-3G), E2 = E(6-31G)-E(STO-3G), E3 = E(6-31G)-E(4-31G), E4 = E(6-31G^+)-E(6-31G); all energy differences are in kcal/mol.

Extending the basis set to 6-31G, which provides an improved description of the core electrons, results in qualitatively the same proton transfer energy curve as that calculated with the 4-31G basis set. Quantitatively, the energies from the 6-31G calculation are about 440 kcal/mol lower than the 4-31G results [Table II and Fig. 4(b)]. Unlike the transition from a minimal basis set (STO-3G) to the split-valence basis set (4-31G), there is little improvement in the description of the zwitterion with respect to the neutral form in going from 4-31G to 6-31G. Consequently, the activation energy, E^{act}, is the same in the calculations with both basis sets, and the stabilization energy, E^{stab}, is only 3.3 kcal/mol smaller in the case of the 6-31G basis set. Augmenting the 6-31G basis set with 3d functions produces a dramatic change in the calculated values of E^{act} and E^{stab}, which are increased by 10.7 and 12.3 kcal/mol, respectively. The characteristics of the proton transfer curve calculated with these basis sets are presented in Table I and Figure 3. The results presented in Table II and Figure 4(b) show that with the augmented basis set, the neutral form is more stabilized than the zwitterion. Because this result could be due to the unequal representation of the sulfur and the nitrogen by the basis set augmented only on the sulfur, it is noteworthy that even with the 6-31G^* basis set the neutral form is stabilized more than the zwitterion relative to the results with the 6-31G basis set (Table III). The values of E^{act} and E^{stab} calculated with the 6-31G^+ and the 6-31G^* basis sets are very similar, differing only by 3.2 and 3.5 kcal/mol, respectively (Table IV). These results indicate that the 3d functions on the sulfur are the main contributors to the improvement produced by the addition of polarization functions.
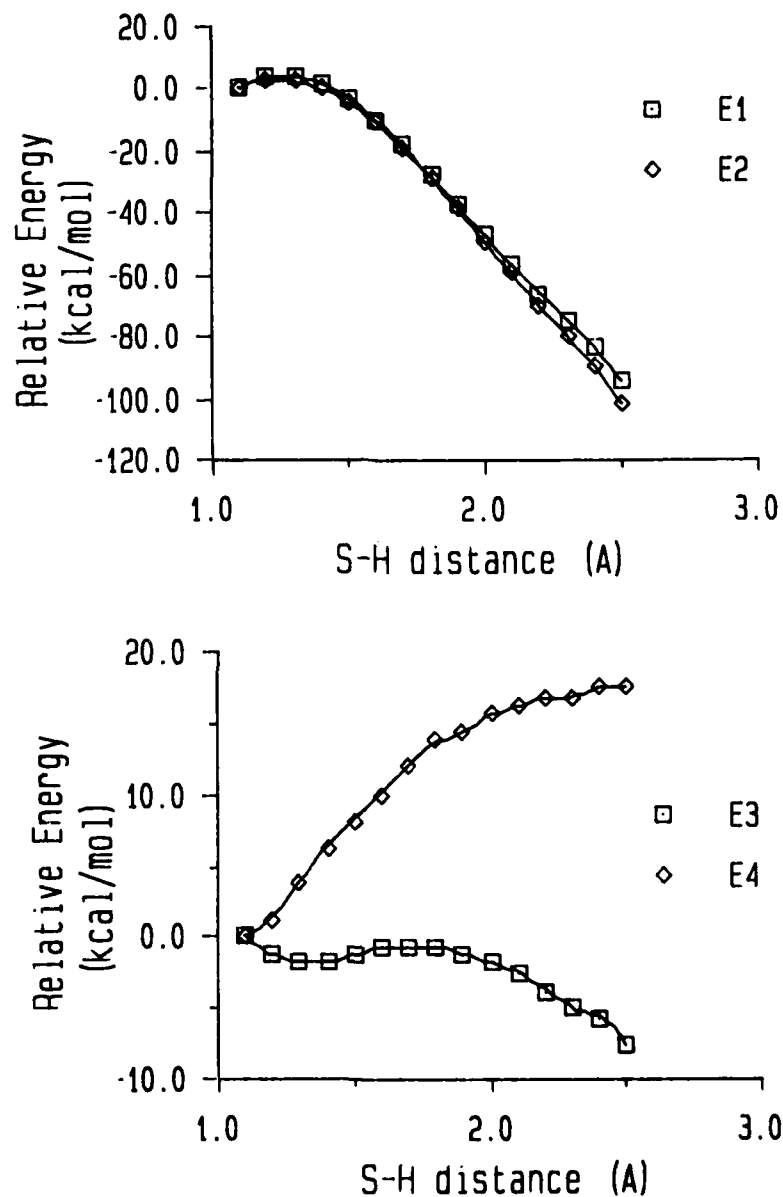
Figure 4. Differences of the total energies of the model active site of papain (see Fig. 1) calculated with different basis sets. The values are shown relative to the energy difference calculated for an S-H distance of 1.1 Å with the corresponding basis sets. (a) The effect of split-valence representations: Difference between energies obtained with double zeta basis sets and the minimal basis set: E1 = E(4-31G) − E(STO-3G), E2 = E(6-31G E(STO-3G); (b) The effect of improved core in the split-valence representation, and the el ... t of polarization functions: E3 = E(6-31G) − E(4-31G), E4 = E(6-31G^*) − E(6-31G).

TABLE III. Additional energy stabilization with respect to results obtained with the 6-31G basis set provided by polarization functions on sulfur alone (6-31G^+) and on all heavy atoms (6-31G^*).[a]

	Basis set	
R(S—H)	6-31G^+	6-31G^*
1.4	−23.6	−99.0
2.0	−13.9	−86.1
2.3	−12.4	−84.3

[a] R(S—H) is the distance of the proton from the sulfur in Å. The three points correspond to the extrema in the proton transfer curve calculated with the 6-31G basis set. Energies are in kcal/mol.

Correlation energy corrections at the MP2 and the MP3 level calculated with the 6-31G and the 6-31G^+ basis set for the two minima and the transition state are shown in Table IV. The MP2 correction calculated with the 6-31G basis set lowers E^{act} by 4.1 kcal/mol, to a new value of 16.7 kcal/mol. The MP3 contribution has the opposite sign and raises E^{act} back to 18.1 kcal/mol. Notably, MinII is almost eliminated by the correlation contributions. To evaluate the possibility that the position of the second minimum was shifted from the previously observed minimum at 2.3 Å by

TABLE IV. Values of the activation energies (E^{act}) and stabilization energies (E^{stab}) calculated with various basis sets and correlation energy corrections.[a]

	E^{act}	E^{stab}
6-31G	20.8	13.2
6-31G^+	30.4	24.3
6-31G^*	33.6	27.8
6-31G/MP2	16.7	18.3
6-31G/MP3	18.2	17.7
6-31G +/MP2	24.9	27.5

[a] Results are taken from calculations at the three distances of the moving proton from the sulfur, at which the extrema of the potential energy curve for proton transfer occur in the calculation with the 6-31G basis set: R = 1.4Å, 2.0Å, and 2.3Å. Energies are in kcal/mole.

the addition of the correlation correction, the MP2 contribution was also calculated for the proton closer to S, at an S—H distance of 2.2 Å. The choice of this point was based on the observation that MP2 tends to increase slightly the optimal bond length [22], in this case between the proton and Nd1. The calculated SCF + MP2 energy at the S—H distance of 2.2 Å is only 0.9 kcal/mol lower than the energy of the system at a distance of 2.3 Å.

The effect of correlation in calculations with the 6-31G^+ basis set is nearly the same as for the 6-31G basis set. Due to the prohibitively large computational effort, only the MP2 correction for the activation barrier could be calculated with this basis set. The correlation correction decreases the barrier by 5.5 kcal/mol, which is not very different from the MP2 correction for the activation barrier calculated with 6-31G (4.1 kcal/mol).

Conclusions

The results of this study show that a basis set with at least a split-valence representation is required to describe the proton transfer in the active site of papain. Even the smallest split valence basis set used here (4-31G) yielded a double well potential for the proton transfer. The characteristics of this potential curve, namely, E^{stab} and E^{act}, converged to the values obtained with 6-31G^* as the basis sets were gradually augmented. The results also emphasize the need for a basis set with polarization functions on the sulfur atom for the adequate representation of this system. Because the MP2 correlation corrections are similar in magnitude to the stabilization energy of the zwitterion with respect to the transition state, it is not possible to establish whether the double well nature of the potential will be maintained when electron correlation is completely accounted for.

Acknowledgments

Bibliography

[1] L. Polgar and P. Halasz, Biochem. J. 207, 1 (1982).
[2] A. Fersht, Enzyme Structure and Mechanism (Freeman & Co., New York, 1984).
[3] A. Warshel and S. T. Russell, Quart. Rev. Biophys. 17, 282 (1984).
[4] S. Topiol, G. A. Mercier, R. Osman, and H. Weinstein, J. Comput. Chem. 6, 581 (1985).
[5] P. A. Kollman and D. M. Hayes, J. Am. Chem. Soc. 103, 2955-2961 (1981).
[6] G. Naray-Szabo, A. K. Kapur, P. G. Mezey, and L. Polgar, J. Molec. Struct. 90, 137 (1982).
[7] S. Scheiner, J. Chem. Phys. 80, 5 (1984).
[8] S. Scheiner, Acc. Chem. Res. 18, 174 (1985).

[9] H. Weinstein, A. P. Mazurek, R. Osman, and S. Topiol, Molec. Pharm. **29**, 28 (1986).

[10] J. Drenth, K. H. Kalk, and H. M. Swen, Biochemistry **15**, 371 (1976).

[11] J. Drenth, H. M. Swen, W. Hoogenstraaten, and L. A. A. Sluikterman, Proc. K. Ned. Akad. Wet. **c78**, 104 (1975).

[12] I. G. Kamphuis, K. H. Kalk, M. B. A. Swarte, and J. Drenth, J. Molec. Biol. **179**, 233 (1984).

[13] J. S. Binkley, M. J. Frisch, D. J. DeFrees, K. Raghavachari, R. A. Whiteside, H. B. Whiteside, H. B. Schlegel, E. M. Fluder, and J. A. Pople, GAUSSIAN82, Department of Chemistry, Carnegie-Mellon University, Pittsburgh, PA.

[14] M. Dupuis, J. Rys, and H. F. King, J. Chem. Phys. **65**, 111 (1976).

[15] (a) First-row: W. J. Hehre, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **51**. 2657 (1969); (b) Second-row: W. J. Hehre, R. Ditchfield, R. F. Stewart, and J. A. Pople, J. Chem. Phys. **52**, 2769 (1970); (c) Third-row, main group: W. J. Pietro, B. A. Levi, W. J. Hehre, and R. F. Stewart, Inorg. Chem. **19**, 2225 (1980).

[16] (a) Carbon to Fluorine: R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724 (1971); (b) Phosphorous to Chlorine: W. J. Hehre and W. A. Lathan, J. Chem. Phys. **56**, 5255 (1972).

[17] (a) W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972); (b) 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**, 3654 (1982).

[18] P. C. Hariharan and J. A. Pople, Chem. Phys. Lett. **66**, 217 (1972).

[19] P. T. van Duijnen, B. T. Thole, R. Broer, and W. C. Nieuwpoort, Int. J. Quantum Chem. **7**, 651 (1980).

[20] P. T. van Duijnen and B. T. Thole, Biophys. Chem. **13**, 133 (1981).

[21] E. L. Mehler and C. Paul, Chem. Phys. Lett. **63**, 145–151 (1979).

[22] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory* (Wiley & Sons, New York, 1986).

Drug and Receptors in Molecular Biology

R. H. DAVIES

*Welsh School of Pharmacy, University of Wales Institute of Science and Technology, Cardiff, CF1 3XF,
Wales*

Abstract

Dose–response relationships of drug-receptor binding show that receptor sites are, in many cases, singly
occupied by the drug molecules. Although this single-site occupancy may be demonstrated for bound hor-
mone analogues which inhibit, stimulate, or partially stimulate the response, the molecular occupancy of
the receptor site is essentially statistical in character, and the observed binding constant may represent a
sum of conformer contributions. These conformer contributions are proportionately weighted by the rele-
vant conformer fractions of the drug and receptor for each interaction. In practice, more than one con-
former may bind productively to the receptor, while, on the other hand, even within one identifiable
conformation, restriction on the fraction of molecules eliciting a response could produce partial agonism.
The thermodynamic representation of explicit models of receptor interaction are reviewed taking into ac-
count the reference phase of the receptor environment and its potential heterogeneity. Decomposition of
thermodynamic data for membrane-bound β-adrenoceptor agents shows that referencing the data to a hy-
drocarbon environment produces more comparative insight into enthalpic differences. Differences in the
enthalpies of binding of the phenoxypropanolamine derivatives practolol and propranolol are largely due to
loss of hydration on the amidic carbonyl moiety of practolol. Using this hydrocarbon model reference state
for comparison, major differences in the enthalpies of binding of the amine moiety in phenethanolamines
and phenoxypropanolamines are observed. There is a 6–7 kcal enthalpic loss in substituting a methyl
group on the protonated amine moiety of noradrenaline, and a further similar loss of 6–7 kcal in substitut-
ing t-butyl for the isopropyl group. In contrast, the phenoxypropanolamine derivatives show an approxi-
mately constant mode of binding for these alkyl substituents. The possibilities that the amine moiety is
sited differently in phenethanolamine and phenoxypropanolamine binding, and is multiply hydrogen
bonded to three receptor sites in the natural hormone are explored. The identification of bioactive conform-
ers in intracellular and membrane-bound receptor agents is also reviewed.

Introduction

The pharmacological concept of a receptor is derived from the highly specific
binding exhibited by hormones and drug molecules (concentration 10^{-9} molar) in
given sites coupled with the emission or competitive inhibition of a measurable signal
(response). Dose–response relationships [1–4] show that the sites in many cases are
singly occupied by the drug molecules.

Cellular receptors [5] for hormones and neurotransmitters are subdivided into those
localized within the cell (within the nucleus and the cytoplasm) which are often water
soluble, and the hydrophobic cell membrane surface receptors. The intracellular hor-
mones exhibit diverse roles as may be seen in the regulatory morphological role of
the steroids, androgen and estrogen in the developmental biology of the sexes. In
contrast, a wide variety of ligands for membrane-bound receptors — the peptide and

glycoprotein hormones, the biogenic amines, and the prostaglandins are associated with coupling to the enzyme adenylate cyclase. Therapeutic advantages in changes in the level and degree of hormone regulation are exemplified by antisteroid treatments in prostate and breast cancer, and by selective control of the sympathetic neurotransmitter noradrenaline on the adenylate cyclase coupled β-adrenoceptors, with the appropriate regulation of cardiac, bronchial, and vascular responses.

Although single-site occupancy may, in many cases, be demonstrated for bound hormone analogues which inhibit, stimulate, or partially stimulate the response (in pharmacological terms, antagonists, agonists, and partial agonists, respectively) the molecular occupancy of the receptor sites is essentially statistical in character and the observed binding constant may represent a sum of individual conformer contributions. Thus, more than one conformer may bind productively to the receptor while, on the other hand, even within one identifiable conformation, restriction on the fraction of molecules eliciting a response might produce a population balance for partial agonism. Again, a bioactive conformer may be energetically unfavorable with a minor population in the conformer distribution of the unbound molecules, with resultant loss in potency. Binding constants, it need hardly be emphasized are free energy-related quantities.

Thermodynamic data on β-adrenoceptor binding at the membrane level indicate major differences between the binding of agonists and antagonists. Agonists bind with highly favorable enthalpies but with large negative entropies of complex formation [6–9]. Antagonists show relatively weak bonding but with favorable entropies for receptor binding. More quantitative understanding of the thermodynamic changes involved could obviously be deduced if the full structures of the receptor were known when partition functions for the receptor complex might be written and thermodynamic functions computed. However, even where the receptor structures are not known, comparative binding studies should yield some knowledge of the receptor sites, the appropriate cancellation of terms of the partition function leaving a potential residue of analyzable information.

It is the purpose of this review to examine what structural information is derivable from drug-receptor data when the molecular details of the receptor are not known, and to examine the viability of explicit models of drug-receptor interaction.

Representational Models of Receptor-Stimulus Actions

The most general representational model for relating receptor binding and stimulus action, when the dose-response relation is essentially hyperbolic, is due to Black and Leff [10].

Where the law of mass action pertains for binding of the drug A to the receptors R, then the resultant concentration of the drug AR is given by the normal hyperbolic relation

$$[AR] = \frac{K_A[A]R_T}{1 + K_A[A]} \tag{1}$$

where K_A is the binding constant of the drug and R_T, the total number of receptors. The observed stimulus is some function of the concentration AR. For a hyperbolic

function to exist between the dose A and the stimulus response, two possible relations join $[AR]$ to the observed effect. The effect may be (1) a linear function or (2) a hyperbolic function of $[AR]$.

Thus if the effect E is given by

$$E = \frac{K_E[AR]}{1 + K_E[AR]} \tag{2}$$

where K_E represents some amplification factor of the signal, then

$$E = \frac{K_A[A]\tau}{1 + (\tau + 1)K_A[A]} \quad \text{where } \tau = K_E R_T \tag{3}$$

or

$$E = \epsilon \frac{K_A'[A]}{1 + K_A'[A]} \quad \text{where } \epsilon = \frac{\tau}{\tau + 1} \quad \text{and} \quad K_A' = (\tau + 1)K_A \tag{4}$$

Three parameters, the binding constant K_A, the amplification factor K_E, and the total number of receptors R_T define the cognitive and transducer functions of the drug upon the receptor. To relate this model to molecular mechanisms of drug-receptor action, it is necessary to consider explicit models of molecular interaction.

Molecular Models of Receptor Stimulus Action.
Some Thermodynamic Relations

Since only a fraction of drug molecules may be relevant to binding to the receptor (and only a fraction of the receptor conformations may be relevant to binding to the drug) it is convenient to represent the gross binding constant in a conformer representation [11].

In terms of standard partial free energies, μ° the gross binding constant $\overline{K}$ may be written as:

$$\mu_{AR}^\circ - \mu_A^\circ - \mu_R^\circ = -kT \log \overline{K} \tag{5}$$

where the subscripts AR, A, and R refer to the complex, drug, and receptor, respectively.

Using second indices to identify the conformer i of the drug A engaged in binding, with j^x its receptor (R) counterpart, then, for the ij^x conformer interaction, (1) may be written.

$$\mu_{A_iR_jx}^\circ + (\mu_{AR}^\circ - \mu_{A_iR_jx}^\circ) - \mu_{A_i}^\circ - (\mu_A^\circ - \mu_{A_i}^\circ)$$
$$- \mu_{R_jx}^\circ - (\mu_R^\circ - \mu_{R_jx}^\circ) = -kT \log \overline{K} \tag{6}$$

Using conformer populations f^i of A, and f^{j^x} of R, and the relations

$$\mu_A^\circ - \mu_{A_i}^\circ = kT \log f^i$$

$$\mu_R^\circ - \mu_{R_jx}^\circ = kT \log f^{j^x}$$

and summing over the bound states, then

$$\sum_i \sum_{j^x} K^{ij^x} f^i f^{j^x} = \overline{K} \tag{7}$$

where K^{ij^x} the conformer binding constant, is given by

$$\mu^\circ_{A_i R_{j^x}} - \mu^\circ_{A_i} - \mu^\circ_{R_{j^x}} = -kT \log K^{ij^x} \tag{8}$$

The binding constant is a sum of the conformer binding constants weighted by their appropriate conformer fractions.

It is more convenient to define the conformer binding constant $K^{ij^x}_\square$ referenced to the average states of A and R.

$$\mu^\circ_{A_i R_{j^x}} - \mu^\circ_A - \mu^\circ_R = -kT \log K^{ij^x} f^i f^{j^x} = -kT \log K^{ij^x}_\square \tag{9}$$

The observed standard free energy change on binding can also be written as a weighted sum over the bound states of the conformer binding constant terms, together with the associated entropy of mixing

$$\Delta G^\circ = -RT \sum_i \sum_{j^x} p_{A_i R_{j^x}} \log K^{ij^x} f^i f^{j^x} + RT \sum_i \sum_{j^x} p_{A_i R_{j^x}} \log p_{A_i R_{j^x}}$$

$$= -RT \sum_i \sum_{j^x} p_{A_i R_{j^x}} \log K^{ij^x}_\square + RT \sum_i \sum_{j^x} p_{A_i R_{j^x}} \log p_{A_i R_{j^x}} \tag{10}$$

Differentiating (9) with respect to temperature and using the relations

$$\delta(\log f^i)/\delta T = \frac{H^i - H^\circ_A}{RT^2}, \qquad \delta(\log f^{j^x})/\delta T = \frac{H^{j^x} - H^\circ_R}{RT^2} \tag{11}$$

Then

$$RT^2 \delta(\log K^{ij^x}_\square)/\delta T = \Delta H^{ij^x}_\square = \Delta H^{ij^x} + (H^i - H^\circ_A) + (H^{j^x} - H^\circ_{R_{j^x}})$$

$$\text{or } \Delta K^{ij^x}_\square = H^{ij^x} - H^\circ_A - H^\circ_R \tag{12}$$

Using the standard enthalpy of the bound states

$$H^\circ_{AR} = \sum_i \sum_{j^x} p_{A_i R_{j^x}} H^{ij^x} \tag{13}$$

the standard enthalpy for the binding becomes

$$\Delta H^\circ_{AR} = \sum_i \sum_{j^x} p_{A_i R_{j^x}} \Delta H^{ij^x}_\square \tag{14}$$

The standard entropy change on binding similarly may be written

$$\Delta S^\circ_{AR} = \sum_i \sum_{j^x} \Delta S^{ij^x}_\square - \sum_i \sum_{j^x} R p_{A_i R_{j^x}} \log p_{A_i R_{j^x}} \tag{15}$$

It may be convenient to consider comparative drug binding with a change of reference, to a hydrocarbon lipid phase, L. The standard free energy contribution of A in (6) can then be written

(1) $\qquad -\mu_{A,L}^{\circ} - (\mu_A^{\circ} - \mu_{AL}^{\circ}) - (\mu_{AL}^{\circ} - \mu_{A,L}^{\circ})$

(2) $\qquad -\mu_{A,L}^{\circ} - (\mu_A^{\circ} - \mu_{A_i}^{\circ}) - (\mu_{A_i}^{\circ} - \mu_{A,L}^{\circ})$

(16)

and since

$$\mu_{AL}^{\circ} - \mu_{A,L}^{\circ} = kT \log f_L^i$$

$$\mu_{A_i}^{\circ} - \mu_{A,L}^{\circ} = kT \log P^i$$

$$\mu_A^{\circ} - \mu_L^{\circ} = kT \log P \qquad (17)$$

where f_L^i is the conformer fraction of i in a nonaqueous medium, and P^i is the conformer or micropartition coefficient of the species i [12], it follows that

(1) $\qquad \sum_i \sum_{j^x} K_L^{ij^x} f_L^i f_L^{j^x} P = \overline{K}$

(2) $\qquad \sum_i \sum_{j^x} K_L^{ij^x} f^i f_L^{j^x} P^i = \overline{K}$

(18)

These relations may be observed from the free energy diagram in Figure 1. Appropriate transformation of Eq. (10) may be similarly applied. The utility of such relations is subsequently examined, but it may be observed that defined sets of conditions are
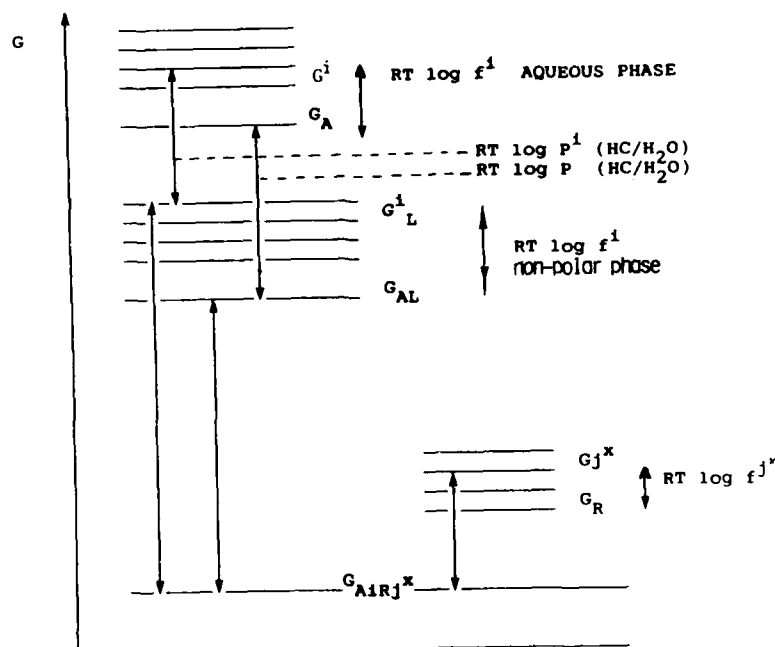
Figure 1. Schematic representation of free energy relations for the conformer i of drug A interacting with the relevant receptor conformer j^x of the receptor protein complex and possible pathways for relating the bound conformer free energy $G_{A_iR_{j^x}}$ to the reference free energy G_A.

required if useful correlations in comparative drug data are to be obtained. Thus in any comparative analogue set $K^{ij^x}f^{j^x}$ should be invariant. A more detailed discussion is given elsewhere [13].

Bioactive Conformer Identification

The influence of conformer fraction on the binding constant of a drug may be exemplified by axial-equatorial preference in a cyclic aliphatic ring system. Figure 2 shows the relative conformer fraction on the $\log_{10}$ scale plotted against the energetic preference of the two conformers attributable to an anomeric effect.

An obvious corollary from the figure is that where bond energetics appear invariant to changes in electron distribution about the bond, the conformation is already energetically dominant. Figure 3 shows a partial interpretation of β-adrenoceptor antagonist action where a dominant conformation about the $-O-CH_2-$ moiety must exist. Table I shows the bond rotational energetics for the $-OCH_3$ group attached to various aromatic and heterocyclic ring systems. For this group (but cf, $N-CH_3$ [17]), STO-3G minimal basis results given reasonable agreement compared with experiment and with better basis set determinations. For β-adrenoceptor binding, the CH_2 moiety must lie planar with the aromatic ring [26].

We have found STO-3G calculations to give useful results in competitive conformer preference in the antiandrogen anilide derivatives used to inhibit the male hormone testosterone in the treatment of prostate cancer (Fig. 4). Table II shows a comparison of predicted and experimental conformer populations. Similar results were obtained
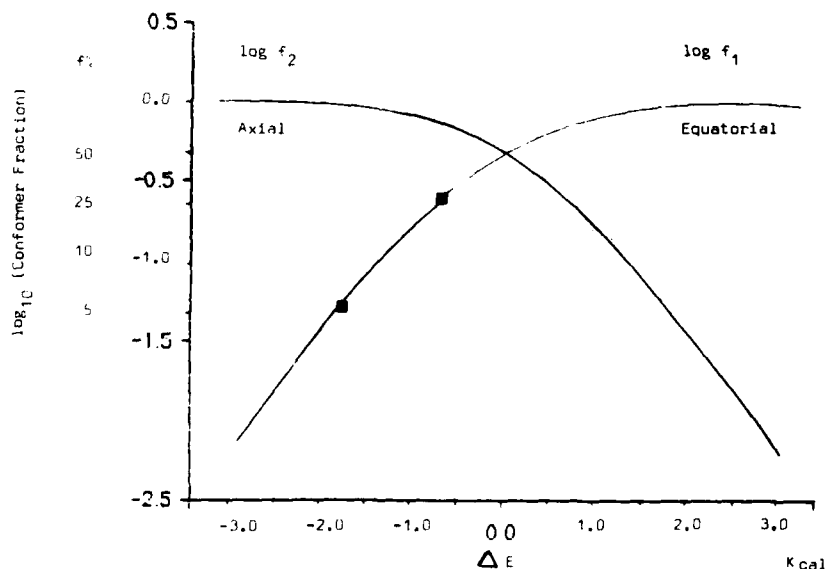
Figure 2.   Relative conformer preference for an axial/equatorial substituent orientation in a cyclic aliphatic ring system. The conformer fractions on the $\log_{10}$ scale are plotted against the energetic difference for the two conformers attributable to an anomeric effect.
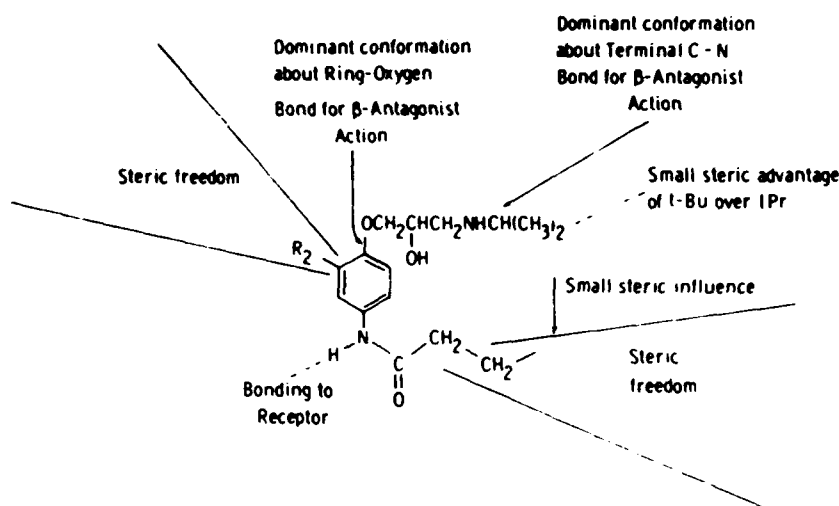
Figure 3. Partial interpretation of the action of p-acylamino phenoxypropanolamine derivatives on the cardiac β-adrenoceptor [13].

with a 4-31G basis set on a reduced molecule (R_1, R_2 = H). Receptor binding correlates with conformer 1, where convenient excitation of the hydrogen bond proton donor interaction from the aromatic ring is attainable [18].

When the fraction of the biologically active conformer is only minor, on the other hand, marked sensitivity to the size of the fraction pertains. A possible example is contained in data on analogues of the central nervous system agent viloxazine (R = OC_2H_5, Fig. 5) which inhibits neuronal reuptake of biogenic amines.
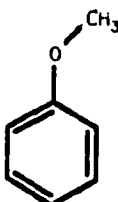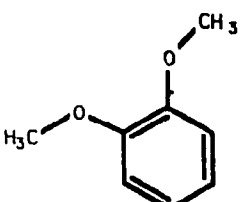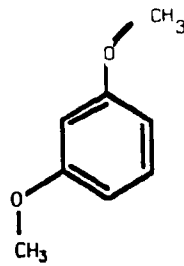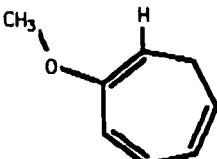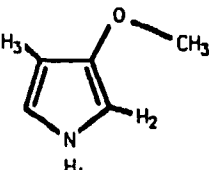
A plot of potency *in vivo* against a partitioning effect is given in Figure 6. If the residual variation from this data is plotted against the conformer fraction where the side chain is now perpendicular to the aromatic ring (Fig. 7) then a simple correlation is observed.

Such interpretations rely on a single conformer of the drug binding productively at the receptor. If more than one conformer binds to the receptor, then a range of behavior from the highly stereospecific to the nonspecific may occur. Other effects are possible. If one of the conformers is the arbiter of stimulus action, partial agonism may result as a competitive occupancy of the receptors by active and inactive conformers. The model for this action will be reviewed, but it is first necessary to consider the choice of reference state for comparing drug data in 'he case of membrane-bound receptors.

Thermodynamic Parameters of Ligand Binding to the β-Adrenergic Receptor. Choice of Reference Phase for Comparative Data

As seen from the previous example of viloxazine data, more stringent conditions are required to identify a bioactive conformer when more than one conformer may be acting upon the hormone receptors. Bound conformations may, of course, be inferred

TABLE I

	Conformational Energy Preference Planar/Perpendicular	Concentration Difference 37°C
	1.2 kcal[a] (STO-3G) 0.7 kcal (4-31G)[b]	5:1
	0.0 kcal (STO-3G)	1:1
	2.0 kcal (STO-3G)	25:1
	1.3 kcal[c] Synperiplanar to H Relative to Antiperiplanar (in CDCl₃, NMR)	8:1
	1.2 kcal [b](4-31G) SYNPERIPLANAR to H₂ or L[r] ANTIPERIPLANAR less favored	7:7:1

[a] From Ref. 14.
[b] From Ref. 15.
[c] From Ref. 16.

ANDROGEN   TESTOSTERONE

ANTI ANDROGEN   ANILIDE DERIVATIVE
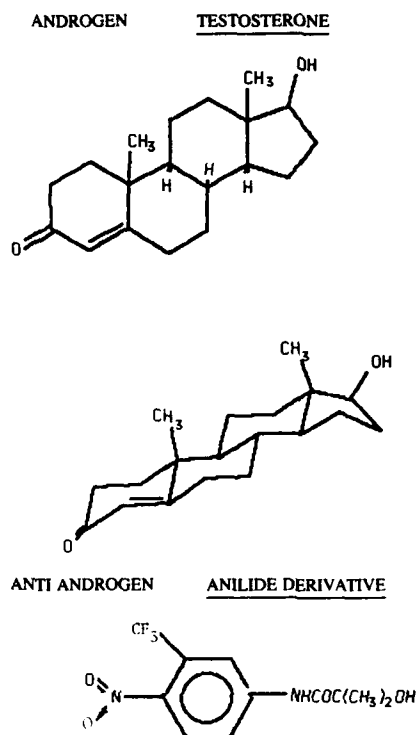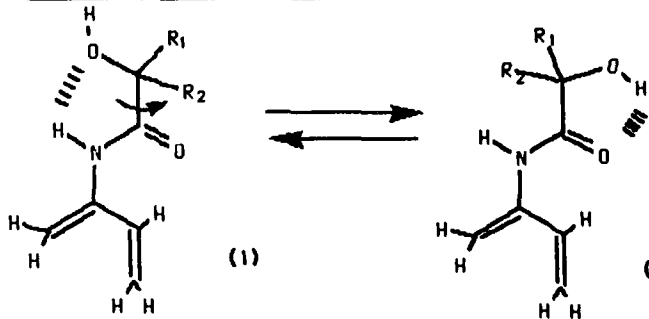
Figure 4.   The steroid androgen testosterone, and the antagonist anilide used in the treatment of prostate cancer.

TABLE II

R_1	R_2	STO-3G ΔE(kcal)	Relative conformer populations F_i on the $\log_{10}$ scale at 310°K $\log_{10}(F_1/F_2)$	Relative conformer population from infrared[a] measurements $\log_{10}(F_1/F_2)$
CH_3	CH_3	+1.5	+1.05	>1.0
CH_3	CF_3	−0.14	−0.10	−0.18
CF_3	CF_3	−1.8	−1.3	Anion at physiological pH

[a] From Ref. 18.
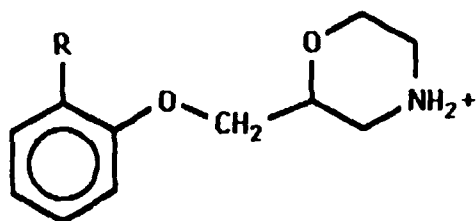
R	
2 OCH$_3$	Does Not Release Catecholamine From Neuronal Stores. Inhibits Reuptake.
3 OCH$_3$	Releases Catecholamine. Effective Reuptake.
3 OC$_2$H$_5$	Does Not Release Catecholamine.

Figure 5. Some properties of closely related analogues of the central nervous system agent, viloxazine (R = 2OC$_2$H$_5$) used to inhibit biogenic amine reuptake.
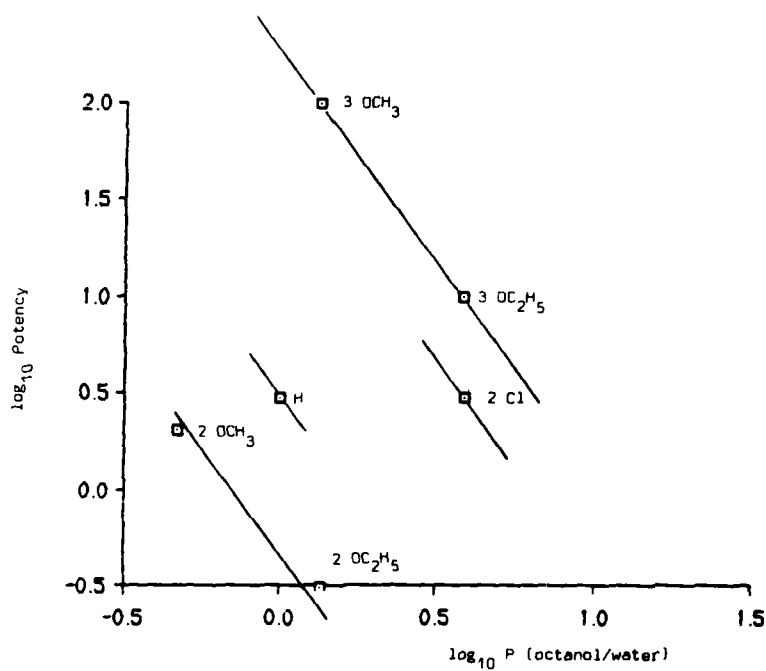
Figure 6. Potency *in vivo* of viloxazine analogues plotted against a partitioning effect using the octanol/water model on the log$_{10}$ scale.

F Conformer Fraction with Side Chain Perpendicular to Ring
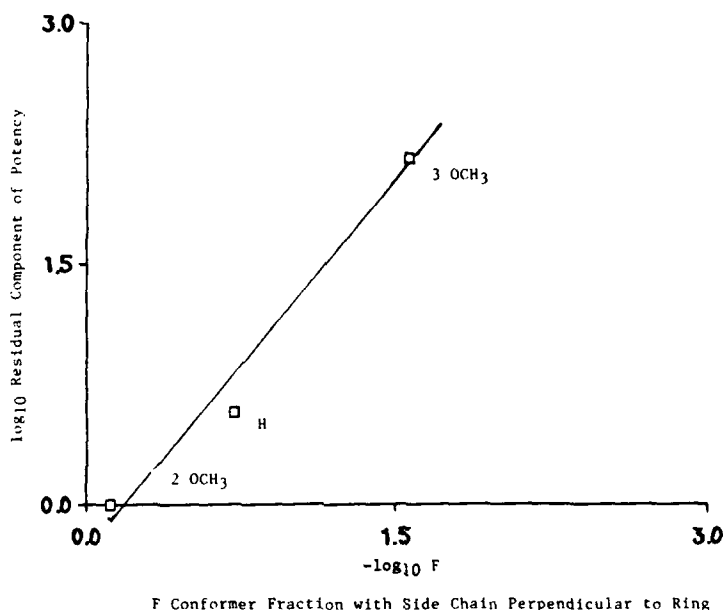
Figure 7. Residual variation in potency of viloxazine analogues after allowance for a partitioning effect plotted on the $\log_{10}$ scale against the fraction of the conformers having the side chain perpendicular to the ring.

from fixed conformation active isomers, but the choice of reference state for comparing drug data is crucial for correct identification of active conformers in flexible molecules.

Previous evidence has shown that the phase environment in given regions around the bound β-adrenoceptor antagonist and partial agonist molecules is hydrophobic in character, and that modelling of substituent effects in these regions with an effective hydrocarbon solvent model produces correlations of unit slope [13]. Even for alkyl substitution on the protonated amine moiety, where the alkyl hydrogen atoms are relatively acidic, a simple correlation appears evident in phenoxypropanolamine derivatives (Fig. 8). However, the limit of the environment in the region of the bound protonated amine moiety with its potential for attendant water molecules is unknown in this charged species and an aqueous rather than a hydrophobic medium might still be the most relevant for modelling this region of the bound molecule, apart from the need to site receptor counter ions. The alternative forms of Eq. (18) allow scope for incorporating such potential heterogeneity, provided that the partition coefficient for the relevant conformer can be estimated satisfactorily.

It should thus, be informative to remove the intrinsic variation in the data due to hydrophobic interactions by either referencing the data to a theoretical hydrocarbon environment, or by use of substituent corrections suggested by Eq. (18) in order that a clearer picture of the factors giving rise to the differences in the ΔH's might be identified.
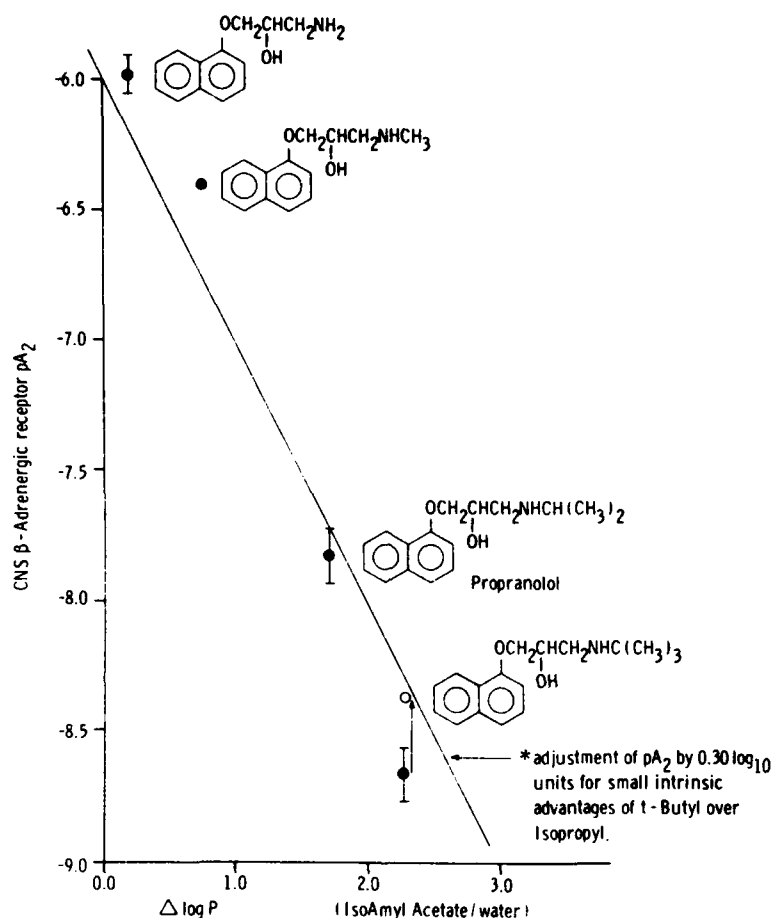
Figure 8. The influence of alkyl substitution on the amine moiety in propranolol analogues. Potencies on the β-adrenoceptor in isolated central nervous system membranes [19] plotted against a simple partitioning effect on the $\log_{10}$ scale using the iso-amyl acetate/water model.

Thermodynamic data on ligand binding to the β-adrenergic receptor of turkey erythrocytes is given in Table III [6]. A key of the ligands which comprise phenethanolamine and phenoxypropanolamine derivatives is in Table IV. Agonists bind with large negative entropies while antagonists show relatively weak ΔH's but favorable entropies for complex formation. Partial agonists show an intermediate behavior.

While bound conformations of phenethanolamines have been deduced from fixed conformation active isomers [20], the active conformers of phenoxypropanolamines upon the β-adrenoceptor are less certain [21–23]. To attempt to identify the bound active conformers using Eq. (18), the requisite thermodynamic model for accounting for the hydrophobic interactions may be considered.

TABLE III.   Thermodynamic parameters of ligand binding to the β-adrenergic receptor
of turkey erythrocytes at 37°C.[a]

	$\Delta G°$ (kcal mol^{-1})	$\Delta H°$ (kcal mol^{-1})	$\Delta S°$ (entropy units)
Agonists			
(−)isoprenaline	−9.39	−13.39	−12.9
(−)noradrenaline	−7.91	−18.86	−35.3
(−)adrenaline	−7.50	−12.75	−16.9
Partial agonists			
Soterenol	−8.23	−7.84	+1.26
Metaproterenol	−6.78	−10.83	−13.06
Terbutaline	−6.19	−4.13	+6.65
Antagonists			
(−)propranolol	−12.51	−3.85	+27.9
Pindolol	−11.85	−5.08	+21.8
Zinterol	−9.13	−3.06	+19.6
Metoprolol	−8.36	−0.66	+24.8
Sotalol	−8.21	−2.15	+19.5
Practolol	−7.46	+3.92	+36.7

[a] From Ref. 6.

The thermodynamics of transfer of the -CH$_2$-group into cyclohexane and water are given in Table V. The two processes are enthalpic and entropic controlled processes, respectively, but the transfer between the two phases produces net changes of particular simplicity, the favorable $\delta\Delta G$ having approximately equal enthalpic and entropic contributions. The position holds for a number of nonaqueous solvents [24]. This relationship will be utilized for estimating nonpolar substituent transfer between water and a hydrophobic environment.

For thermodynamic data on transfer of polar moieties from water to a nonaqueous environment, partition data using a suitable hydrocarbon solvent model may similarly be made [25]. In the case of hydrogen bond interaction with the receptor, the most suitable model may be to mimic the hydrogen bond interaction with a suitable solvent of appropriate bond strength, so that nonspecific effects of other groups may be reasonably estimated. In the case of practolol where hydrogen bond proton donor interaction with the receptor from the p-acylamino group has been demonstrated, a suitable model for phase transfer is a long-chain ester, the receptor bonding of the proton donor being of similar strength on the free energy scale.

Tables VI and VII show predictions of adrenaline from isoprenaline and of propranolol from practolol allowing for simple phase transfer of the substituent groups. Self consistent results are obtained, the observed ΔH for practolol and propranolol being dominated by loss of hydration on the practolol amidic carbonyl moiety. The observed free energy difference referenced to the nonaqueous ester model is 0.5–0.6 kcal at 37°C.
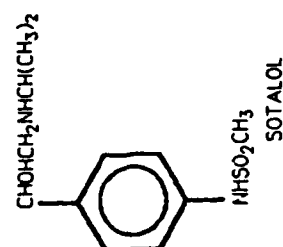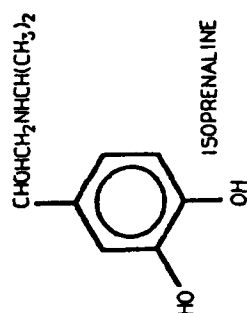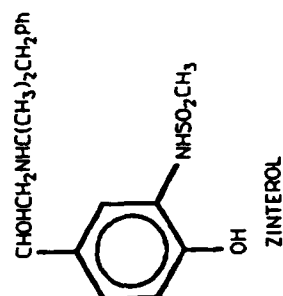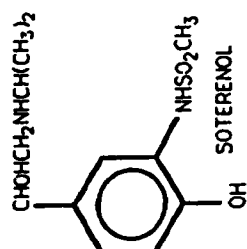
TABLE IV

OCH$_2$CHOHCH$_2$NHCH(CH$_3$)$_2$
PROPRANOLOL

OCH$_2$CHOHCH$_2$NHCH(CH$_3$)$_2$
NHCOCH$_3$
PRACTOLOL

CHOHCH$_2$NHCH(CH$_3$)$_2$
OH
META-PROTERENOL

CHOHCH$_2$NHC(CH$_3$)$_2$
OH
TERBUTALINE

CHOHCH$_2$NH$_2$
OH
NOR ADRENALINE

CHOHCH$_2$NHCH$_3$
OH
ADRENALINE
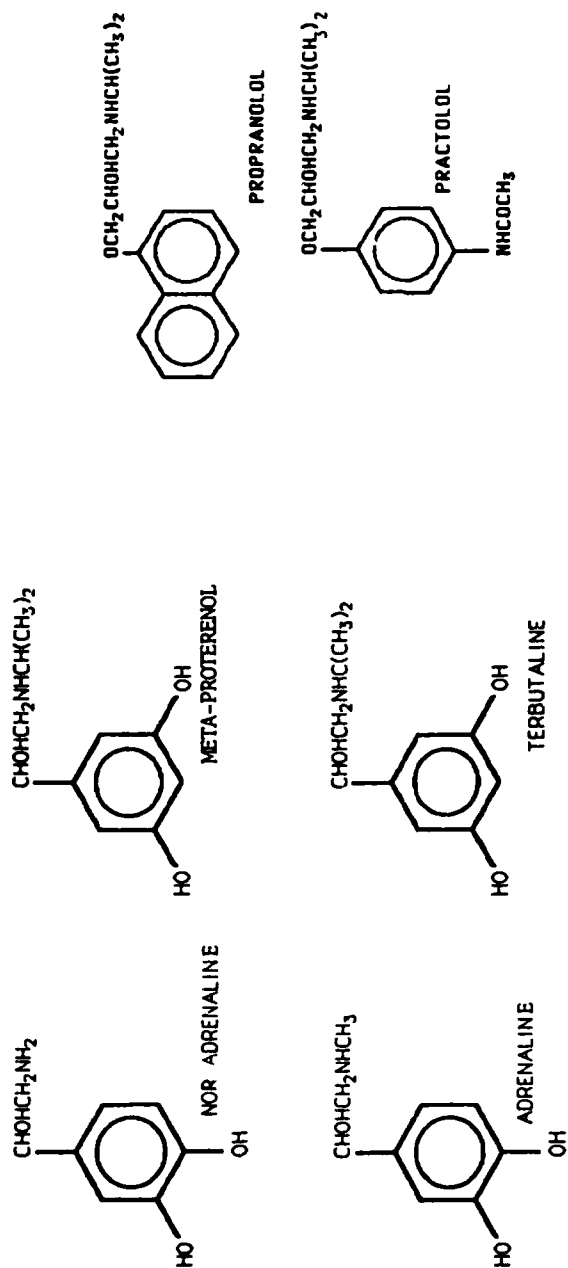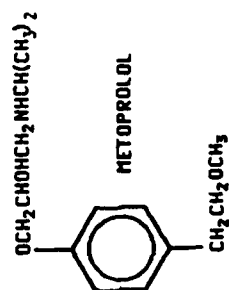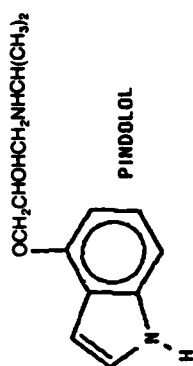
TABLE V. Incremental thermodynamics of partitioning of the $-CH_2$-group.[a]

| Partitioning | 310°K | Kcal | |
Phases	$\delta\Delta G$	$\delta\Delta H$	$-T\delta\Delta S$
1. Cyclohexane/gas	-0.76	-1.12	+0.36
2. H_2O/gas	+0.18	-0.67	+0.85
Cyclohexane/H_2O	-0.94	-0.45	-0.49

[a]From Ref. 24

TABLE VI. Thermodynamic changes on alkylamino substitution in phenethanolamines.

| | 310°K | Kcal | |
	ΔG°	ΔH°	$-T\Delta S^\circ$
Isoprenaline	-9.39	-13.39	-4.00
$\delta(CH(CH_3)_2-CH_3)$[a]	-1.70	-0.85	-0.85
Adrenaline prediction	-7.69	-12.54	+4.85
Experiment	-7.50	-12.75	+5.24
Error	+0.19	-0.21	+0.39
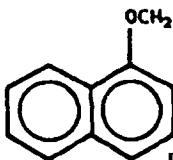
[a]Group contribution based on cyclohexane/H_2O partitioning

Where self-consistency in prediction is not obtained, comparative differences should highlight differences in binding requirements of the ligand to the receptor. Table VIII shows differences affecting progressive alkyl substitution on the amine moiety in phenethanolamines. There is an enthalpic loss of 6.5 kcal in going from noradrenaline to adrenaline, and a further loss of 6.5 kcal in going from isopropylamino to t-butylamino substitution. The free energy losses are also similar of ~1.5 kcal. This behavior is in striking contrast to phenoxypropanolamine derivatives (Fig. 8), where the moieties H, CH_3, $CH(CH_3)_2$, $C(CH_3)_3$ have little effect upon the potency when referenced to a hydrocarbon environment.

Table IX shows an approximate estimate of differences in bonding when predicting a phenethanolamine derivative from phenoxypropanolamine data. While the predictions require experimental verification, the deviation is of a similar order, giving $\delta\Delta H$ of 6–7 kcal.

One possibility is that the amine moiety is sited differently in phenethanolamine and phenoxypropanolamine binding, and is multiply hydrogen bonded to three receptor sites in the natural hormone. Phenoxypropanolamines in contrast would appear to be dominantly singly bonded to the receptor. If the agonist activity is primarily attributable to the siting of the amine moiety, then this group would be sited differently for antagonist action in phenoxypropanolamines. A possible hypothesis is given in Figure 9 where phenoxypropanolamines are compared to a fixed agonist isomer of a phenethanolamine [20]. The alternative hypothesis is that similar conformations exist for antagonist and agonist action, and only a subset of the conformer space is available for triggering agonist activity. Current work on the problem may be examined.

TABLE VII. Self-consistency of thermodynamic changes in Phenoxybropanolamines.

	310°K $\Delta G°$	kcal $\Delta H°$	$-T\Delta S°$
Practolol	−7.46	+3.92	−11.40
−p-NHCOCH₃	−2.72	−6.6ᵃ	+3.9ᵃ
OCH₂CHOHCH₂NHR₁ Prediction (1)	−10.2	−2.7	−7.5
Entropy correction for specific- NHCO conformation	0.42		0.42
OCH₂CHOHCH₂NHR₁ Prediction (2)	−10.6	−2.7	−7.9
	−2.06	−1.03	−1.03
OCH₂CHOHCH₂NHR₁ Prediction	−12.7	−3.7	−8.9
Experiment	−12.5	−3.85	−8.65

ᵃ Estimate.
$R_1 = CH(CH_3)_2$.

Molecular Models of Stimulus Action and the Black and Leff Operational Model

Using the definitions of Sections 2 and 3, and under given conditions, the efficiency of a partial agonist ϵ_B can be written [1]

$$\frac{\epsilon_B}{1 - \epsilon_B} = \tau_B \frac{K^{ijx} f^i f^{jx}}{K_B} \tag{19}$$

where K^{ijx} is the conformer binding constant of the agonist conformer (the assumption is made that a single conformer controls agonist activity), and K_B is the binding constant of the drug B.

TABLE VIII. Thermodynamic changes on alkylamino substitution in phenethanolamines II.

	310°K $\Delta G°$	kcal $\Delta H°$	$-T\Delta S°$
a) Noradrenaline ⟶ Adrenaline			
Noradrenaline	−7.91	−18.86	+10.9
+δ(CH₃ — H)	−0.84	−0.42	−0.42
Adrenaline Prediction	−8.7	−19.3	+10.5
Experiment	−7.5	−12.75	+5.2
Δa	+1.2	+6.5	−5.3
b) Metaproterenol ⟶ Terbutaline			
Metaproterenol	−6.78	−10.83	+4.05
+δ(CH₃ — H)	−0.86	−0.43	−0.43
Terbutaline Prediction	−7.64	−11.26	+3.62
Experiment	−6.19	−4.13	−2.06
Δb	+1.5	+7.1	−5.7
c) Soterenol ⟶ Zinterol			
Soterenol	−8.23	−7.84	−0.39
+δ(CH₂φ — H)	−3.43	−1.72	−1.72
Zinterol Prediction	−11.86	−9.56	−2.11
Experiment	−9.13	−3.06	−6.08
Δc	+2.7	+6.5	−4.0

If the fraction of receptor conformers f^{j^x} involved in agonist binding is approximately constant for the close set of drug analogues under comparison, Eq. (19) may be written on the logarithmic scale

$$\log K_B + \log \frac{\epsilon_B}{1 - \epsilon_B} - \log \tau_B = \log K_{\blacksquare}^{ij^x} + \log f^i$$

$$\text{where } K_{\blacksquare}^{ij^x} = K^{ij^x} f^{j^x} \tag{20}$$

Using a linear model, the efficacy term reduces to log ϵ_B. For a correlation to be observed between the agonist effect, and the conformer fraction f^i of the drugs, a stringent set of conditions is required. The drug analogues must all bind in the same way (K_B constant), and create a stimulant effect under equivalent conditions (K^{ij}, τ_B constant).

One way in which such conditions might be attained is by entropic restriction in a set of close analogues. A possible example is given in the series of β-adrenoceptor agents in Table X [26]. These phenoxypropanolamine derivatives which show identical binding constants when referenced to a nonpolar hydrocarbon environment have a stimulus effect controlled by the effective size of the ortho substituent. The data are based on a sample of four rats, the maximal incremental effect on heart rate due to β-adrenoceptor stimulation being ~230beats/minute. It may be noted that the change in electron distribution in the derivatives R = H, OCH₃, CH₃, C₂H₅ are relatively minor compared to the variation in stimulus effect.

TABLE IX. Differences between phenethanolamines and phenoxypropanolamines.

	310°K $\Delta G°$	kcal $\Delta H°$	$-T\Delta S°$
OCH$_2$CHOHCH$_2$NHR$_1$ (phenyl)	-10.6	-2.7	-7.9
$-\delta(OCH_3(Aryl)-H)^a$	$\underline{\sim 0.17}$	$\underline{+0.18}$	$\underline{-0.34}$
CHOHCH$_2$NHR$_1$ (phenyl) Prediction (1)	-10.8	-2.5	-8.2
CHOHCH$_2$NHR$_1$ (phenyl, NHSO$_2$CH$_3$)	-8.21	-2.15	-6.05
$-\delta(NHSO_2CH_3-H)$	-2.45	-5.94^b	$+3.5^b$
CHOHCH$_2$NHR$_1$ (phenyl) Prediction (2)	$\underline{-10.7}$	$\underline{-8.09}$	$\underline{-2.55}$
$\delta(2)-(1)$	0.1	+5.6	-5.6

^{a}(Hydrocarbon/H$_2$O model).
^{b}estimate.
$R_1 = CH(CH_3)_2$.

The possible conformations of the phenoxypropanolamines are shown in Figure 10. The intramolecular hydrogen-bonded species I, IV, and II, V have the amine moiety quite closely coincident with the expected position found in phenethanolamine derivatives. Table XI shows conformer populations of the protonated species in D$_2$0 [11]. A potential candidate as the arbiter of the stimulus effect is conformer I, IV, the
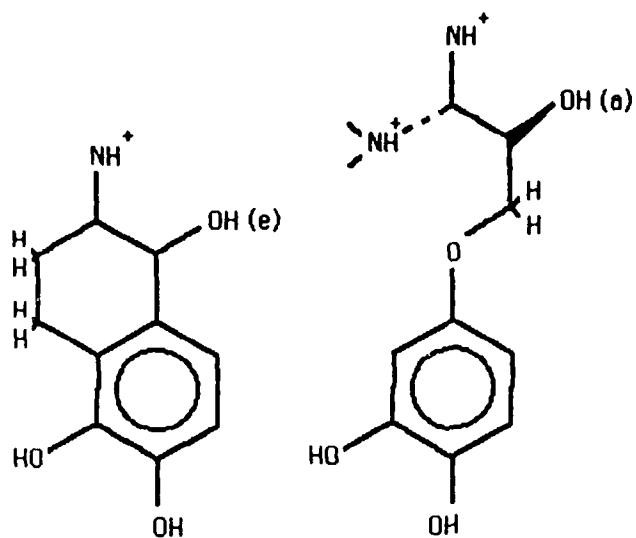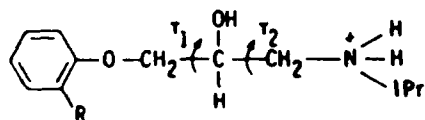
Figure 9. Comparison of a fixed active isomer of a phenethanolamine [20] with phenoxypropanolamine conformers. The dotted line of the phenoxypropanolamine conformer denotes siting of the amine moiety closely coincident to the position in phenethanolamines, and a potential stimulant conformer.

TABLE X

R Substituent	Intrinsic sympathomimetic activity, heart rate (beats/min)
H	104 ± 7
CH$_3$	65 ± 11
OCH$_3$	101 ± 7
C$_2$H$_5$	29 ± 7
CH$_2$CH:CH$_2$	32 ± 5
COCH$_3$	74 ± 5
OCH$_2$CH:CH$_2$	56 ± 5
CF$_3$	90 ± 5
NO$_2$	66 ± 5
F	117 ± 2
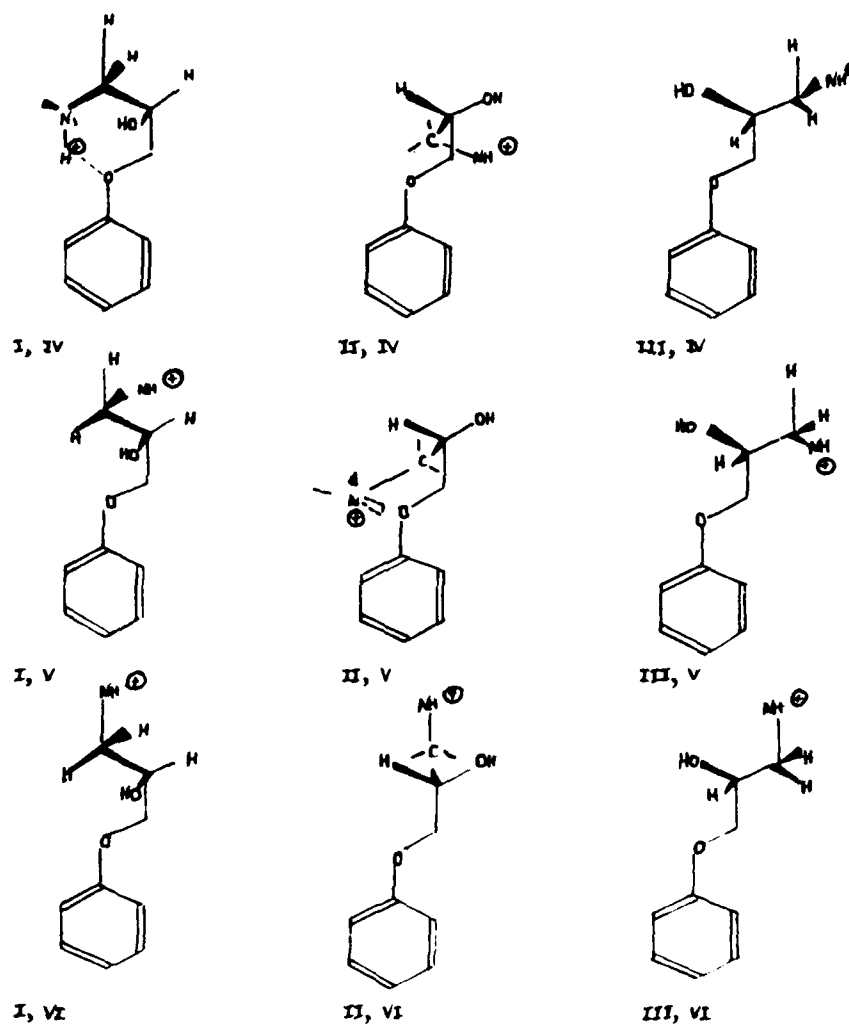
Figure 10. Primary conformers of phenoxypropanolamines.

TABLE XI

			Populations in D_2O/DCl (NMR)			
	F	H	CH_3	$CH_2CH:CH_2$	$C_2H_4(1)$	$C_2H_4(2)$
P I, IV	0.10	0.15	0.07	0.07	0.05	0.07
P I, VI	0.45	0.39	0.46	0.45	0.49	0.46
P II, V	0.12	0.01	0.12	0.09	0.09	0.06
P II, VI	0.02	0.11	0.03	0.09	0.03	0.09
P III, VI	0.34	0.34	0.33	0.31	0.33	0.32

likely contender for the antagonist conformer being I, VI, which has an approximately invariant conformer population within the set of analogues. This latter conformer has previously been proposed from X-ray studies on phenethanolamine and phenoxypropanolamine derivatives [21].

Individual conformer populations were evaluated, taking into account the coupling between the dihedral angles due to the intramolecular hydrogen bonding, utilizing the five most probable conformers [27]. Even so, variations between conformer populations assuming dihedral coupling or independence were quite minor. Most variation was shown in the coupling constant for the $R = C_2H_5$ derivative, and these measurements are shown separately in Table XI.

The model for evaluating conformer populations is not, however, complete for correlation studies on the binding constants of this set of analogues, as stated above, shows that the *ortho* substituents are surrounded by hydrophobic interactions in these drug-receptor complexes. From Eq. (18), the relevant conformer fractions in aqueous and nonaqueous phases (12) are given by the relation

$$f_L^i P = f^i P^i$$

and the ratio of the gross (P) to conformer partition coefficient (P^i) is required. In the intramolecular hydrogen-bonded species I, IV, the conformer partition coefficient P^i is sensitive to electronic changes in *ortho* substitution with the direct electrostatic effect on the protonated amine moiety and would require experimental model isomers for estimation. We have, therefore, considered it worthwhile to evaluate f_L^i theoretically by computing the intrinsic relative strengths of the intramolecular hydrogen bond I, IV on the free energy scale. Preliminary studies on the problem evaluating the partition function by direct integration techniques on intermolecular hydrogen bonding in sterically hindered phenols showed the methodology to be feasible [28].

7. Conclusions

The ability to derive useful structural information from data on drug-receptor interactions when details of the receptor-protein complex are not known has been exemplified in intracellular and cell membrane receptor applications.

The representational model of a binding constant as a sum of individual conformer binding constants weighted by the relevant conformer fractions of drug and receptor highlights one of the difficulties in working with partial information. An inherent difficulty is that the relevant conformer j^x of the receptor interacting with the ith conformer of the drug is unknown. Thus in maximizing the effective binding by increasing the conformer binding constant K^{ij^x} and the relevant conformer fraction f^i of the drug, no information is gained on increasing the relevant receptor conformer f^{j^x}. A random as opposed to a logical change in structure of the drug might prove more effective by increasing f^{j^x}, and the development of an improved structure. It is at least logical in the design of a drug to increase K^{ij^x} and f^i as efficiently as possible, in order that the number of random changes that may be introduced can be maximized for a given effort. On the more optimistic side, it is of course possible to bring a potential drug compound to within therapeutic limits by the use of such models.

The methodologies for doing so to achieve quantitative predictions appear reasonably established.

References

[1] E. J. Ariens, Arch. Int. Pharmacodyn. **99**, 32–49 (1954).

[2] D. R. Waud, Pharmacol. Rev. **20**, 49–88 (168).

[3] H. P. Rang, Br. J. Pharmacol. **48**, 475–486 (1973).

[4] R. A. Ferguson, H. H. Robertson, A. J. Bilski, and J. L. Wale, Clin. Exp. Pharm. Physiol. **6**, 21–29 (1979).

[5] D. Schulster and A. Levitzki, Eds., *Cellular Receptors* (J. Wiley and Sons, New York, 1980).

[6] G. A. Weiland, K. P. Minneman, and P. B. Molinoff, Nature **281**, 114–117 (1979).

[7] G. A. Weiland, K. P. Minneman, and P. B. Molinoff, Mol. Pharm. **18**, 341–347 (1980).

[8] M. L. Contreras, B. B. Wolfe, and P. B. Molinoff, J. Pharm. Exp. Ther. **237**, 154–164 (1986).

[9] M. L. Contreras, B. B. Wolfe, and P. B. Molinoff, J. Pharm. Exp. Ther. **237**, 165–172 (1986).

[10] J. W. Black and P. Leff, Proc. Roy. Soc. Lond. **B 220**, 141–162 (1983).

[11] R. H. Davies, K. J. Broadley, and D. R. Kelly, J. Med. Chem. (in preparation).

[12] R. H. Davies, B. Sheard, and P. J. Taylor, J. Pharm. Sci. **68**, 396–397 (1979).

[13] R. H. Davies, Int. J. Quant. Chem. **QBS4**, 413–442 (1977).

[14] G. M. Anderson, P. A. Kollman, L. N. Domelsmith, and K. N. Houk, J. Am. Chem. Soc. **101**, 2344–2352 (1979).

[15] I. H. Hillier, A. D. Patterson, and M. Scanlan, unpublished results.

[16] J. D. Mersh and J. K. M. Sanders, Tetrahedron Lett. **22**, 4029–4032 (1981).

[17] A. Wolf, W. Voets, and H-H Schmidtke, Theor. Chim. Acta. **54**, 229–238 (1980).

[18] A. T. Glen, L. R. Hughes, J. J. Morris, and P. J. Taylor, 3rd SCI-RSC Med. Chem. Symp. (ed. R. W. Lamberts, Chem. Soc. London 1986) p. 345–361.

[19] D. N. Middlemiss, personal communication.

[20] M. Nishikawa, M. Kanno, H. Kuriki, H. Sugihara, M. Motohashi, K. Itoh, O. Myashita, Y. Oka, and Y. Sanno, Life Sci. **16**, 305–314 (1975).

[21] H. L. Ammon, A. Balsamo, B. Macchia, F. Macchia, D.-B. Howe, and W. E. Keefe, Experientia 644–646 (1975).

[22] J. Zaagsma, J. Med. Chem. **22**, 441–449 (1979).

[23] A. Balsamo, A. Lapucci, B. Macchia, and A. Martinelli, J. Med. Chem. **28**, 153–160 (1985).

[24] M. H. Abraham, J. Am. Chem. Soc. **104**, 2085–2094 (1982).

[25] W. Riebesehl, E. Tomlinson, and H. J. M. Grunbauer, J. Phys. Chem. **88**, 4775–4779 (1984).

[26] R. H. Davies and L. H. Smith, Int. J. Quant. Chem. **QBS7**, 331–345 (1980).

[27] V. Barone, F. Lelj, and G. Abbate, Magn. Res. Chem. **23**, 715–719 (1985).

[28] R. H. Davies, R. C. Mason, D. A. Smith, D. J. McNeillie, and R. James, Int. J. Quant. Chem. **QBS5**, 221–243 (1979).

The Search for Active Substructures in Structure-Activity Studies*

MILAN RANDIĆ

Department of Mathematics and Computer Science, Drake University, Des Moines, Iowa 50311; and Ames Laboratory-DOE,[†] Iowa State University, Ames, Iowa 50011, U.S.A.

BORKA JERMAN-BLAŽIČ

Institut Jožef Štefan, 61001 Ljubljana, Slovenia, Yugoslavia

DENNIS H. ROUVRAY

Department of Chemistry, University of Georgia, Athens, Georgia 30602, U.S.A.

PAUL G. SEYBOLD

Department of Chemistry, Wright State University, Dayton, Ohio 45435, U.S.A.

STUART C. GROSSMAN

Omaha North High School, 37th and Ames Avenue, Omaha, Nebraska 68111, U.S.A.

Abstract

We consider the problem of locating an active fragment (substructure) common to a class of biologically active compounds and presumed responsible for their biological activity (therapeutic or toxic). Our approach is graph-theoretical in that molecules are represented by suitable graph-theoretical invariants. Specially weighted paths in the molecular graph are adopted as descriptive elements. By selecting different sets of atoms one searches for a fragment that best represents the relative activities of the compounds. As an illustration we consider a dozen nitrosamine mutagens and analyze the cases of five-, six-, and seven-atom fragments. The approach clearly indicates that a specific seven-atom fragment (for molecules with up to 11 nonhydrogen atoms) can account for the relative mutagenic activities of the nitrosamines considered.

Introduction

The importance of structure–activity relationships (SARs) is generally accepted today, and considerable effort is now aimed at elaborating the developing schemes which can lead to quantitative characterizations and predictions of activity. The key to the SAR problem has always been the characterization of chemical structure. For convenience, existing approaches to the problem may be classified as: (a) *structure cryptic*, namely approaches in which limited structural information is explicitly used; (b) *structure implicit*, schemes such as quantum mechanical models, which yield

---

information not directly interpretable in terms of structural components; and (c) *structure explicit;* schemes which make sole use of structural information in the comparison of results [1]. Graph-theoretical approaches are all structure explicit because one starts by selecting appropriate structural (graph-theoretical) invariants and then uses these in making comparative studies. Here we continue to explore and develop graph-theoretic methodology and will show not only how selected graph invariants can be used to represent molecules (chemicals), but also how they can be employed for the characterization of *local* molecular features. In particular, we focus attention on the search for substructures that can be identified as responsible for molecular activity in families of closely related structures.

A number of active molecular fragments have been identified empirically. For example, the so-called "morphine" rule [2] defines a fragment found in various morphine-related molecules, and similarly dopamines appear to have a common, characteristic grouping of atoms. Inspection of molecular diagrams sometimes suggests such fragments and visual verification suffices to establish them. However, in other instances, the fragments may be unknown or largely obscured, and in need of identification. We now outline a graph-theoretical procedure that identifies potentially active molecular fragments or substructures within families of structurally related compounds displaying similar therapeutic (toxic) features. In the next section we first discuss the graph-theoretical representation of structures adopted here, and then follow this with an illustration of the search procedure for an active substructure in a set of nitrosamine mutagens.

Graph-Theoretical Representation of a Structure

Mathematical problems are frequently solved by adopting suitable coordinates, and sometimes the crucial step in solving a problem involves the proper choice of coordinates. Conversely, a poor selection of coordinates can lead to tedious calculations and frustration. A prudent selection of graph invariants can similarly facilitate structure activity analyses, though even these analyses can become time-consuming and unproductive if the molecules studied are represented by inappropriate descriptors. Experience suggests that certain graph invariants are likely to be useful representatives of molecular structure. In particular, we may mention the Wiener number, W, which represents the count of all paths in a structure [3]; the Hosoya number, Z, which totals the number of nonadjacent bonds in a structure, and as such represents the first graph-theoretical index designed to represent a structure by a *single* number [4]; and, finally, the connectivity index, χ [5], introduced by one of the present authors [6] as a bond-additive quantity which differentiates contributions from bonds of different type. More recently, additional descriptors were introduced, for example, the molecular identification number which represents the sum of suitably weighted molecular paths [7]. This number is analogous to the Wiener number, but employs weights shown to be useful in applications of the connectivity index, χ.

The limitations of a single-parameter representation of structure are obvious, even to those not familiar with the field of structure-property studies. What is *surprising,* and sometimes overlooked, however, is *how much* structural information can often be condensed into a single number. Nonetheless, if one wants to go beyond a certain

level, the next logical step is to consider several parameters (invariants). The first issue to be settled here is the selection of the invariants. That atomic and bond contributions play a dominant role in most applications is self-evident, but what should the next contribution be? Nonbonded interactions? Next nearest neighbors? Bond–bond interactions? While each of these plays a significant role to a differing extent, such individual improvements suffer from the same defect, namely that, if they cannot account for all the deviations observed, the question again arises as to what should be considered next. If one wishes to develop a systematic multiparametric approach to SAR, two basic possibilities present themselves: one considers either an *ordered sequence* of numbers, or a *collection* (a set) of numbers (parameters). Each choice has certain advantages, and we now examine these two alternatives.

However, even when one considers such a route, we recommend that the invariants (parameters) be *naturally* ordered, so that one can successively introduce higher terms as needed and eliminate those terms that are not warranted. One of the advantages of the connectivity index is that it allows such a natural extension [8]. One can consider two-bond fragments, three-bond fragments, and so on, and introduce for each the corresponding (higher) connectivity value. These indices can subsequently be used as a sequence or set of parameters to improve the correlations. Such a generalization of an index is lacking in the case of both the Hosoya, Z, and Wiener, W, numbers, as pointed out in the literature [8].

More generally, one can select paths of different length and use *path numbers, p_k,* (which designate the number of paths of length k in a molecule) as molecular descriptors [9]. Although the suggestion that path numbers may be useful molecular descriptors can be traced back to the work of Platt [10], it was not widely recognized or adopted. Paths are defined as self-avoiding walks, that is, as sequences of vertices and incident edges, that cannot be repeated. There are alternative ways of characterizing structure, for instance, the use of self-returning or random walks rather than paths. A random walk on the other hand allows one to use the same edge and vertex several times. An advantage of random walks (and self-returning random walks, which are walks that start and end on the same atom) is that they are easy to compute — they are given by the entries of A^k (the k-th power of the adjacency matrix) — but they tend to grow rapidly with molecular size. In contrast, path numbers do not grow so fast, though their computation (except in the case of very small and simple graphs) is tedious and requires computer use. A program called ALLPATH is available [10] and gives results for most molecules of chemical interest rapidly. The problem of a path count is inherently NP-hard, that is, it grows exponentially (nonpolynomially, hence the abbreviation NP) with the growth of the size of the graph (measured in terms of the number of vertices, n). Thus, in principle, for sufficiently large graphs, a path count is bound to become impractical. However, for molecules having up to a half a dozen rings and some 50–150 atoms (which covers a lot of chemistry!), the count is usually relatively fast. Accordingly, we shall continue to use path numbers as our basic approach to structure description. The following characteristics suggest the virtues of path numbers:

Apparently (pictorially) similar structures lead to path sequences that are visually and analytically similar [11];

Naturally occurring compounds (such as the terpenes) show greater mutual similarity than they do to artificially constructed structures having the same building blocks (isoprene units) [11];

Structures of similar biological activity show similarities when ordered according to selected (most active) standards [12, 13].

There are also certain drawbacks to the use of path numbers for the characterization of molecules. If molecules have a similar outside shape (periphery), but differ significantly in their internal structure (number of rings or bridges), they will have vastly different counts of paths, even though they possess similar properties. An example is provided by civetone, a macrocyclic musk and a sterol that possesses a decidedly musk-like odor as observed by Prelog and Ružicka [14].
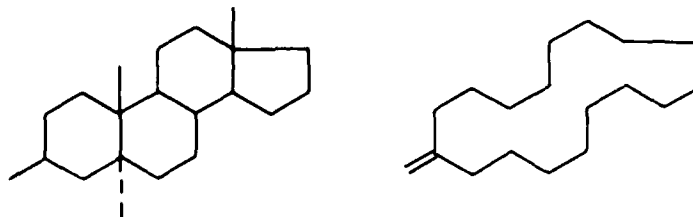
Both have a similar periphery and a similar musk odor, but this similarity is not detectable using an approach based on count of paths.

This kind of problem prompted the need to introduce weighting factors, which can reduce the role of some paths and thereby enhance the significance of others. Comparison of selected physicochemical properties of alkanes has shown that short paths, namely paths of lengths two and three (p_2 and p_3, respectively), play a dominant role in dictating the relative magnitudes and variations in properties among isomers [15–17]. The connectivity index, which focuses on paths of length one (p_1 or bonds), has been shown to be successful in correlating structure with property in the form of a single-number representation [18, 19]. This index is based on special weightings: bonds are classified as (m, n) bond types, where m and n indicate the number of neighbors (valency) for the atoms forming the bond. The bond (m, n) is assigned the weight $1/(m \times n)^{1/2}$, and the contributions of different bonds are added together to yield the molecular connectivity index.

In Table I we show a preselected order of alkane isomers having 4–6 carbon atoms, based on the condition that the numerical values for bond contributions reproduce the ordering for a selected property. The order selected in the order of relative magnitudes for a number of thermodynamic molecular properties, including the boiling points. Hence, it should not be surprising that the connectivity index gives very good correlations — it has been constructed to do just this by requiring that the ordering be reproduced. If the ordering is reproduced, as Heilbronner and Schmelzer have pointed out [20], a relatively high correlation is possible, even with arbitrarily selected parameters! The particular weights $1/(m \times n)^{1/2}$ represent one *solution* to the inequalities, and therefore we here extend their use to weighting paths. The modified program ALLPATH, which from the input list of atomic neighbors gives the count of

TABLE I. Decomposition of bonds in smaller alkanes into various (m, n) bond types and associated inequalities imposed by requiring that the ordering of the isomers follow the ordering of selected thermodynamic properties (the weighting rule $1/(m \times n)^{1/2}$ is a solution to the inequalities).

$$2(1,2)+(2,2) > 3(1,3)$$

$$2(1,2)+2(2,2) > (1,2)+2(1,3)+(2,3) > 4(1,4)$$

$$2(1,2)+3(2,2) > 2(1,2)+(1,3)+2(2,3) > (1,2)+2(1,3)+(2,2)+(2,3) >$$

$$4(1,3)+(3,3) > (1,2)+3(1,4)+(2,4)$$

weighted paths, is available [10] and suitable for applications in SAR. To each bond in a graph the weight $1/(m \times n)^{1/2}$ is assigned, and whenever a path contains that bond its value is multiplied by this factor. Because all such factors are less than one and they are used repeatedly, the scheme drastically reduces the values for path numbers of longer paths, giving more prominence to shorter paths. This appears to be a more natural way to diminish the dominant role of paths of intermediate length (which are the most numerous and tend to obscure the role of shorter paths) than mere truncation.

An Illustration

In Table II we give the computed output of weighted path counts for the molecule methyl-2-oxypropylnitrosoamine (MOP), one of the nitrosoamines that we examine in the next section:
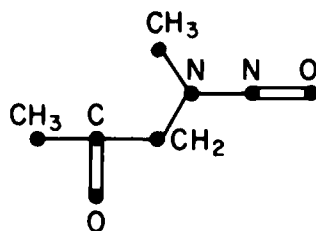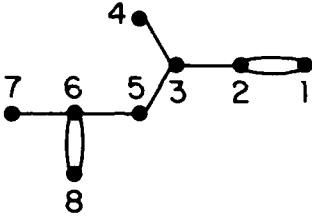
TABLE II. Results of the program ALLPATH modified to include the weighting of bonds as $1/(m \times n)$[a]

Atom	P_0	P_1	P_2	P_3	P_4	P_5	Atomic ID
1	1	0.8164	0.2721	0.2682	0.0392	0.0474	2.4436
2	1	1.1498	0.3285	0.0481	0.0580		2.5845
3	1	1.3189	0.4165	0.1742			2.9096
4	1	0.5773	0.4281	0.2404	0.1005		2.3465
5	1	0.7618	0.7985	0.1111			2.6714
6	1	1.5606	0.1443	0.1314	0.03928		2.8757
7	1	0.5000	0.5303	0.0721	0.0657	0.0196	2.1878
8	1	0.7071	0.6035	0.1020	0.0929	0.0277	2.5334
Molecule:	8	3.6960	1.7610	0.5739	0.1979	0.0474	14.2764

Atom Count	Molecular Connectivity	Higher Connectivities[b]		Molecular ID

[a] Each row gives path numbers for the indicated atom. The sum of the entries in each row gives the atomic identification (ID) number. Summation along individual columns (and dividing the result by two, except for the first column) gives molecular path numbers. The first two members in the molecular path sequence give the number of atoms and the connectivity index, respectively.

[b] Slightly differently defined that in the original reference [8].

Among the molecules considered, this is the most potent mutagen [21]. Our interest will center primarily on *atomic ID* numbers, which represent sums of all weighted paths to individual atoms (when nonweighted paths are used an analogous index has been considered by Balaban [22] and Seybold [23,24]). In Table II the atomic ID values are obtained by adding the entries in each row (with each row corresponding to one atom). If we add numbers in each column we obtain the sequence of numbers shown at the bottom of Table II. Observe that the first entry (i.e., the sum of the first column) is the value of the atom count and the second is the connectivity index for the molecule. Finally, if one adds all the numbers in Table II (all rows and all columns), one obtains the *molecular ID* number, indicated also in the table (as the total number of paths).

The above molecule can now be represented in one of the following ways:

Molecular ID: 14.2764
Connectivity: 3.6960
Weighted paths: 8, 3.6960, 1.7610, 0.5739, 0.1979, 0.0474
Set of atomic IDs: 2.4436; 2.5845; 2.9096; 2.3465; 2.6714; 2.8757; 2.1878;

2.5334; (Elements of set, of course, can be displayed in any arbitrary order.)

Each of these descriptors, as well as corresponding unweighted quantities, have been used in SAR applications. ID numbers have been used to define nonempirical clusterings of selected therapeutically useful drugs [25], weighted paths are used in similarity studies [26,27], and atomic ID numbers in a search for antitumor agents [28]. Here we pursue a similar theme: our search, however, is not for an optimal compound, but for a substructure responsible for the bioactivity in a family of structures. As will be seen, atomic ID numbers are well suited for such an application because they allow some flexibility in probing various molecular fragments.

Searching for Active Substructures

If the compounds considered are structurally similar and differ only modestly in composition and arrangement of their various parts, a graph-theoretical approach based on path enumeration may be appropriate for comparative study and analysis. If the molecules are of different sizes, even when the same pharmacophor is responsible for their activity, any similarity in the path counts can be obscured by the role of extraneous atoms. Clearly one then has to restrict the count of paths to common atoms in all the members of the family. But even such empirical or visual selection of common atoms may still contain irrelevant atoms whose paths reduce the signal-to-noise ratio. Often a trial-and-error approach needs to be adopted. Here we will outline the use of weighted paths in a systematic search for a molecular fragment (common within a family of compounds considered) that can be identified as the underlying, active portion of the molecules under consideration.

We have selected a set of nitrosamine mutagens (mostly propyl derivatives) because they have been previously examined and so comparison is possible with alternative approaches. Moreover, these structures are sufficiently general to illustrate the method. The compounds are shown in Figure 1, abbreviated as in Ref. 21, where interested readers can find their full chemical names and other details. The structures have been *ordered* in Figure 1 according to their relative mutagenic potencies: MOP is the most potent with a relative mutagenic activity of 650. (The data refer to mutagenic activities in the hamster hepatocyte-mediated V79 cell mutagenesis system at a concentration of 0.7 mM, and are expressed as numbers of ouabain-resistant mutants per 10^4 surviving V79 cells [21].) Table III lists atomic ID numbers for seven atoms, all numbered to overlap the numbering shown for MOP in Table II. When extending the path counts to atoms 6 and 7 there are some minor problems due to the ambiguity of alternative labelings. We have shown both alternatives, and, as discussed later, when a choice exists we have selected the one that optimizes similarity. DMN, of course, has only 5 atoms, and is excluded when considerations extend to fragments with 6 or more atoms.

One can see by inspection of Table III that values for the atomic IDs do not vary dramatically within each column, except when the atom in question has (or has not, while others have) another neighbor (substituent). Thus, all the values in the first column are roughly 2.45–2.50, the next column has somewhat larger values, 2.55–2.65, the next values lying between 2.90 and 3.35, and so on. Yet there are variations, and the nature of these path numbers suggests a small but significant role for
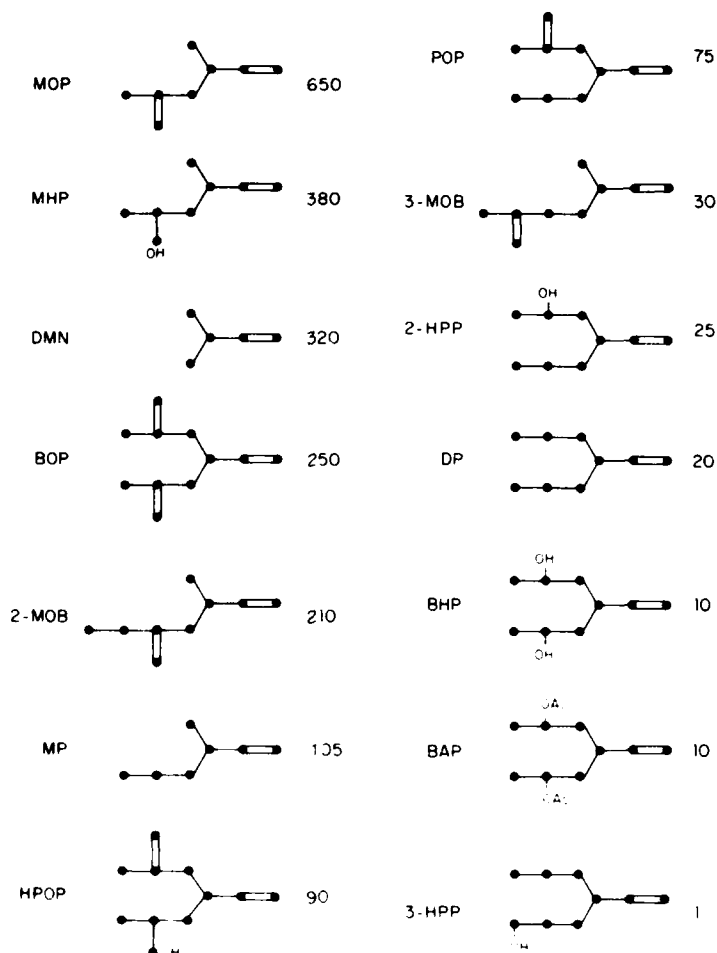
Figure 1. The simplified molecular diagrams for the 13 nitrosamines considered. Their mutagenic activities are shown by the numerical values; the abbreviations are those used for Ref. 21.

more distant neighbors. We now wish to use these minor variations as a basis for tests of relative similarity among different compounds.

Rather than examining all similarities, we shall restrict our attention here to the relative similarities of the compounds to MOP and MHP, the two compounds with the highest mutagenic activities (650 and 380, respectively). Table IV shows numerical estimates of the relative similarities of the compounds to these two standards. The entries in Table IV are obtained by viewing the collection of atomic ID numbers for each structure as a vector, and taking a Euclidean n-dimensional metric to find the distance between such vectors. Consider the "distance" (i.e., measure of similarity or dissimilarity) between MOP and MHP (restricting attention here to the first 5 atoms):

MOP   2.443   2.584   2.909   2.346   2.671

TABLE III. Atomic ID numbers for the 13 nitrosamines considered.[a]

Atom Molecule	1	2	3	4	5	6	7
MOP	2.443	2.584	2.909	2.346	2.671	2.875	2.187
MHP	2.454	2.598	2.950	2.369	2.770	2.926	2.356
DMN	2.402	2.534	2.760	2.260	2.260	—	—
BOP	2.484	2.634	3.054	2.732	2.732	2.897	2.198
2-MOB	2.447	2.584	2.924	2.355	2.708	2.979	2.547
MP	2.451	2.594	2.939	2.363	2.744	2.652	2.375
HPOP	2.495	2.647	3.099	2.749	2.831	2.903	-2.201
							2.951	-2.370
POP	2.492	2.644	3.089	2.744	2.805	2.683	2.397
								2.200
3-MOB	2.455	2.594	2.594	2.372	2.781	2.725	2.894
2-HPP	2.503	2.657	3.129	2.844	2.822	2.691	2.403
								2.373
DP	2.500	2.654	3.118	2.817	2.817	2.684	2.401
BHP	2.506	2.661	3.140	2.848	2.848	2.958	2.374
3-HPP	2.508	2.664	3.148	2.830	2.891	2.835	2.698
								2.405

[a] Only data for the 7 common nonhydrogen atoms are represented. In some cases there are two alternative choices for the 7 atoms and both alternatives are shown.

MHP   2.454   2.598   2.950   2.369   2.770

One takes the difference between each pair of values and squares it thus:

$$0.011^2 + 0.014^2 + 0.041^2 + 0.023^2 + 0.99^2 .$$

Upon adding these contributions one obtains the result 0.0123, the square root of which is 0.1110, the number used for comparison of the structures. As may be seen from Table IV (which gives such results for atoms 1-5, for atoms 1-6, and finally for atoms 1-7) all the numbers are relatively small and differ at most by an order of magnitude. Nevertheless, differences do exist, and, as will be seen, these reflect differences in the structures considered.

In order to make useful deductions from Table IV, we now examine the cases of 5-, 6-, and 7-atom fragments by *ordering* the compounds relative to the selected standards. This is, admittedly, a step in the direction from *quantitative* to *qualitative* analysis. But later on we shall revisit the data and examine them quantitatively. Ordering will give us some insights into the complexities and difficulties of correlating data and structures. In Figures 2–4 we show for the three cases (fragments having 5, 6, and 7 atoms) an ordering of the structures with respect to MOP (top row) and MHP (bottom row). After arriving at these orderings (as previously outlined in several papers on path comparison of structures in SAR [12]), one uses a line to connect the same structures. From such a diagram it is not difficult to extract all the *partial orders*, and these are shown by diagrams in which we have inserted the values of the mutagenic activities, rather than the molecular symbols (except for the case of the 5-atom fragments, where we have for the sake of clarity shown both). If all the num-

TABLE IV. Part of the similarity table for nitrosamines.[a]

5-Atom Fragment Similarity to MOP and MHP

	MOP	MHP	DMN	BOP	2-MOB	MO	HPOP	POP	3-MOB	2-HPP	DP	BHP	3-HPP
MOP	*	0.111	0.450	0.426	0.041	0.081	0.480	0.463	0.389	0.572	0.543	0.506	0.591
MHP	0.111	*	0.561	0.383	0.069	0.029	0.417	0.405	0.012	0.515	0.486	0.548	0.523

6-Atom Fragment Similarity to MOP and MHP

	MOP	MHP	DMN	BOP	2-MOB	MO	HPOP	POP	3-MOB	2-HPP	DP	BHP	3-HPP
MOP	*	0.122	—	0.427	0.112	0.238	0.481	0.502	0.417	0.602	0.574	0.513	0.593
MHP	0.122	*	—	0.385	0.087	0.276	0.418	0.473	0.201	0.567	0.541	0.549	0.530

7-Atom Fragment Similarity to MOP and MHP

	MOP	MHP	DMN	BOP	2-MOB	MO	HPOP	POP	3-MOB	2-HPP	DP	BHP	3-HPP
MOP	*	0.209	—	0.428	0.377	0.548	0.482	0.502	0.807	0.630	0.616	0.546	0.632
MHP	0.209	*	—	0.416	0.210	0.276	0.418	0.475	0.574	0.567	0.543	0.550	0.533

[a] Only similarities with respect to the two most potent mutagenic compounds, MOP and MHP, are shown. The similarity is based on comparison of atomic ID numbers for the compounds considered, using a Euclidean n-dimensional distance formula to derive the numerical similarity values.
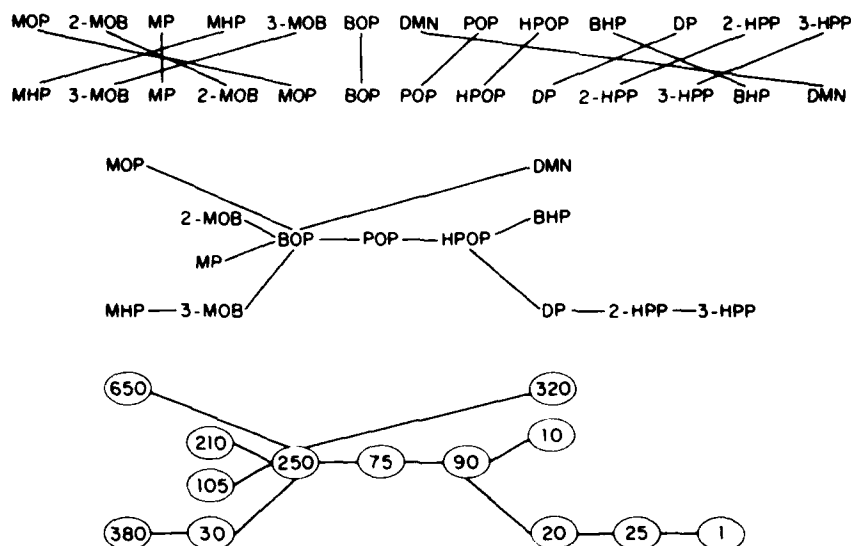
Figure 2. Ordering of the nitrosamines relative to MOP and MHP (top) and the derived partial order (constructed by first pairing same labels in the top two rows and then examining relations for which no crossing of lines occur). The bottom diagram replaces molecular species by their mutagenic activity values. Ordering is based on common 5-atom fragments.
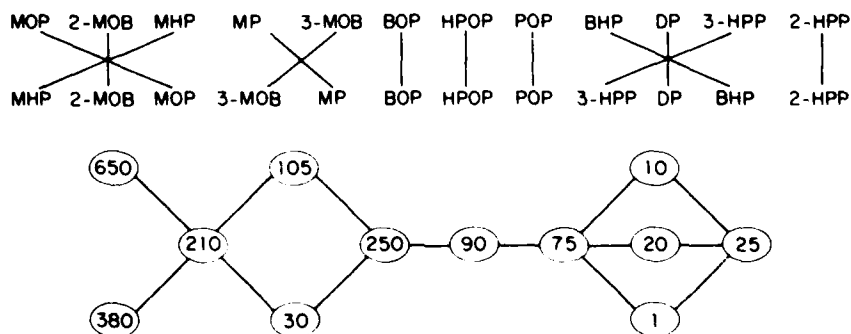

Figure 3. Ordering of nitrosamines based on 6-atom fragments and the corresponding partial order with mutagenic activity values displayed.

bers in a partially ordered diagram were to follow monotonically from left to right, we could claim that the approach points to a specific fragment as the substructure responsible for bioactivity. A look at Figure 2 is, however, disappointing: BOP (250) and DMN (320) do not fit the scheme at all, and in addition, the small value of 30 (for 3-MOB) appears "too early" in the diagram. Observe, however, that when we add on additional (sixth) atom (Fig. 3), a considerably better diagram results. Of course, we had to eliminate DMN (which has only 5 nonhydrogen atoms), but although the diagram is improved, there are still several discrepancies. Now the values
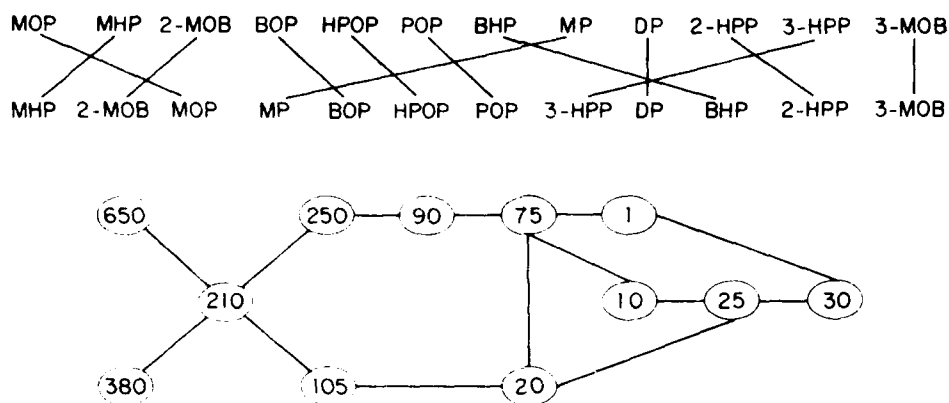
Figure 4. Ordering of nitrosamines based on 7-atom fragments and the corresponding partial order with mutagenic activity values displayed.

105 and 30 are "pushed" a bit to the right, and those of HPOP and POP (90 and 75, respectively) are ordered correctly. All this suggests that atom 6 is *relevant*, though we still cannot claim a clear cut result.

Extending considerations to a fragment having seven atoms, we obtain the diagram shown in Figure 4. It shows dramatic improvement: observe that 3-MOB with its low value (30) has been pushed to the end of the diagram and the relative orders of BOP (250) and MP (105), which were previously incorrect, have now been resolved by placing them in different branches of the lattice. Although the diagram of Figure 4 is still not "perfect" [e.g., there remains an inversion of BOP (250) and 2-MOB (210)], if we view these discrepancies as minor (recalling the considerable experimental uncertainty) we can see definite progress and claim that the approach now appears to have captured the main features of the pertinent molecular fragment responsible for biological (mutagenic) activity.

In order to see even better, and on a *quantitative* scale, the substance of the present approach, we illustrate in Figures 5–7 the relationships between the mutagenic activities (y axis) and similarities to MOP (the most potent mutagenic structure, x axis). For an ideal theoretical model one should obtain a correlation curve which decreases from the top left of the figure to bottom right. Instead, in Figure 5 we see a widely scattered set of points in the middle of the figure, indicating a failure of the assumption that a 5-atom fragment can explain the mutagenic activity. In Figure 6, when a sixth atom is taken into account, we notice some narrowing of the middle section, with the same three structures (BOP, 2-MOB, and MP) as outliers. Figure 6 shows improvement over Figure 5, but remains unsatisfactory. Finally, in Figure 7, we see an essentially valid correlation, with perhaps only 3-MOB at the end of the plot being somewhat out of line. This latter observation does not represent serious misbehavior, in part because we are primarily interested in (i) similarity to the most potent compound MOB; and (ii) the question whether activity can be traced to a fragment present in the most potent compounds. If a compound shows little similarity to the
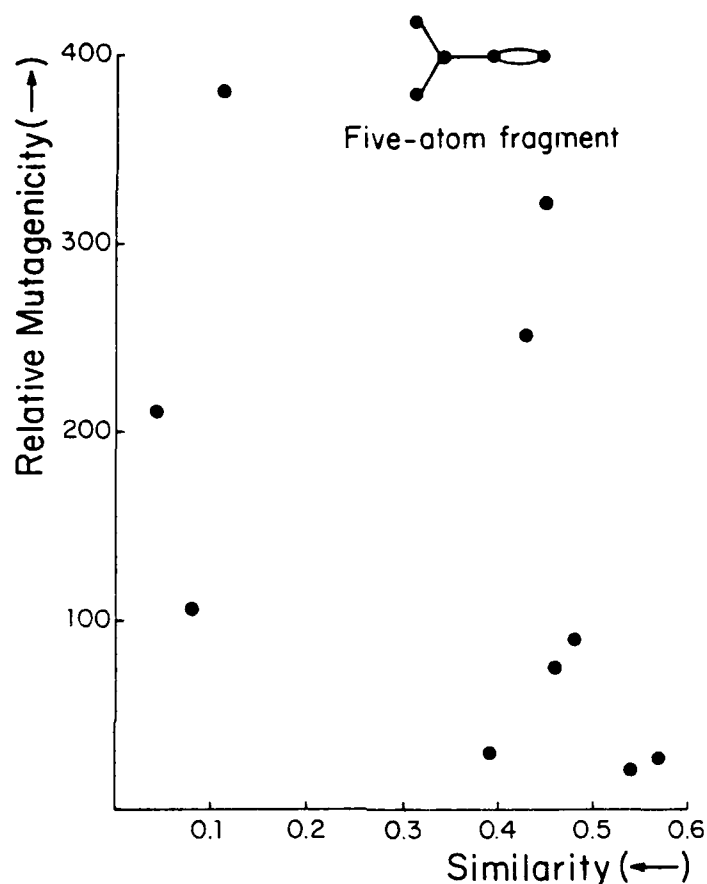
Figure 5.   Relationships between mutagenicities and similarities to MOP, based on 5-atom
fragments.

standard (as is the case with 3-MOB, which is the structure least similar to MOP) then other factors may influence its activity and cause observed deviations.

Conclusions

Figures 4 and 7 suggest that for the dozen nitrosamines considered, one can at least partly associate mutagenic potency with the presence of a particular 7-atom substructure. The limited success of other approaches, for example, use of the connectivity index (which gives a fair correlation), may well lie in the fact that the nitrosamines have superfluous atoms which may reduce the signal-to-noise ratio. We have described how a search for an active fragment can be carried out. Our outline employs nitrosamines as an illustration, but the methodology can readily be applied to other cases, and in particular to situations in which the bioactive fragment is not known. Such cases may require a trial-and-error approach, but with the accumulation of rele-
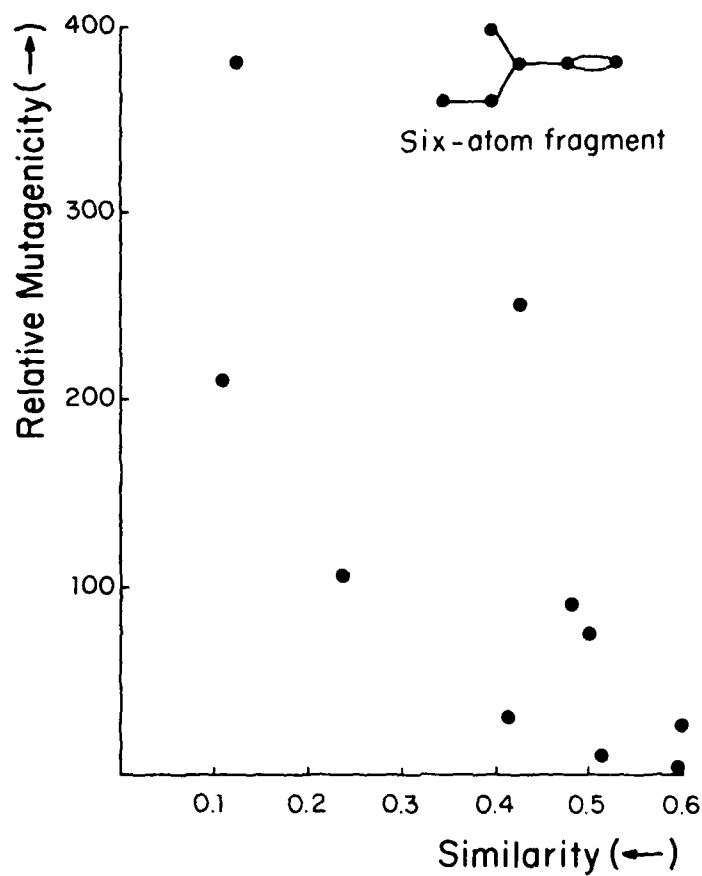
RANDIĆ ET AL.

Figure 6. Relationships between the mutagenicities of nitrosamines and their similarities to MOP, based on 6-atom fragments. Observe some narrowing of the dispersion compared to that in Figure 5.
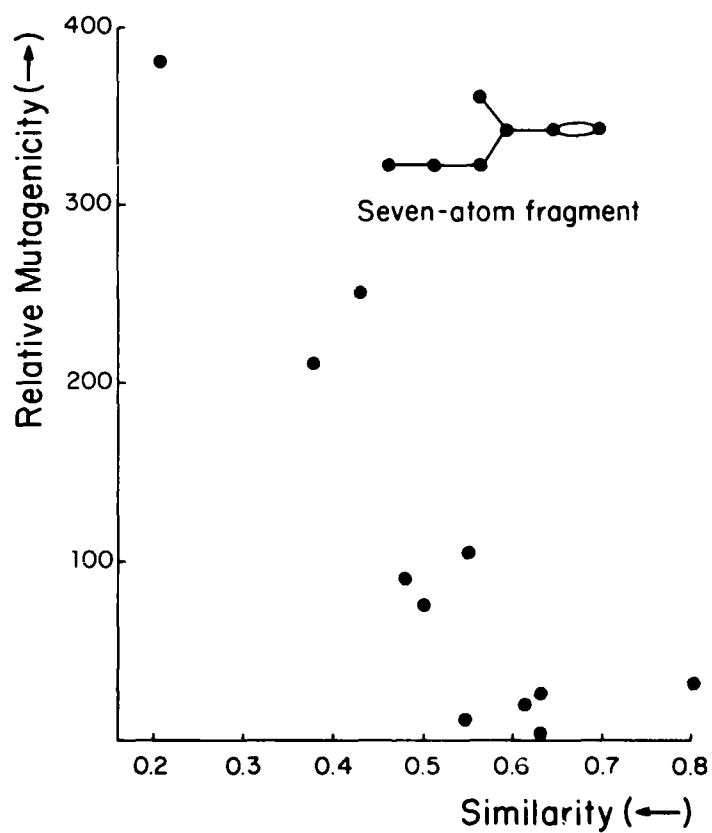
Figure 7. Correlation between the mutagenicities of the nitrosamines and their similarities to MOP, based on the 7-atom fragments. Observe the improvement in the correlation relative to that in Figures 5 and 6.

vant information it should be possible to restrict the number of candidate fragments and determine which of these is most likely to be the underlying bioactive component.

Acknowledgments

References

[1] N. Trinajstić, M. Randić, and D. J. Klein, Acta Pharm. Jugosl. 36, 267 (1986).

[2] A. Burger, *A Guide to the Chemical Basis of Drug Design*. (Wiley, New York, 1983) p. 61.

[3] H. Wiener, J. Am. Chem. Soc. 69, 17 (1947).

[4] H. Hosoya, Bull. Chem. Soc. Japan 44, 2332 (1971).

[5] L. B. Kier, L. H. Hall, W. J. Murray, and M. Randić, J. Pharm. Sci. 64, 1971 (1975).

[6] M. Randić, J. Am. Chem. Soc. 97, 6609 (1975).

[7] M. Randić, J. Chem. Inf. Comput. Sci. 24, 164 (1984).

[8] L. B. Kier, W. J. Murray, M. Randić, and L. H. Hall, J. Pharm. Sci. 65, 1226 (1976).

[9] M. Randić, Math. Chem. 7, 5 (1979).

[10] The program is written in BASIC and operated on an Apple II/e computer. The original ALLPATH program is listed in M. Randić, G. M. Brissey, R. B. Spencer, and C. L. Wilkins, Computers and Chem. 3, 5 (1979) and is also reproduced in P. C. Jurs, *Computer Software Applications in Chemistry*. (Wiley and Sons, New York, 1986).

[11] M. Randić and C. L. Wilkins, J. Chem. Inf. Comp. Sic. 19, 31 (1979).

[12] M. Randić and C. L. Wilkins, Int. J. Quant. Chem. Quant. Biol. Symp. 6, 55 (1979).

[13] B. Jerman-Blažić and M. Randić, Proc. Int. AMSE Conf., "Modelling and Simulation," Nice, Sept. (1983), 5, p. 161.

[14] V. Prelog and L. Ružicka, Helv. Chim. Acta 27, 61 (1944); 27, 66 (1944).

[15] M. Randić and C. L. Wilkins, Chem. Phys. Lett. 63, 332 (1979).

[16] M. Randić and C. L. Wilkins, J. Phys. Chem. 83, 1525 (1979).

[17] M. Randić and N. Trinajstić, Math. Chem. 13, 271 (1982).

[18] D. H. Rouvray, Sci. Am. 254(9), 40 (1986).

[19] D. H. Rouvray, Acta Pharm. Jugosl. 36, 239 (1986).

[20] E. Heilbronner and A. Schmelzer, Nouv. J. Chim. 4, 23 (1980).

[21] R. Langenbach, C. Kruszynski, R. Gingell, T. Lawson, D. Nagel, P. Pour, and S. C. Nesnow, in *Structure-Activity Correlation as a Predictive Tool in Toxicology*. L. Goldberg. Ed. (McGraw-Hill, New York, 1983) Chap. 16.

[22] A. T. Balaban, Chem. Phys. Lett. 89, 399 (1982).

[23] P. G. Seybold, Int. J.Quant. Chem.: Quant. Biol. Symp. 10, 95 (1983).

[24] P. G. Seybold, Int. J. Quant. Chem.: Quant. Biol. Symp. 10, 103 (1983).

[25] M. Randić, Int. J. Quant. Chem.: Quant. Biol. Symp. 11, 137 (1984).

[26] S. C. Grossman, B. Jerman-Blažič Džonova, and M. Randić, Int. J. Quant. Chem.: Quant. Biol. Symp. 12, 123 (1986).

[27] M. Randić, B. Jerman-Blažič, S. C. Grossman, and D. H. Rouvray, Proc. Nat. Conf. Mathematical Modelling, San Francisco (1985) p. 571.

[28] M. Randić, in *Molecular Basis of Cancer*. Part A, *Macromolecular Structure, Carcinogens, and Oncogens*, R. Rein Ed. (Alan R. Liss, Inc., New York, 1985) p. 309.

Entropic Elastomeric Force in Protein Structure/Function

DAN W. URRY

Laboratory of Molecular Biophysics, The University of Alabama at Birmingham, University Station/P.O. Box 311, Birmingham, Alabama 35294, U.S.A.

Abstract

Briefly noting earlier studies on the polypentapeptide of elastin, $(Val^1-Pro^2-Gly^3-Val^4-Gly^5)_n$, and on elastin, it is emphasized that entropic elastomeric force can be exhibited by nonrandom, anisotropic polypeptide systems and therefore that entropic elastomeric force does *not* necessarily result from isotropic random chain networks as required by the classical theory of rubber elasticity, nor does it result from solvent entropy effects as deduced from the slow loss of elastomeric force on thermal denaturation. Instead entropic protein elasticity can be the result of internal chain dynamics, specifically of librational processes that become damped on chain extension. This new mechanism of entropic protein elasticity allows for an understanding not only of elastin but also of the passive tension of striated muscle, of the voltage-dependent interconversion between open and closed conductance states in the sodium channel of squid nerve, and of protein elastic forces producing strain in a substrate bond during enzyme catalysis. Because entropic elastomeric force develops as a result of an inverse temperature transition, it becomes possible to shift the temperature of the transition to higher or lower temperatures by decreasing or increasing, respectively, the hydrophobicity of the elastomeric polypeptide chain. In warm-blooded animals this allows for biochemical modulation of the relaxation or development of entropic elastomeric force by an enzymatically modulated decrease or increase of the hydrophobicity, as for example, by phosphorylation or dephosphorylation of the elastomeric polypeptide chain. This understanding provides a mechanism for modulating protein function, whether for example enzymatic or channel, a mechanism for the remarkable reversible structural processes that attend parturition, and a mechanism for the connective tissue anomalies of wound repair and environmentally induced lung disease.

Introduction

Presently recognized as the primary elastomeric protein of warm-blooded animals, elastin is the second most prevalent protein in the extracellular matrix, only collagen is more common [1]. The nature of the elastomeric force was demonstrated by Hoeve and Flory in 1958 to be dominantly entropic in origin [2]. This is an important statement as it provides an understanding of the durability of elastin where single elastin fibers can last the lifetime of an individual, which when used in the human vascular system means undergoing more than one billion stress/strain cycles. That elastin is a dominantly entropic elastomer was reaffirmed by Hoeve and Flory in 1974 when they continued to insist that "A network of random chains within elastin fibers, like that in a typical rubber, is clearly indicated" [3]. This perspective has dominated thinking with respect to protein elasticity for nearly three decades and remains a staunchly held perspective [4–11]. Accordingly, the insistence that entropic elastomeric force

requires a random network of chains has precluded application to protein systems known to be *nonrandom* chain networks.

Studying the molecular structure and function of the polypentapeptide of elastin, this laboratory has demonstrated a new mechanism of entropic elasticity for this most striking primary structural feature of elastin [12], occurring within the longest sequence between cross-links, a sequence twice as long as any other possible elastomeric sequence between cross-links [12, 13], and has demonstrated its applicability to the elastin fiber as a whole [14, 15]. The mechanism derives from internal chain dynamics and is called the librational entropy mechanism of elasticity. In this report the new mechanism of entropic elasticity is considered relative to other protein systems where elastomeric force is implicated but where the proteins cannot be described as random chain networks.

In particular, the identification and possible origins of entropic elastomeric force are considered briefly. The applicability of internal chain dynamics, that is, librational processes, to protein elasticity as newly understood in elastin is extended to an understanding of the passive tension in muscle, of changing conductance states of channels, and of enzyme mechanisms. Furthermore, the relevance of structural transitions to and from the elastomeric state is considered in regard to elastogenesis, wound repair, fibrotic lung disease, and to processes attending parturition and their reversal, that is, cervical ripening and pubic ligament formation.

Possible Origins of Entropic Elastomeric Force in Proteins

Elasticity, of course, is the property whereby a material resists and recovers from deformation. The elastomeric force, f, can be considered to be comprised of two components: an internal energy component, f_e, and an entropy component, f_s, or

$$f = f_e + f_s \qquad (1)$$

Following Flory and colleagues, the relative magnitudes of the internal energy and entropy components can be determined by means of thermoelasticity studies [16]. In these studies the elastomer is extended to a fixed length and the elastomeric force is measured as a function of temperature. A plot of $\ln[f/T(°K)]$ versus temperature allows evaluation of the f_e/f ratio, or

$$\frac{f_e}{f} = -T \frac{\partial \ln(f/T)}{\partial T}_{P,L,eq} - \frac{\beta_{eq}T}{\alpha^3(V_i/V) - 1} \qquad (5)$$

where the experiment is carried out at constant pressure, P, with the elastomer at fixed length, L, and with the elastomeric matrix in equilibrium, eq, with the solvent. The second term in Eq. (2) is a correction term allowing the analysis to proceed at constant pressure rather than constant volume, and in equilibrium with solvent rather than at constant composition [17]. In this term $\beta_{eq} = (\partial \ln V/\partial T)_{P,L,eq}$ is the thermal expansion coefficient; α is the fractional increase in length; and V_i and V are the elastomer volumes before and after elongation. This correction term is of the order of 0.1 for elastin [18] as well as for the polypentapeptide of elastin [19]. Figure 1 shows thermoelasticity studies for elastin and for the polypentapeptide of elastin, where particularly for the latter, the near zero slope argues for a dominantly entropic elas-
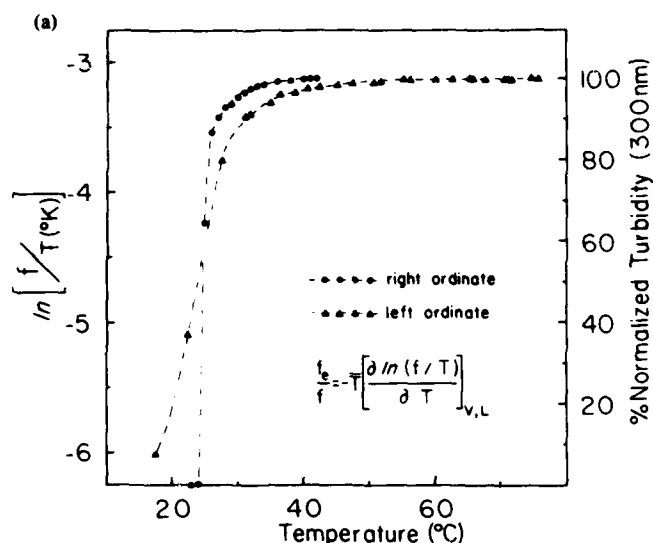
Figure 1. Thermoelasticity studies: Temperature dependence of elastomeric force at fixed extension. (A) Polypentapeptide of elastin cross-linked by 20 Mrads of γ-irradiation while in the coacervate state which is obtained by raising the temperature of solutions of polypentapeptide plus water from 20°C to 40°C to form a dense viscoelastic phase that is 62% water, 38% peptide by weight. The sample is extended to 60% at 40°C and then the force is measured as a function of temperature. In going from 20 to 40°C there is an abrupt development of elastomeric force, but above 40°C the plot of ln[force/T(K)] versus temperature exhibits a near zero slope. Since the slope is proportional to the f_e/f ratio and since this is near zero, it can be argued that the polypentapeptide of elastin exhibits dominantly entropic elastomeric force in the temperature range above 40°C. The development of elastomeric force in the 20–40°C range correlates with an inverse temperature observed by numerous physical methods and seen to be a process of self-assembly into fibers. Therefore the polypentapeptide of elastin is an ansiotropic, entropic elastomer. (B) Ligamentum nuchae elastin exhibits a similar development of elastomeric force on raising the temperature over a somewhat broader temperature range, but at higher temperatures the slope approaches zero and a dominantly entropic elastomer has been concluded. This conclusion is assisted by carrying out the study in 30% ethylene glycol in water which shifts the transition to lower temperature giving a wider temperature range where the slope is near zero. In both cases there is plotted on the right-hand side the temperature profile for aggregation, actually for fiber formation as observed by microscopy, for the constituent peptide on protein. (●—●) Right ordinate; (▲—▲) left ordinate. Reproduced with permission from Ref. 20.

tomeric force [20]. On changing the solvent to ethylene glycol-water, 3:7 by volume, the rapid rise in elastomeric force is shifted to lower temperature and the near zero slope becomes more apparent for elastin [unpublished data, 2, 3]. Furthermore, a near zero slope for elastin has been found in triethylene glycol [10]. Thus elastin and the polypentapeptide of elastin are considered to be dominantly entropic elastomers.
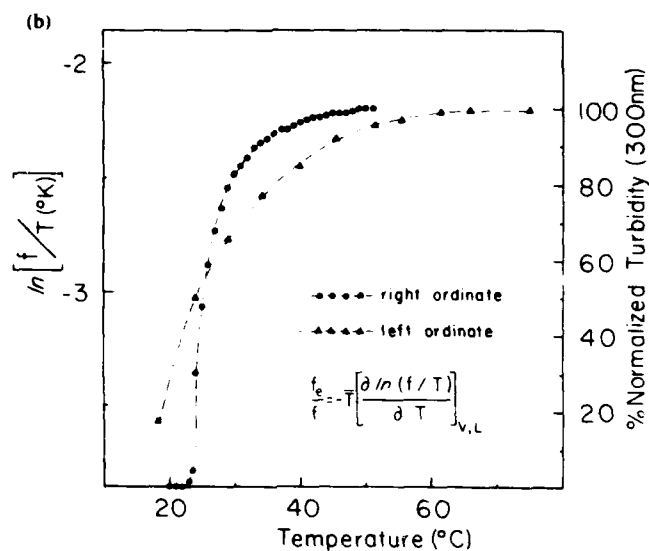
Figure 1.   (Continued.)

The Classical Theory of Rubber Elasticity for Random Chain Networks

The classical or statistical theory of rubber elasticity holds that entropic elastomeric force derives from random chain networks [21–23]. At rest the network is described as being comprised of a random distribution of chain end-to-end lengths. This is the highest entropy state. On stretching, the distribution of end-to-end lengths is displaced from that of highest entropy. This decrease in entropy provides the resistance to deformation and the driving force for recovery. A representative distribution of chain end-to-end lengths is given in Figure 2 where $W(r)$ is the probability distribution of the end-to-end lengths, r, in nm. In this theory the f_e/f ratio is given by $d \ln\langle r^2 \rangle o/dT$ where $\langle r^2 \rangle o$ is the mean square end-to-end chain length.

Solvent Entropy

When the elastomer is comprised of hydrophobic groups that become exposed to polar solvents such as water on extension, several workers [24–27] have suggested that the formation of clathrate-like water surrounding these exposed hydrophobic groups constitutes a decrease in entropy that would provide an entropic restoring force.

Internal Chain Dynamics: Librational Process

Another source of decrease in entropy on extension has been derived from studies on the polypentapeptide of elastin [14, 15, 28–31] but it is an entirely general mechanism. It asserts that chain segments within a bulk matrix have freedom to undergo rocking motions. Since the chain segments in the dense, cross-linked bulk matrix will be essentially immobilized at their ends, motion occurs by rotation about one bond
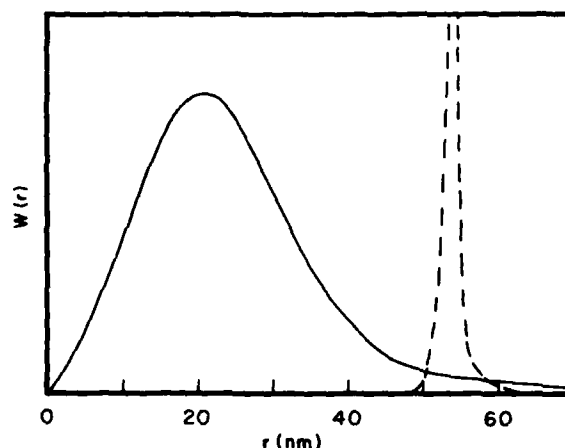
Figure 2. Probability distribution, $W(r)$, of chain end-to-end lengths r in nm. The solid line gives the distribution for a freely jointed chain with 10,000 segments of 0.25 nm each [22]. This is a random distribution of end-to-end lengths representing the highest entropy state. On stretching of a bulk cross-linked matrix of such a collection of chains, the distribution is displaced from that of a random chain network. The decrease in entropy provides a resistance to deformation and a restoring force. This is a description of the classical theory of rubber elasticity. The dashed curve represents a possible distribution of chain end-to-end lengths where the chains are nearly the same length. In this case an entropic restoring force can derive from the damping of internal chain dynamics on extension. This has been referred to as the librational entropy mechanism of elasticity which as represented, can occur with anisotropic fibrillar elastomers.

being paired with compensating rotations about one or more other bonds such that rocking motions or librational processes occur. On stretching these librational motions become damped. This has been termed the librational entropy mechanism of elasticity [29].

Elastomeric Processes in Protein Systems

The Polypentapeptide of Elastin

As shown in Figure 1(A), when the polypentapeptide of elastin is γ-irradiation cross-linked at a concentration of about 40% peptide, 60% water by weight, the resulting elastomer exhibits dominantly entropic elastomeric force above 40°C. On raising the temperature from 20° to 40°C, however, there is a dramatic development of elastomeric force. This development has been demonstrated by five independent physical methods — nuclear magnetic resonance structural and relaxation studies, dielectric relaxation studies, circular dichroism studies, microscopic characterization, and composition studies — to correlate with development of molecular order, that is, to correlate with an inverse temperature transition [14, 19, 32]. In the 20–40°C temperature range, development of molecular order correlates with development of elastomeric force. That the entropically elastomeric state above 40°C is an ordered state is further demonstrated by thermal denaturation followed by circular dichroism [19],

by extrusion of water [15, 19] and most directly by the slow loss of elastomeric force and of elastic modulus [15, 33], all demonstrated by heating at 80°C. As the elastomeric state is not a random chain network and since at 80°C destructuring of clathrate-like water would occur with time constant of the order of nanoseconds or less whereas the loss of elastic modulus at 80°C occurs with a half-life of days, the entropic elastomeric force must be due to internal chain dynamics.

The proposed elastomeric structure of the polypentapeptide of elastin is given in Figure 3 [31, 34, 35, 28] and the effect of stretching on the damping of the librational
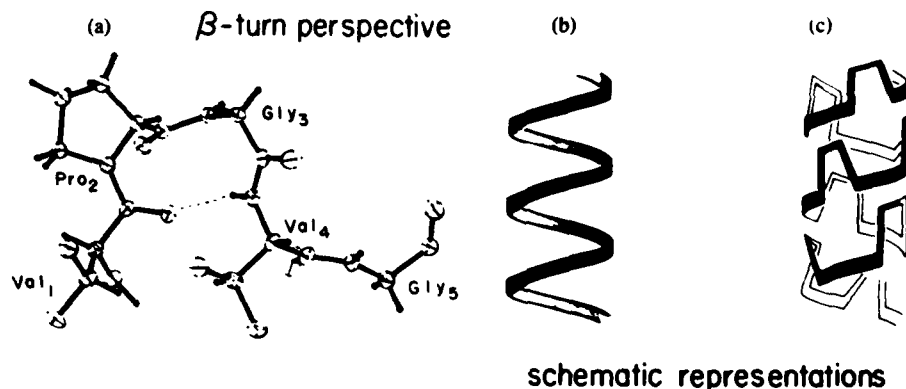

Figure 3. Proposed conformation of the elastomeric state of the polypentapeptide of elastin: (A) β-turn perspective showing the 10-atom hydrogen-bonded ring which utilizes the Val[1]C—O···HN Val[4] hydrogen bond. This conformation was first developed in solution using NMR methods and then demonstrated in the crystal for the cyclopentadecapeptide which was shown to be the cyclic conformational correlate of the polypentapeptide of elastin. Reproduced with permission from Ref. 34. (B) and (C) Schematic representations of the helical state (β spiral) of the polypentapeptide of elastin which is the elastomeric state. In (C) the β turns are included showing them to function as spacers with hydrophobic contacts between the turns of the spiral. Reproduced with permission from Ref. 31. (D) Detailed stereo pair of the spiral axis view showing space for water within the β spiral and showing suspended segments between the β turn. The suspended segment runs from the α-carbon of Val[4] to the α-carbon of Val[1] and is referred to as the Val[4]-Gly[5]-Val[1] suspended segment. It is within the segment where the large amplitude, low frequency librational motions are most pronounced (see Figs. 4 and 5). Reproduced with permission from Ref. 31. (E) Stereo pair of the side view of the β spiral of the polypentapeptide of elastin. This is one of a family of closely related β spirals. Seen here are gaps in the surface of the β spiral on each side of the suspended segments. The contacts between turns of the spiral utilize the Val and Pro hydrophobic side chains. The structure in (E) is displayed the same as in the schematic representation in (C). It is the optimization of intramolecular hydrophobic interactions that is responsible for β-spiral formation. Reproduced with permission from Ref. 35. (F) and (G) Supercoiling of β spirals to form twisted filaments of dimensions similar to those observed in transmission electron micrographs of negatively stained polypentapeptide, α-elastin and tropoelastin coacervates [14, 30, 40] and of negatively stained elastin. The structure is given in (F) in α-carbon to α-carbon virtual bond representation and in (G) in terms of spheres of different sizes centered at the α-carbon locations. Reproduced with permission from Ref. 28.
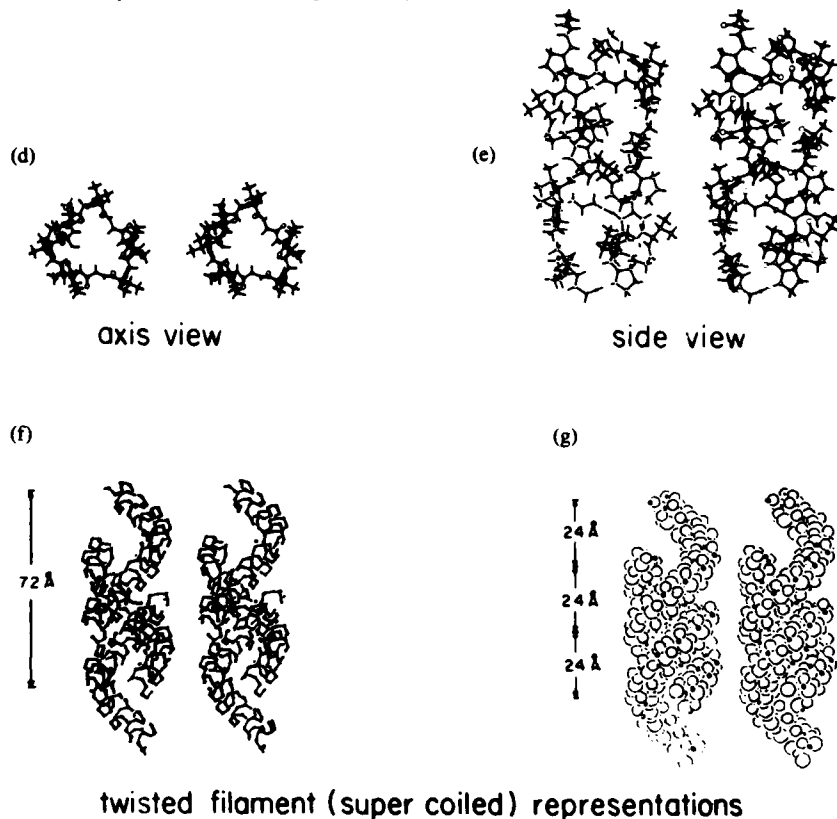
β-spiral of the polypentapeptide of elastin

(d)

(e)

axis view

side view

(f)

(g)

72 Å

24 Å

24 Å

24 Å

twisted filament (super coiled) representations

Figure 3. (*Continued.*)

motions is shown in Figure 4 [29]. The regularly repeating structure of the polypen-tapeptide provided the opportunity to demonstrate unequivocally that entropic elas-tomeric force occurs on formation of a regular nonrandom structure. One of the particularly interesting demonstrations is provided by dielectric relaxation studies [36]. At 20°C where there is minimal elastomeric force, the real part of the dielectric permittivity in the 1 GHz to 1 MHz frequency range exhibits a monotonically in-creasing curve. This is shown in Figure 5. As the temperature is raised and elas-tomeric force develops, there develops a localized Debye-type relaxation centered near 20 MHz. This has been assigned to a peptide librational mode [14, 36]. The in-tensity at 40°C, $\Delta\varepsilon \simeq 70$, and the localized nature of the relaxation require a regular nonrandom elastomeric state and the relaxation identifies a backbone (peptide) libra-tional mode that is directly contributing to the high entropy of the relaxed state. While the phenomenology enumerated above require setting aside the random chain network analysis and the elimination of solvent entropy as a consideration, this ex-periment allows direct observation of the responsible internal chain dynamics. This is the remarkable contribution of the polypentapeptide of elastin.
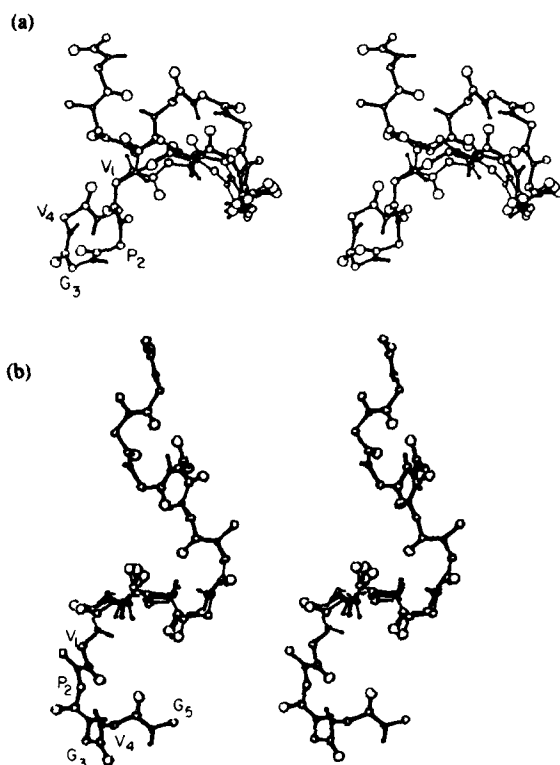
(a)

(b)

Figure 4.   Stereo pair view of a pentadecapeptide segment in the β-spiral conformation of Fig. 3(E) in which the central Val^1 α-carbon to Val^1 α-carbon pentamer has been allowed to assume conformations within a 2 kcal/mole residue cut-off energy. What is observed is a rocking motion of peptide moieties. In the relaxed state in (A), large librational motions are observed whereas in an extended state, in (B) at 130% extension, the librational amplitudes are markedly damped. This decrease in amplitude of the librations and possibly an associated increase in the frequency of the librational motions on extension is the decrease in entropy that resists elongation and that provides the restoring force. This is called the librational entropy mechanism of elasticity and this mechanism for developing entropic elastomeric force can occur in any polypeptide segment wherein the structure favors librational processes. Reproduced with permission from Ref. 29.

The Elastin Fiber

In the case of the elastin fiber, 3 of the 5 physical methods, utilized to demonstrate that increase in elastomeric force in the below 40°C temperature range correlates with increase in molecular order in the polypentapeptide, have been applied to elastin, to the precursor protein, or to a chemical fragmentation product of elastin. Those physical methods are microscopy [37–40], dielectric relaxation [41], and circular dichroism [42]. Furthermore, thermal denaturation has been directly observed on elastin, as on the polypentapeptide of elastin, by following the slow loss of elastomeric force in
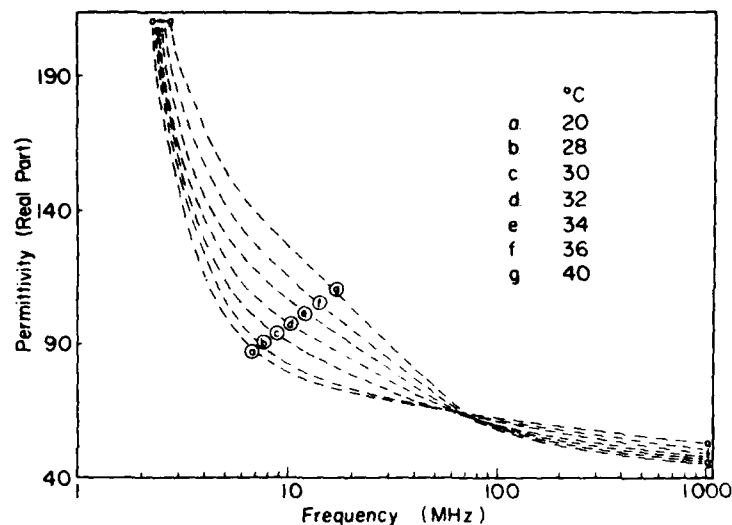
Figure 5. Dielectric permittivity (real part) of the polypentapeptide of elastin coacervate which is 38% peptide and 62% water by weight. On raising the temperature from 20 to 40°C there develops an intense, localized, Debye-type relaxation near 20 MHz. As the only dipolar entities are water and peptide moieties and because the intensity of the relaxation is so large and the frequency relatively low with a low temperature dependence, the relaxation is assigned to a peptide librational motion. Because the relaxation is at a localized frequency the polypentapeptides must be developing a regular structure as the temperature is raised from 20 to 40°C. The development of this relaxation correlates with the development of elastomeric force observed in Figure 1(A). The relaxation is taken to be due to the librational motions shown in Figure 4(A). Reproduced with permission from Ref. 36.

a thermoelasticity study and the slow loss of elastic modulus monitored by stress/strain curves at 37°C which resulted from heating at 80°C [15, 33]. Therefore, the entropic elastomeric force exhibited by this protein is not due to a random chain network nor is it due to the formation of clathrate-like water structures, rather it too must derive from internal chain dynamics. It may be noted that the slow thermal denaturation is in the practical sense irreversible in water. Here again the internal chain dynamics can, with the insight of the studies on the polypentapeptide of elastin and with awareness that the most prominent sequence between cross-links is where the polypentapeptide resides, be directly observed by dielectric relaxation studies on α-elastin (the chemical fragmentation product of elastin) in the 1 GHz to 1 MHz frequency range as shown in Figure 6 [41]. While the intensity of the relaxation is less, as expected with the polypentapeptide being a fractional component of α-elastin, a relaxation is again observed near 20 MHz.

Elastomeric Filaments of Muscle

Studies of Maruyama et al. [43, 44] and of Wang et al. [45, 46] have resulted in the

TEMPERATURE DEPENDENCE
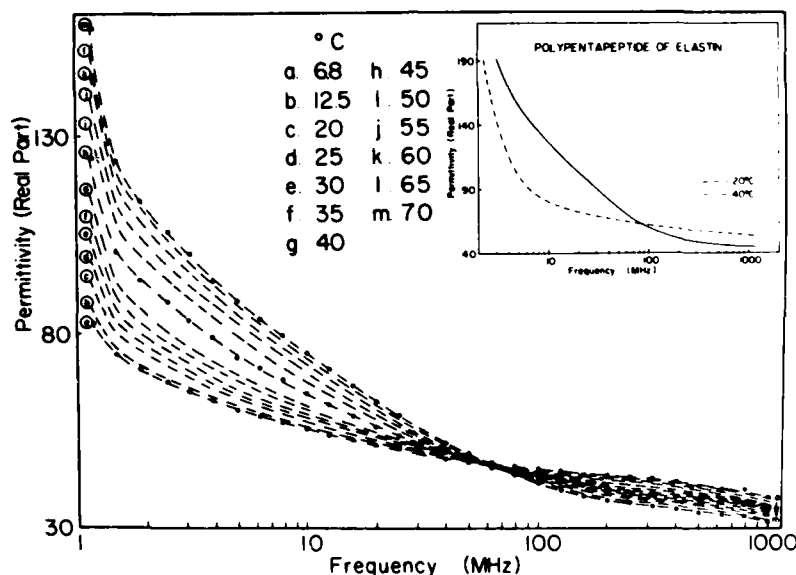## DIELECTRIC RELAXATION SPECTRUM OF
α-ELASTIN COACERVATE

Figure 6. Dielectric permittivity (real part) of the coacervate state of α-elastin which is a 70,000 molecular weight chemical fragmentation product of elastin. Below 15°C there is a monotonically increasing permittivity from several hundred MHz to 1 MHz. But as the temperature is raised there develops a relaxation near 20 MHz. A α-elastin contains the polypentapeptide of elastin which exhibits a similar relaxation (see inset and Fig. 5) this relaxation in α-elastin has been assigned to the same or similar peptide librational processes. The development of the relaxation with temperature in the 15° to 45°C temperature range correlates with the development of elastomeric force over the same temperature range as seen in Figure 1(B). Thus this along with considerable other data on elastin, α-elastin and tropoelastin including thermal denaturation of elastomeric force and elastic modulus of elastin at 80°C [33] allows the conclusion that elastin too is a nonrandom entropic elastomer. Reproduced with permission from Ref. 41.

isolation of a several million molecular weight elastic protein from muscle. Efforts to characterize this protein microscopically have demonstrated the protein to be filamentous [43]. This protein becomes a possible explanation for the passive tension of muscle and for the residual passive force exhibited when the sarcomere length has been extended beyond the point where the thick and thin filaments no longer overlap. Microscopic studies on pulled fibers have led to the identification of long narrow filaments either connecting the thick filaments to the Z lines or directly running from Z line to Z line [47,48]. Consistent with an effort to understand elastomeric force in terms of random networks, it has been suggested that the stretching itself causes the filaments to form from a gel state (see discussion following Ref. 47). Consistent with the random chain network theory of entropic elasticity, efforts are made to understand

elasticity in terms of an isotropic gel state rather than in terms of the anisotropic filaments observed microscopically both on the pulled fibers and for the isolated elastic protein of muscle. Having demonstrated internal chain dynamics to be the source of entropic elastomeric force, it now becomes possible to understand durable nonrandom, anisotropic elastomeric filaments, and thereby to accept the microscopic observations of the isolated elastic protein of muscle and of the pulled muscle fibers.

Interconversion of Sodium Channel Conductance States

To a physical chemist, one of the very challenging aspects of biology is understanding the molecular structure and mechanisms of ion-selective, voltage-dependent transmembrane channels. The conductance state — open, closed, refractory — depends on the transmembrane potential. It is of fundamental interest, for example, to understand what structural changes and processes result in changing the conductance state. This issue has been addressed in an interesting way by Rubinson [49] who modelled the sodium channel opening/closing equilibrium of squid nerve "as a charged region of a macromolecule moving under the influence of the applied field and confined elastically by interconnection with other masses." The result was the characterization of the mechanical properties of the polypeptide chain segment which controlled the gating process as rubber-like with an elastic modulus in the range of that of elastin. Taking the elastic modulus to be 5×10^6 dynes/cm^2 as for elastin, the ratio of the cross-sectional area to length ($\sim$400Å) of the connecting chain segment would not be unlike that of the polypentapeptide β-spiral in Figure 3. This is not to imply in any way that a β-spiral like that of the polypentapeptide of elastin actually exists in the sodium channel but rather to emphasize that internal chain dynamics and specifically librational processes rather than random chain networks would be required to understand this elastomeric process.

Enzyme Mechanisms

Several aspects of enzyme mechanisms may involve entropic elastomeric forces within the protein, for example, the structural rearrangements resulting from the binding of an allosteric effector [50], induced fit elements of substrate binding [51], and the catalytic process itself. In the former two processes it is apparent that binding to the surface of a viscoelastic protein could result in stretch-damping of librational motions within proximal regions of the active site. In addition the catalytic process itself has been considered in terms of elastic forces. Recalled, for example, is the elastomeric "rack" of Eyring et al. [52]. A recent elegant description of this element of enzyme catalysis has been presented by Gavish [53] in an exposition of "molecular dynamics and the transient strain model of enzyme catalysis." With emphasis on the viscoelastic properties of proteins [54], Gavish described a detailed model for stress and strain in the enzyme-substrate complex. The protein exerts an elastic force on the scissile bond of the substrate resulting in a strain that contributes to the potential energy required for bond cleavage. An effective means of increasing the rate of the catalytic process would seem to be to employ an entropic elastomeric force to induce strain in a substrate. Gavish states [53] "factors that dominate structural mobility in proteins should affect enzyme catalysis." On the basis of the new

understanding of entropic protein elasticity it might be said that *factors that modulate entropic elastomeric force should modulate enzyme catalysis*. For entropic elastomeric force as demonstrated by the polypentapeptide of elastin, it is not mobility *per se*, but rather it is mobility arising from a regularity of structure that gives rise to force capable of inducing significant strain. As shown by the nuclear magnetic resonance (NMR)-derived rotational correlation times [15], the mean mobility of the peptide moieties is greater at 25°C before the inverse temperature transition than at 37°C after the inverse temperature transition, yet the entropic elastomeric force is minimal at 25°C and dramatically increases until 37°C [15]. Thus it is not motion *per se* but the nature of the motion. In the dielectric relaxation studies at 25°C there is no localized relaxation in the 1 GHz to 1 MHz frequency range, but as the temperature is raised to 40°C there develops in concert with the development of elastomeric force an intense, Debye-type relaxation near 20 MHz, indicating motion within a regular structure [36]. Thus it is coherent motion (e.g., a librational mode) within a regular structure that gives rise to entropic elastomeric force. This provides for an anisotropic structure capable of producing a strain in an enzyme substrate by means of an entropic elastomeric force.

Modulation of Transitions in the Elastomeric State: Turning Entropic Elastomeric Force On and Off

In the preceding discussion of elastomeric processes in protein systems it was generally the elastomeric state itself that was considered, but the modulation of the transition to and from the elastomeric state can be an effective means of turning on and off an entropic elastomeric force. The modulation can be biochemical and it can be involved in such disparate processes as the modulation of enzyme catalysis, wound repair, the destruction of elastic tissue in environmentally induced lung disease, and relaxin-induced cervical ripening and pubic ligament formation attending parturition and their reversal.

Elastogenesis

Before addressing the more biomedical issues, it is necessary to consider the implications arising from the fact that, for elastin and the polypentapeptide of elastin, elastogenesis arises out of an inverse temperature transition and is therefore dependent on the hydrophobicity of the chains which are to constitute the elastomer. Generally, elastogenesis of elastin has been considered to be the physical process of fiber formation but as will be seen below it is simultaneously fiber formation *and* the development of elastomeric force. This is not possible within the constraints of the classical theory of rubber elasticity requiring, as it does, random chain networks, because the formation of an isotropic random chain network could not result in the formation of anisotropic fibers. *Once the random chain network perspective is set aside, it becomes apparent that modulation of elastomeric force in homoiothermic animals can be achieved by shifting the temperature range in which the inverse temperature transition occurs.*

Effect of Changing the Hydrophobicity. Using the polypentapeptide of elastin as the model elastomer, analogs can be prepared in which the hydrophobicity of

the repeating unit is changed. Three physical characterizations can be compared: (1) the temperature profile for aggregation, which is actually the temperature profile for fiber formation, (2) the temperature dependence of conformational change followed by circular dichroism, and (3) the temperature dependence of elastomeric force of the γ-irradiation cross-linked analog which has been stretched to a fixed length at 40°C. As shown in Figure 7, these transitions occur near 30°C for $(Val^1-Pro^2-Gly^3-Val^4-Gly^5)n$, the polypentapeptide of elastin. When the hydrophobicity of the repeating unit is increased as in $(Ile^1-Pro^2-Gly^3-Val^4-Gly^5)n$, the Ile^1 polypentapeptide, the temperature of the transition, as followed by all three means, shifts to lower temperature by some 20°C to near 10°C [55]. When the hydrophobicity of the repeating unit is decreased as in $(Val^1-Pro^2-Gly^3-Gly^4)n$ where the Val^4 residue has been deleted, the temperature of the transition shifts some 20°C higher to a temperature near 50°C [56]. These shifts are proportional to the hydrophobicity of the repeating unit as estimated by the Nozaki and Tanford [57] and the Bull and Breese [58] scales. This reaffirms the transition to be an inverse temperature transition, with a temperature inversely proportional to the hydrophobicity of the repeating unit. It is to be emphasized that the transition for the development of elastomeric force follows the hydrophobicity shifts; this further reaffirms development of elastomeric force to be the result of an inverse temperature transition leading to increased order for the elastomeric state [55,56].

Effect of the Transition on the Length of the Elastomer. The steepness of the curve for the development of elastomeric force of the Ile^1-polypentapeptide near 10°C [see Fig. 7(C)] is the result of matrix shortening and the fact that this sample had been stretched to 40% elongation at 40°C whereas the other samples had been stretched to 60% elongation at 40°C. As reflected in the temperature profiles of aggregation, the noncross-linked polypeptide is soluble in all proportions at a temperature below the onset of the inverse temperature transition [19]. This means that the cross-linked elastomers would dissolve on lowering the temperature below the transition if it were not for the cross-links. Instead of dissolving, the cross-linked polypeptides simply swell to the limit allowed by the cross-links and by the structural transition. This results in remarkable changes in the length of the cross-linked matrix as shown in Figure 8 where the length is measured as a function of temperature under zero load [59]. For 20 Mrad cross-linked polypentapeptide, the length of a strip of matrix increases 2.2 fold as the temperature is decreased from 40 to 20°C. Elastin shows analogous but less dramatic lengthening; a classical rubber such as latex, of course, shortens on lowering the temperature under zero load.

Biochemical Modulation of Hydrophobicity (i.e., of Transition Temperature). Rather than decreasing the temperature to relax the elastomeric force, it is possible to modify enzymatically the hydrophobicity of the elastomeric polypeptides and thereby to shift the temperature of the inverse temperature transition. This shift in temperature of the inverse temperature transition has been demonstrated with the enzyme prolyl hydroxylase. As shown in Figure 9, when the polypentapeptide is exposed to prolyl hydroxylase with the resulting hydroxylation of some of the Pro residues, this decrease in hydrophobicity causes the temperature profile for aggregation [60], equivalently for fiber formation and for elastomeric force development, to shift to higher temperature. This shift occurs with only about one Pro in 10 hydroxylated; this is only 1 hydroxylation in 50 residues. Thus enzymatic prolyl hydroxylation with a
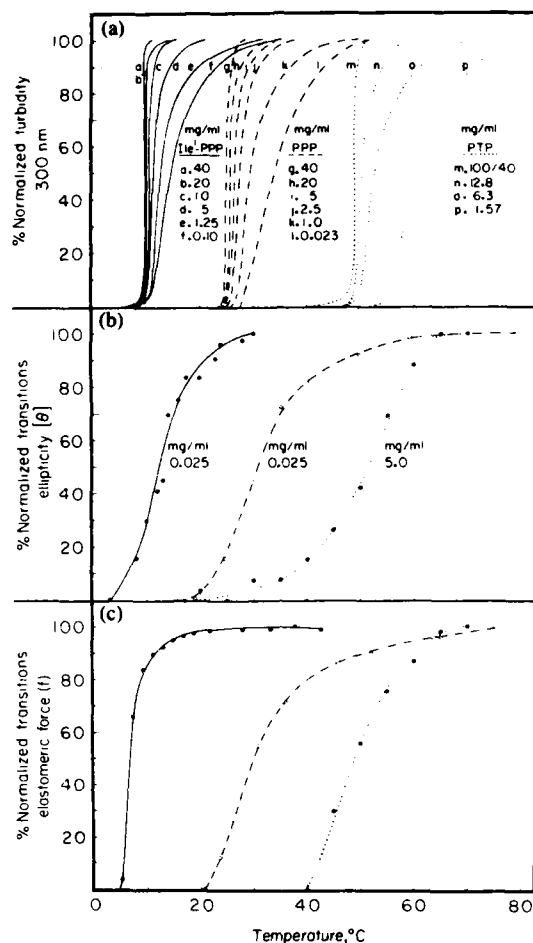
Figure 7. Comparison of a series of studies on a related series of elastomeric sequential polypeptides: Ile^1-PPP is $(Ile^1$-Pro^2-Gly^3-Val^4-$Gly^5)_n$; PPP is the polypentapeptide of elastin, $(Val^1$-Pro^2-Gly^3-Val^4-$Gly^5)_n$; and PTP is $(Val^1$-Pro^2-Gly^3-$Gly^4)_n$. These are all high polymers with molecular weights greater than 50,000. (A) Temperature profiles for aggregation which have been shown to be temperature profiles for fiber formation, that is, fiber formation occurs by an inverse temperature transition utilizing intermolecular hydrophobic interactions. Increasing the hydrophobicity of the repeating unit as in Ile^1-PPP causes the transition, i.e., fiber formation, to occur at lower temperature than for PPP; Ile is more hydrophobic than Val. Decreasing the hydrophobicity of the repeating unit as in PTP causes the aggregations, i.e., fiber formations, to occur at higher temperature. (B) The conformation of each of the sequential polypeptides is followed by circular dichroism of suspensions wherein the concentration was kept low enough so that the particulate distortions due to the small suspended aggregates were not significant. Observed in each case is an increase in intramolecular order as the temperature is raised through the transition. (C) Temperature dependence of elastomeric force, when the γ-irradiation cross-linked coacervates are set at a fixed extension, is followed. The development of elastomeric force is found to have shifted to the temperature range of the inverse temperature transition. This is a clear demonstration that elastomeric force develops as the result of an inverse temperature transition dependent on the hydrophobicity of the polypeptide. The elastomeric state is the more-ordered state and loss of elastomeric force can be achieved by decreasing order. The temperature range of the inverse temperature transition can be shifted by changing hydrophobicity of the polypeptide. If the temperature range of the transition could be reversibly shifted at body temperature then elastomeric force could be turned on and off. Adapted with permission from Refs. 55 and 56.
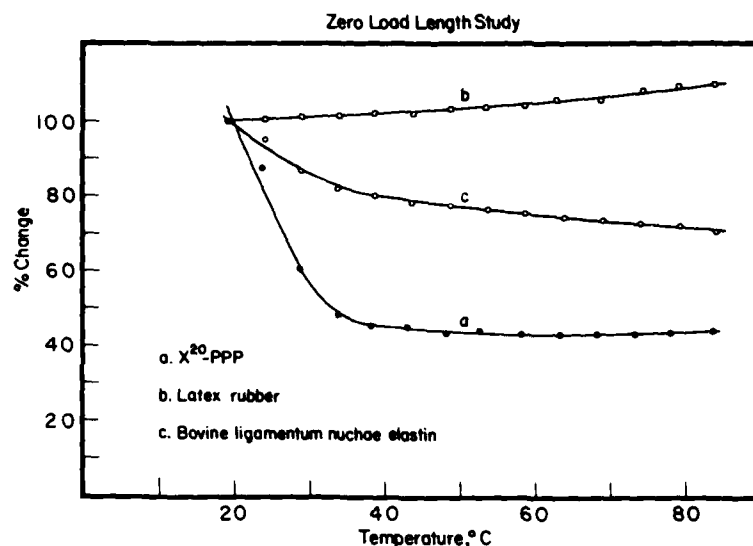
Zero Load Length Study

Figure 8. Effect of inverse temperature transition on the length of the elastomer. On raising the temperature from 20 to 40°C the 20 MRad γ-irradiation cross-linked polypentapeptide of elastin, X^{20}-PPP (●—●), undergoes a dramatic shortening to 45% of its 20°C length. This study is carried out at zero load (zero force). The structuring that occurs during the inverse temperature transition to form the β-spiral type of structure results in a shortening of the strip of X^{20}-PPP. A similar but less dramatic and more gradual shortening is observed for bovine ligamentum nuchae elastin (□—□). Typical of rubbers, latex (O—O) expands on raising the temperature. Thus elastomeric force is lost in part due to the structural transition. If by making the polypeptide less hydrophobic, the transition temperature range should shift to higher temperature and the elastomer would at body temperature lengthen and release or relax the force between two contact points. Reproduced with permission from Ref. 59.

sample of X^{20}-PPP held extended at 37°C should result in a decrease in elastomeric force when held at constant length and an elongation of the sample when maintained at a constant force.

While hydroxylation is an irreversible process, it becomes a trivial conceptual step to consider an elastomer with occasional serine or threonine residues that could be phosphorylated by a kinase causing the elastomer to extend (i.e., to relax) and that could be dephosphorylated by a phosphatase causing the elastomer to shorten and elastomeric force to again develop. It is suggested that such processes could be involved in the relaxin-induced cervical ripening and interpubic ligament formation and their reversal after parturition. Phosphorylation of enzymes and other proteins such as channels could be expected to have analogous effects on polypeptide segments capable of exerting entropic elastomeric force.

Biomedical Relevance

Wound Repair. In scar tissue there is a preponderance of collagen fibers with few or no elastin fibers [61]. In optimizing wound repair which involves sewing the breach together with collagen fibers, high levels of prolyl hydroxylase occur. Hydroxylation of proline residues in collagen is necessary for release of collagen from the cell; it is required to stabilize the collagen triple-stranded helix, and it protects collagen from nonspecific proteolysis (see references within Ref. 62). The same en-

Temperature Profile of Coacervation (1mg/ml)
Effect of Prolyl Hydroxylation in HCO-(Val₁-*Pro₂-Gly₃-Val₄-Gly₅)ₙ-Val-OMe

a. 0% Hyp
b. 1% Hyp
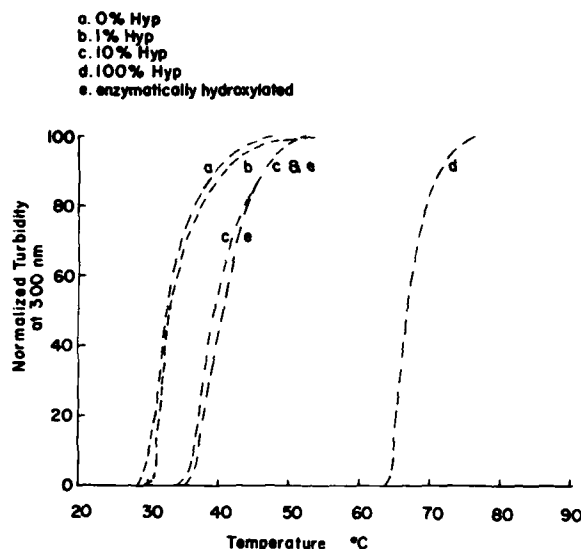c. 10% Hyp
d. 100% Hyp
e. enzymatically hydroxylated

Figure 9. Prolyl hydroxylation of the polypentapeptide of elastin by the enzyme prolyl hydroxylase decreases the hydrophobicity of the polypeptide and shifts the temperature range for the inverse temperature transition 10°C to higher temperatures. Using synthetic polypentapeptide in which 10% of the pentamers contained hydroxyproline instead of proline causes a similar shift. Of the order of one hydroxyl introduced in 50 residues causes a substantial shift in the transition, as much as 10°C. Considered in terms of Figure 7(C), this would shift the development of elastomeric force to a higher temperature. Considered in terms of Figure 8, this prolyl hydroxylation would at 37°C result in a lengthening of the elastomer. Thus an enzymatic modification is expected to cause a relaxation of elastomeric force at body temperature. If the enzymatic modification were phosphorylation and dephosphorylation then entropic elastomeric force could be turned off and on as desired for changing structural states in connective tissue and elastomeric components of muscle or for changing the functional state of an enzyme or channel, for example.

zyme hydroxylates proline residues in tropoelastin, the single precursor protein of elastin fibers. Based on the shift to higher temperatures of the temperature profile for fiber formation of the polypentapeptide of elastin (see Fig. 9) that results from prolyl hydroxylation, this decrease in hydrophobicity of tropoelastin would be expected to have a similar effect. The result would be less elastic fiber formation and the fiber formed would be in a more nearly relaxed state and unable to provide an appropriate entropic elastomeric restoring force. This has been demonstrated in cell cultures of aortic smooth muscle cells induced to high levels of hydroxylation by the addition of ascorbic acid required by prolyl hydroxylase [63].

Environmentally Induced Lung Disease. In environmentally induced lung disease, such as pulmonary emphysema, the elastin fibers are fragmented and dysfunctional. When the lung is challenged by toxic substances, it is proposed that the

ensuing repair response results in the elaboration of high levels of prolyl hydroxylase. The consequence of overhydroxylation of tropoelastin would limit elastin fiber formation; those fibers that did form would be able to exert a more limited elastomeric function because of the shift to higher temperature of the inverse temperature transition; and it is not unreasonable to expect that the poorly formed fibers would be more susceptible to proteolytic degradation [62]. In general, any process, such as inhalation of cigarette smoke, that resulted in oxidation of the elastomeric chains in elastin would cause a loss of elastic recoil.

Events Attending Parturition and Their Reversal. *Interpubic Ligament Formation.* There are remarkable deformations and restoring forces attending and following parturition. In mice and guinea pigs [64, 65] and in some women there is the development of an interpubic ligament in the days prior to delivery. In mice, for example, the pubic symphysis is normally less than 2 mm in width. In the days before delivery an interpubic ligament develops becoming 5–6 mm in length allowing for enlargement of the birth canal. By the third or fourth day after delivery the gap between the pubic bones is drawn back to 2 mm [65]. What connective tissue processes could allow this elongation, and then within the time period of a few days what restoring forces could result in the shortening? The above mentioned biochemical process of decreasing the hydrophobicity by phosphorylation could lead to lengthening by shifting of the temperature range of the inverse temperature transition for the development of elastomeric force to higher temperature. The result would be a biochemically controlled relaxation of elastomeric force. Subsequent removal of the phosphate moieties by phosphatases would result in a restoration of elastomeric force and a shortening of the elastomer. Interestingly, the shortening from about 5 mm to 2 mm is similar to the shortening of the cross-linked polypentapeptide (seen in Fig. 8) on going from the relaxed state at 20°C to the elastomeric state at 37°C. A 20°C increase in the temperature range of the inverse temperature transition by decreasing hydrophobicity due to phosphorylation could result in the lengthening and then dephosphorylation could return the transition temperature to its normal physiological range being completed as it is just at body temperature.

Cervical Ripening. The relaxing and softening of the cervix is referred to as cervical ripening. This occurs in the hours preceding delivery and is thought to be under the control of the hormone relaxin [66–68]. Here one could employ elastin fibers as considered for the interpubic ligament formation. However, if uterine smooth muscle fibers contained elastomeric filaments as observed in striated muscles, then phosphorylation and dephosphorylation of intracellular elastomeric filaments could readily be considered as a potential mechanism. This is a particularly attractive hypothesis as the mechanism of action of relaxin is considered to involve the activation of kinases and phosphatases in a time-dependent manner [69]. Once such a hypothesis is raised involving uterine smooth muscle cells it is natural to inquire whether such a process could be operative in vascular smooth muscle cells and be relevant to some forms of essential hypertension.

Requiem for the Random Chain Network Theory of Entropic Protein Elasticity

One of the purposes of the above limited enumeration of the possible roles of en-

tropic elastomeric force in protein structure and function is to demonstrate the reasoning that becomes possible once the shackles of the classical theory of rubber elasticity (requiring as it does random chain networks) are removed from consideration of entropic protein elasticity. Useful approaches of three decades ago should give way to more accurate descriptions, made possible by improvements in physical methods and their interpretation. These more correct descriptions can lead to new contributions, to new concepts of mechanism that can be tested by a wide range of experimental approaches. It is pernicious to hold that polypeptide backbone motions of the order of nanoseconds can only be achieved by random chain networks. It is contrary to progress in understanding protein structure and function to assume that the only examples of ordered polypeptide states are α-helix, β-sheet, and triple-standed helix and that all else is random. It is particularly curious to see protein structure deduced on the basis of a theoretical approach that has found it necessary to invoke phantom chains that occupy no space and that can pass through one another [70]. Once the random chain network theory of entropic protein elasticity is set aside, progress in understanding many fundamental processes utilizing entropic protein elasticity can occur more readily.

Acknowledgment

Bibliography

[1] L. B. Sandberg, W. R. Gray, and C. Franzblau, Eds., *Elastin and Elastic Tissue* (Plenum Press, New York, 1977).

[2] C. A. J. Hoeve and P. J. Flory, J. Am. Chem. Soc. **80**, 6523 (1958).

[3] C. A. J. Hoeve and P. J. Flory, Biopolymers **13**, 677 (1974).

[4] B. B. Aaron and J. M. Gosline, Nature **287**, 865 (1980).

[5] B. B. Aaron and J. M. Gosline, Biopolymers **20**, 1247 (1981).

[6] D. A. Torchia, L. S. Batchelder, W. W. Fleming, L. W. Jelinski, S. K. Sarkar, and C. E. Sullivan, *Mobility and Function in Proteins and Nucleic Acids* (Pitman, London, 1983).

[7] D. A. Torchia and K. A. Piez, J. Mol. Biol. **76**, 419 (1973).

[8] J. R. Lyerla and D. A. Torchia, Biochemistry **13**, 5175 (1975).

[9] J. M. Gosline and J. Rosenbloom, *Extracellular Matrix Biochemistry* (Elsevier Press, New York, 1984).

[10] A. L. Andrady and J. E. Mark, Biopolymers **19**, 849 (1980).

[11] L. DeBolt and J. E. Mark, Polym. Preprints **25**, 193 (1984).

[12] L. B. Sandberg, N. T. Soskel, and J. B. Leslie, N. Engl. J. Med. **304**, 566 (1981).

[13] J. Rosenbloom, private communication.

[14] Dan W. Urry, J. Protein Chem. **3**, 403 (1984).

[15] Dan W. Urry, *Elastin and Elastases*, L. Robert and W. Hornebeck, Eds. (CRC Press, Boca Raton, in press).

[16] P. J. Flory, A. Ciferri, and C. A. J. Hoeve, J. Polym. Sci. **45**, 235 (1960).

[17] K. L. Dorrington and N. H. McCrum, Biopolymers **16**, 1201 (1977).

[18] F. Mistrali, D. Volpin, B. G. Garibaldo, and A. Ciferri, J. Phys. Chem. **75**, 142 (1971).

[19] D. W. Urry, T. L. Trapane, and K. U. Prasad, Biopolymers **24**, 2345 (1985).

[20] D. W. Urry, C. M. Venkatachalam, S. A. Wood, and K. U. Prasad, *Structure and Motion: Membranes, Nucleic Acids and Proteins*, E. Clementi, G. Corongiu, M. H. Sarma, and R. Sarma, Eds.,

(Adenine Press, Guilderland, New York, 1985) p. 185.

[21] P. J. Flory, Principles of Polymer Chemistry (Cornell University Press, Ithaca, 1953).

[22] J. P. Queslel and J. E. Mark, Encyclopedia of Polymer Science and Engineering, 5, 365 (1986).

[23] L. Mandelkern, An Introduction to Macromolecules, 2nd ed. (Springer Verlag, Inc., New York, 1983).

[24] T. Weis-Fogh and S. O. Andersen, Nature 227, 718 (1970).

[25] J. M. Gosline, Biopolymers 17, 677 (1978).

[26] J. M. Gosline, The Mechanical Properties of Biological Materials, J. F. V. Vincent and J. D. Currey, Eds. (Cambridge Univ. Press, 1980), p. 331.

[27] W. R. Gray, L. B. Sandberg, and J. A. Foster, Nature 246, 461 (1973).

[28] D. W. Urry, C. M. Venkatachalam, M. M. Long, and K. U. Prasad, Conformation in Biology, R. Srinivasan and R. H. Sarma, Eds. (Adenine Press, Guilderland, New York, 1982), p. 11.

[29] D. W. Urry and C. M. Venkatachalam, Int. J. Quantum Chem.: Quantum Biol. Symp. No. 10, 81 (1983).

[30] D. W. Urry, Methods in Enzymology, L. W. Cunningham and D. W. Frederiksen, Eds. (Academic Press, Inc., New York, New York, 1982), pp. 82, 673.

[31] D. W. Urry, Ultrastruct. Pathol. 4, 227 (1983).

[32] D. W. Urry, Biomolecular Stereodynamics III, R. H. Sarma and M. H. Sarma, Eds. (Adenine Press, Guilderland, New York, 1985), p. 173.

[33] D. W. Urry, B. Haynes, and R. D. Harris, in preparation.

[34] W. J. Cook, H. M. Einspahr, T. L. Trapane, D. W. Urry, and C. E. Bugg, J. Am. Chem. Soc. 102, 5502 (1980).

[35] C. M. Venkatachalam and D. W. Urry, Macromolecules 14, 1225 (1981).

[36] R. Henze and D. W. Urry, J. Am. Chem. Soc. 107, 2991 (1985).

[37] L. Gotte, D. Volpin, R. W. Horne, and M. Mammi, Micron 7, 95 (1976).

[38] B. A. Cox, B. C. Starcher, and D. W. Urry, Biochim. Biophys. Acta 317, 209 (1973).

[39] B. A. Cox, B. C. Starcher, and D. W. Urry, J. Biol. Chem. 249, 997 (1974).

[40] D. Volpin, D. W. Urry, B. A. Cox, and L. Gotte, Biochim. Biophys. Acta 439, 253 (1976).

[41] D. W. Urry, R. Henze, P. Redington, M. M. Long, and K. U. Prasad, Biochem. Biophys. Res. Commun. 128, 1000 (1985).

[42] B. C. Starcher and D. W. Urry, Biochem. Biophys. Res. Commun. 53, 210 (1973).

[43] K. Maruyama, S. Kimura, H. Yoshidomi, H. Sawada, and M. Kikuchi, J. Biochem. 95, 1423 (1984).

[44] J. Maruyama, S. Kimura, K. Ohashi, and Y. Kuwano, J. Biochem. 89, 701 (1981).

[45] K. Wang, J. McClure, and A. Tu, Proc. Natl. Acad. Sci. USA 76, 3698 (1979).

[46] K. Wang, R. Ramirez-Mitchell, and D. Palter, Proc. Natl. Acad. Sci. USA 81, 3685 (1984).

[47] A. Magid, H. P. Ting-Beall, M. Carvell, T. Kontis, and C. Lucaveche, Contractile Mechanisms in Muscle, G. H. Pollack and H. Sugi, Eds. (Plenum Publishing, New York, New York, 1984), p. 307.

[48] K. Wang, Cell and Muscle Mobility, J. Shay, Eds. 315 (Plenum, New York, 1985), p. 6.

[49] K. Rubinson, Biophys. Chem. 25, 57 (1986).

[50] J. Monod, J. Wyman, and J. P. Changeux, J. Mol. Biol. 12, 88 (1965).

[51] D. E. Koshland, Jr., G. Nemethy, and D. Filmer, Biochemistry 5, 365 (1966).

[52] H. Eyring, R. Lumry, and J. D. Spikes, The Mechanism of Enzyme Action, W. D. McElroy and B. Glass, Eds. (Johns Hopkins Press, Baltimore, 1954), p. 123.

[53] B. Gavish, The Fluctuating Enzyme (John Wiley and Sons, Inc., New York, 1986).

[54] H. Frauenfelder, G. A. Petsko, and D. Tsernoglou, Nature 558, 280 (1979).

[55] D. W. Urry, M. M. Long, R. D. Harris, and K. U. Prasad, Biopolymers 25, 1939 (1986).

[56] D. W. Urry, R. D. Harris, M. M. Long, and K. U. Prasad, Int. J. Pept. and Protein Res. 28, 649 (1986).

[57] Y. Nozaki and C. Tanford, J. Biol. Chem. 246, 2211 (1971).

[58] H. B. Bull and K. Breese, Arch. Biochem. Biophys. 161, 665 (1974).

[59] D. W. Urry, B. Haynes, and R. D. Harris, Biochem. and Biophys. Res. Commun. 141, 749 (1986).

[60] D. W. Urry, H. Sugano, K. U. Prasad, M. M. Long, and R. S. Bhatnagar, Biochem. Biophys. Res. Commun. 90, 194 (1979).

[61] E. E. Peacock, Jr., Wound Repair (W. B. Saunders Company, Philadelphia, 1984), p. 3.

[62] D. W. Urry, R. S. Bhatnagar, H. Sugano, K. U. Prasad, and R. S. Rapaka, *Molecular Basis of Environmental Toxicity*, R. S. Bhatnagar, Ed. (Ann Arbor Science Publishers, Inc., Ann Arbor, MI, 1980), p. 515.

[63] L. M. Barone, B. Faris, S. D. Chipman, P. Toselli, B. W. Oakes, and C. Franzblau, Biochim. Biophys. Acta **840**, 245 (1985).

[64] F. L. Hisaw, Proceedings of the Society for Experimental Biological Medicine **23**, 661 (1926).

[65] K. Hall, J. Endocrinol. **5**, 174 (1947).

[66] A. H. MacLennan, R. C. Green, G. C. Bryant-Greenwood, F. C. Greenwood, and R. F. Seamark, Lancet **1**, 220 (1980).

[67] B. E. Kemp and H. D. Niall, *Vitamins and Hormones* (Academic Press, Inc., New York, New York, 1984), p. 79.

[68] G. Weiss, Ann. Rev. Physiol. **46**, 43 (1984).

[69] K. Nishikori, N. W. Weisbrodt, O. D. Sherwood, and B. M. Sanborn, Endocrinology **111**, 1743 (1982).

[70] R. T. Deam and S. F. Edwards, Phil. Trans. Roy. Soc. (London), **280**, 317 (1976).

Multistep Modeling of Protein Structure: Application Towards Refinement of Tyr-tRNA Synthetase

SUBHASHINI SRINIVASAN, MASAYUKI SHIBATA, MIHIR
ROYCHOUDHURY,* AND ROBERT REIN

*Unit of Theoretical Biology, Department of Biophysics, Roswell Park Memorial Institute, Buffalo,
New York 14263, U.S.A.*

Abstract

The scope of multistep modeling (MSM) is expanded by adding a least-squares minimization step in the procedure to fit backbone reconstruction consistent with a set of C-alpha coordinates. The analytical solution of Phi and Psi angles, that fits a C-alpha x-ray coordinate [1] is used for tyr-tRNA synthetase. Phi and Psi angles for the region where the above mentioned method fails, are obtained by minimizing the difference in C-alpha distances between the computed model and the crystal structure in a least-squares sense. We present a stepwise application of this part of MSM to the determination of the complete backbone geometry of the 321 N terminal residues of tyrosine tRNA synthetase to a root mean square deviation of 0.47 Å from the crystallographic C-alpha coordinates.

Introduction

Our recent papers [2, 3] concern the prediction of macromolecular structure from partial experimental data, for example, medium resolution crystallographic C-alpha coordinates. Purisima and Scheraga [1] approached this problem by using analytical method to find the set of Phi and Psi angles of the polypeptide backbone when threaded through the C-alpha trace. These authors pointed out that the method fails under certain conditions. Purisima and Scheraga [1] used a least-squares fitting of the C-alpha atoms of the polypeptide chain onto the C-alpha atoms of the crystal structure in the regions where the analytical solution failed [4] by optimizing a function

$$F = \sum_{i=1}^{N} a_i^2$$

where N is the number of guide points (C-alpha atoms in this case) and $a_i = r_i - r_{io}$. In the fixed coordinate system, the position of any atom with respect to the origin is represented by vectors r_i and r_{io}, respectively. Where the former corresponds to the i-th atoms in the computed conformation and the latter position in the x-ray structure. Our attempt to use the analytical method to refine the C-alpha trace of 321 N terminal residues of tyr-tRNA synthetase gave Phi and Psi values for 85% of the residues.

---

*On leave from the Department of Physics, University of Gorakhpur, Gorakhpur 273001 India.

This paper presents the use of least-squares minimization of the respective distances between C-alpha atoms in the model polypeptide and the crystallographic guide points to obtain the backbone geometry for the rest of the 15% residues consistent with the rest of the backbone. We further refine the model by computer graphics and energy optimization. We illustrate our results in the following sections along with a brief description of our method and experience gained during the assignment of Phi and Psi angles for all 321 residues.

Method and Results

The C-alpha coordinates of the 321 N terminal residues of tyr-tRNA synthetase are obtained from Brookhaven Data Bank. The method and results are subdivided into the following four sections:

Analytical Method

CONVRT and CONMAP, developed by Purisima and Scheraga [1], are two programs (Document No. NAPS-04122) obtained from ASIS National Auxiliary Publication Service which give all possible sets of Phi and Psi angles consistent with a given set of C-alpha coordinates. A score of one is assigned to all the Phi and Psi angles if the values are within the Ramachandran allowed region, or it is assigned a score of zero. The sum of this score for every solution is then used as a criteria to select one from many sets of Phi and Psi satisfying the same C-alphas. Various segments of the model which fit on the C-alpha trace of crystal structure are shown in Figure 1. There are 11 separate regions where the backbone dihedral angles Phi and Psi remain undetermined. We refer to these regions as "gaps" in later sections. Table I lists the first and last residue numbers of each segment.

TABLE I.   Segments for which Phi and Psi Angles are obtained using CONVRT and CONMAP.

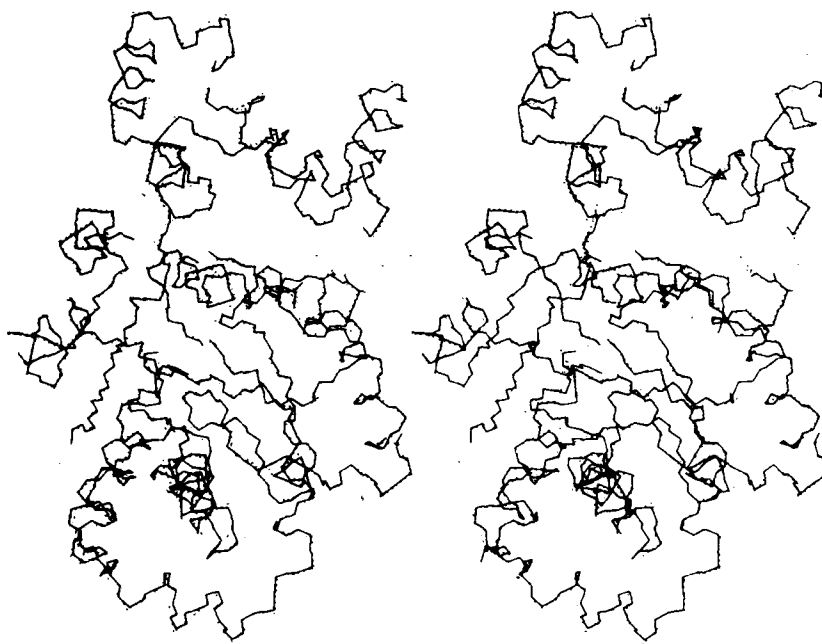| Segment no.	N-terminal residue no.	C-terminal residue no.
1	3	20
2	21	44
3	46	54
4	57	76
5	79	81
6	83	111
7	115	190
8	192	212
9	214	225
10	232	237
11	244	292
12	294	318

Figure 1. Dotted line = C-alpha trace; solid line = backbone segments obtained from
analytical method.

Optimization

The various segments of the structure for which Phi and Psi angles are obtained
from the previous step, lose their relative spatial orientation when arbitrary values are
assigned for the Phi and Psi angles of gap residues. These segments on both sides of
each gap are brought to their relative orientation by optimizing a minimal set of dis-
tances between the C-alpha atoms in the model to the corresponding distances in the
crystal structure. A minimal set of distances is a subset of all distances between the
C-alpha atoms which is necessary and sufficient to preserve a 3D configuration. This
choice is based on the concept that the position of a point is uniquely defined if its
distance from four non-coplanar points is known [5]. This is a special case of more
general mathematical conditions for the contraction of dimensionality and is stated in
more physical terms by Purisima and Scheraga [6]. Gaps are treated one or two at a
time depending on their separation along the chain. The variable parameters for the
optimization are the Phi and Psi angles of the residues in the gap along with the Phi
and Psi angles of three adjacent residues on both sides of the gap. The set of dis-
tances chosen are the distances ˜f all the C-alpha atoms in the moving segment from
four nonplanar C-alpha atoms of the fixed segment. Moving segment consists of all
the C-alpha atoms on the C terminal side of the gap and the C-alpha atoms of the
residues whose Phi and Psi angles are assigned variables. The fixed segment is the N
terminal side of the gap under consideration. Figure 2 shows the fixed and moving
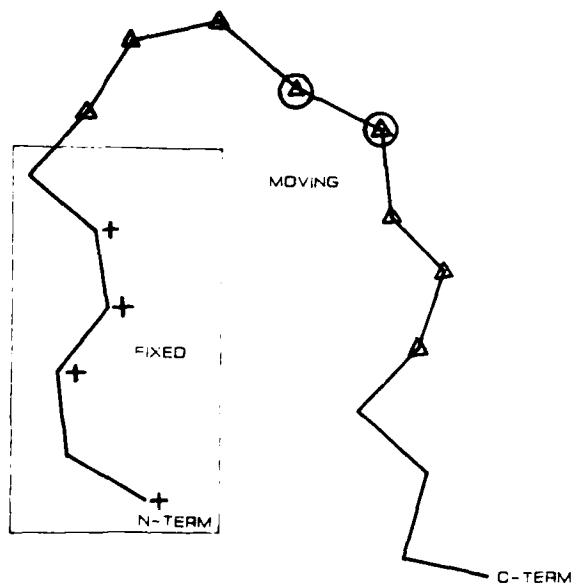
SRINIVASAN ET AL.

Figure 2.   Schematic showing fixed and moving regions. (Residues in the gap are circled)

segments in our definition. The function to be optimized is given by

$$F = \sum_{k=1}^{M} (\mathbf{D}_k - \mathbf{d}_k)^2$$

where M is the number of distances picked by the criteria described above. $\mathbf{D}$ is the vector of selected distances in the model and $\mathbf{d}$ is the vector of corresponding distances in the crystal structure. The optimization procedure adopted is the well known damped least-squares procedure of Levenberg [7]. The iterative equation for the i-th iteration is given by

$$\phi_{(i+1)} - \phi_i = (JJ - pI)^{-1} JF_k$$

where ϕ is the variable parameter, J is the rectangular matrix $(M \times N)$ of derivative $\delta F/\delta\phi$, I is the square identity matrix of order $(N \times N)$, p is a damping constant, and N is the number of variables.

The optimization of 7 of 11 gaps closed rapidly, showing steady convergence of the standard derivation. However, two other segments gave local mirror structure near the gap. This results because: (1) the starting geometry is far from minimum, and (2) the choice of the four fixed C-alphas on the N terminal segment may be very close to each other. However, this problem is readily overcome by using the negative of the Phi and Psi angles obtained from the ill optimized structure for residues where the chain direction is opposite. This set of changed Phi and Psi angles is then used as a starting model for closing the gaps. Figure 3 shows the result of this step.

Figure 3.   Dotted line = C-alpha trace; broken line = backbone after first stage of optimization; and solid line = backbone after second cycle of optimization.

Computer Graphics Modeling

The remaining two gaps, with 6 continuous residues with undefined Phi and Psi angles pose problems in convergence. These segments are modeled using our graphics program (MOSES) [8] and then used as an input for optimization step. Figure 4 shows the juxtaposed structure of this region before and after modeling.

The set of Phi and Psi angles for all the 321 residues obtained by the application of the above steps gave a root mean square deviation of 1.7 Å when least-squares juxtaposed on the crystal structure. This structure is further optimized for all distances closer than 10 Å in the crystal C-alpha coordinates and changing all 319 Phi and Psi angles of the built model. The root mean square improved to 0.47 Å. Figure 5 shows the complete backbone structure juxtaposed on C-alpha trace of the crystal structure.

Energy Minimization

A rough analysis of the set of Phi and Psi angles shows that 15–20% of the residues fall outside the Ramachandran allowed region. This is not unexpected as it is known that 10% of the Phi and Psi angles in the protein crystal structure are outside this allowed region. In crystal structure this is due to the considerable deviation of the peptide geometry from standard values. Here it could be because of the attempt to fit a standard geometry to 3 Å resolution structure. Forty-five cycles of energy optimiza-
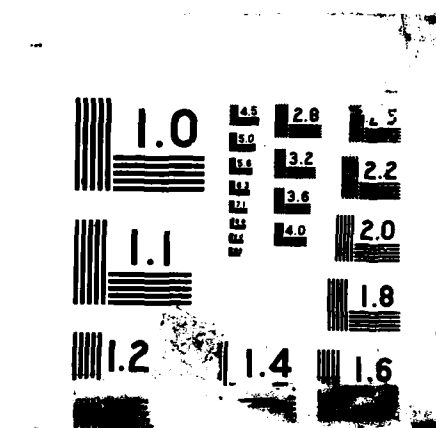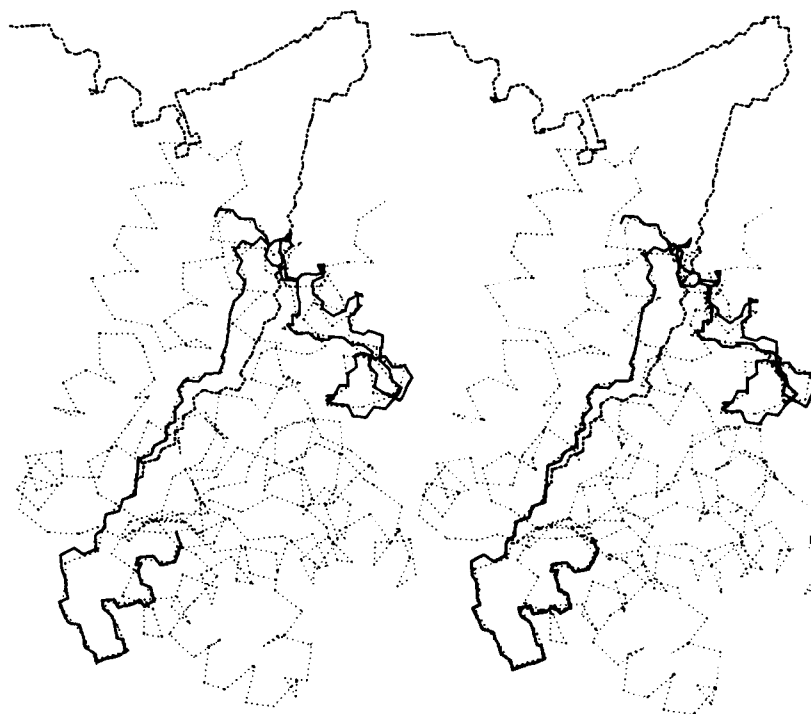
END
DATE
FILMED
10 88

Figure 4. Dotted line = C-alpha trace; broken line = model with two gaps of length of 6 residues whose values are assigned 180–180; and solid line = fitted model using computer graphics.

tion on the entire backbone with Ala in place of all other residues except Gly and Pro have been performed. The purpose of this step is to remove backbone short contacts which may exist in the final backbone and to improve the Phi values of the prolines. The position of the C-alpha atoms were kept fixed during optimization. In spite of the constraint on the C-alpha atoms, the deviation of the bond lengths and bond angles from the standard ECEPP geometry is <0.05 Å and <10.0°, respectively. The maximum out-of-plane deformation of the *trans* peptide bond is 18 degrees, while only 11 of them are >10 degrees. Figure 6 shows the final energy optimized structure fitted through the C-alpha trace. With the constraint on the C-alpha atoms and the limited cycles of energy optimization the value of the 6–12 components of the total energy is 78 Kcal for nearly two million atom pairs.

Discussion

Grafting of the respective side chain residues instead of alanine and refinement of the full structure using energy optimization is in progress. Although the structure with all the side chains is not yet known, we use the developed model to describe the
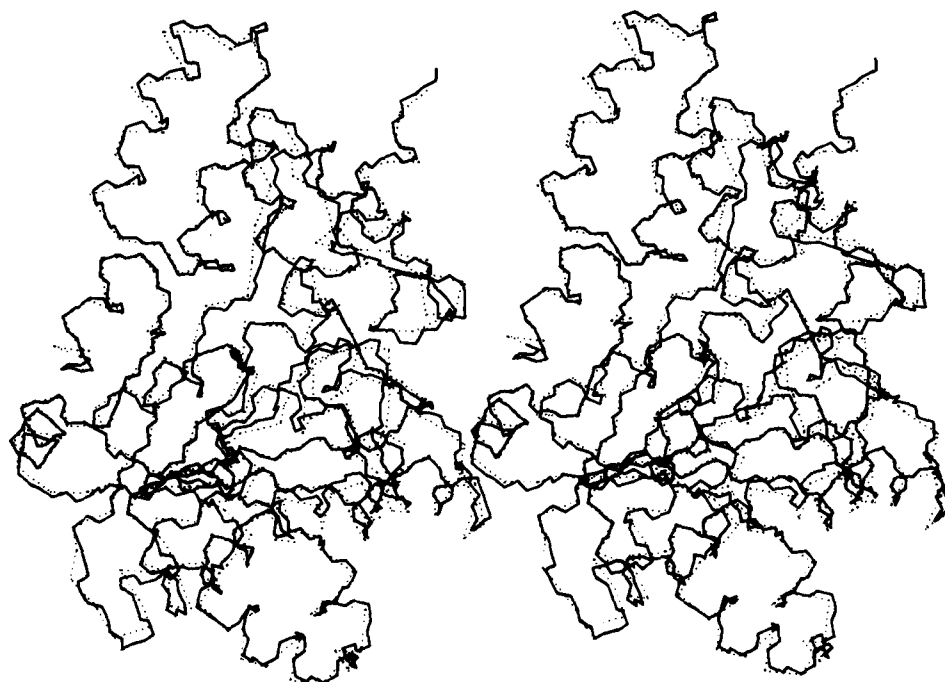
Figure 5.   Dotted line = C-alpha trace; and solid line = complete fitted backbone (root
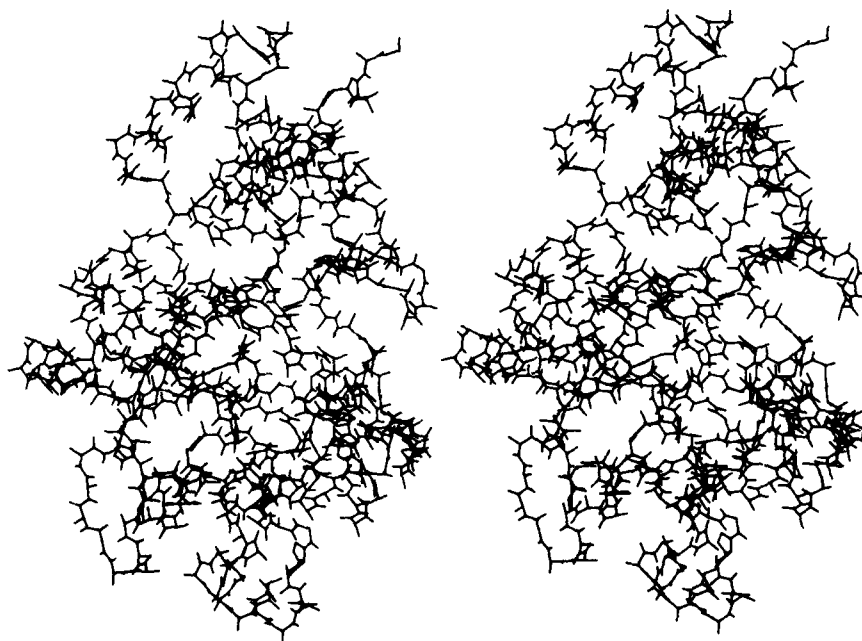mean square = 0.47 Å).

Figure 6.   Energy-refined backbone with alanine for all residues except glycines and
prolines.

HELICAL DOMAIN —— Ω-LOOP
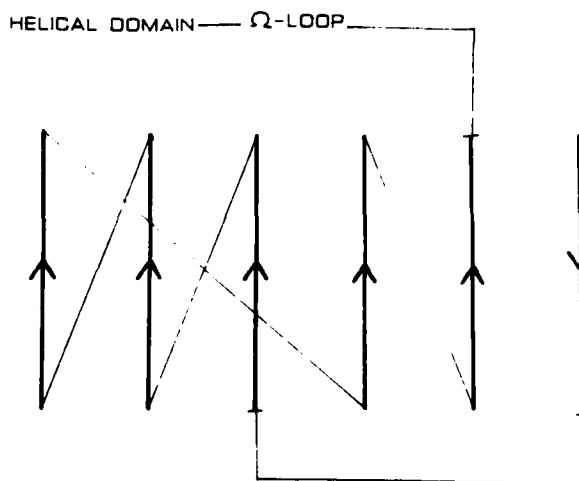
Figure 7.   Topology diagram for the alpha-beta region.

topology of this protein following the description of Richardson [9]. The topological diagram (Fig. 7) of the beta-extended and alpha-helical regions resembles that of dihydrofolate reductase and glutathione reductase [9]. Recognition of such homologous feature provide additional template to be identified in proteins and can be used in comparative modeling.

Acknowledgments

Bibliography

[1] E. O. Purisima and H. A. Scheraga, Biopolymers 23, 1207 (1984).
[2] S. Srinivasan, G. Raghunathan, M. Shibata, and R. Rein, Int. J. Quantum Chem. QBS 12, 217(1986).
[3] S. Srinivasan, M. Shibata, and R. Rein, Int. J. Quantum Chem. QBS 13, 167 (1986).
[4] P. K. Warme, N. Go., H. A. Scheraga, J. Comp. Phys. 9, 303 (1972).
[5] N. S. Goel, B. Rouyanian, and M. Sanati, J. Theor. Biol. 99, 705 (1982).
[6] E. O. Purisima and H. A. Scheraga, Proc. Natl. Acad. Sci. USA 83, 2782 (1986).
[7] K. Levenberg, Quart. Appl. Math. 2, 164 (1944).
[8] S. Srinivasan, D. McGroder, M. Shibata, and R. Rein, Biomolecular Stereodynamics III: Proceedings of the Fourth Conversation in Biomolecular Stereodynamics, State University of New York, Albany, R. H. Sarma and M. H. Sarma, Eds. (Adenine Press, New York, 1986b), pp. 299–304.
[9] J. S. Richardson, Adv. Protein Chem. 34, 167 (1981).

Received March 30, 1987

Stabilization of Alpha Helices by Ion Pairs

M. SUNDARALINGAM * AND Y. C. SEKHARUDU

Department of Biochemistry, College of Agriculture and Life Sciences, University of Wisconsin-Madison, 420 Henry Mall, Madison, Wisconsin 53706, U.S.A.

N. YATHINDRA[+] AND V. RAVICHANDRAN

Department of Crystallography and Biophysics, University of Madras, Madras 600 025, India

Abstract

A survey of 50 protein structures (47 globular and 3 fibrous) indicates that intrahelical ion pairs between oppositely charged residues (Glu$^-$, Asp$^-$/Lys$^+$, Arg$^+$) 3 or 4 residues apart along the helix may have a stabilizing effect on alpha helices exposed to solvent. It is found that the $i, i \pm 3/4$ types of ion pairs are the most predominant, and their observed frequencies are significantly greater than their expected frequencies. Such a preference is not seen for the like-charged pairs which served as a control. It was found that the normalized frequencies of these ion pairs increased with the helix length. An analysis of the distances between the charged groups in ion pairs suggests that only about 20% of the ion pairs are stabilized by hydrogen bonding (salt bridged), about 40% by electrostatic interactions, and the remaining may be stabilized by solvation: forming water bridges or plumes of water molecules around the charged groups. The fibrous proteins, which have a proportionately larger solvent exposed area than the globular proteins, have a higher density of intrahelical or secondary structural ion pairs. They are distinguished from the globular proteins which contain fewer ion pairs/charged residues because of their smaller solvent exposed area. The results indicate that the ion pairs may have a stabilizing effect on alpha helices exposed to solvent.

Introduction

One of the surprising features of the dumbbell-shaped molecular structure of skeletal muscle troponin C [1, 2] is the stability of the solvent-exposed alpha-helical "handle." It is found that the amino acid composition of the "handle" is rich in charged residues, Glu$^-$, Lys$^+$, Asp$^-$, and Arg$^+$ [3]. Furthermore, it is found that there are several ion pairs involving oppositely charged residues located 3 or 4 residues apart along the helix handle. The ion pairs may be visualized to be involved in (a) direct hydrogen bonding or salt bridges, (b) attractive electrostatic interactions, and (c) bridged by water molecules or surrounded by plumes of water molecules. It is surmised that the juxtaposition of oppositely charged residues or ion pairs can protect and screen the solvent-exposed helix from denaturation or "erosion" [3].

Results

Fibrous Proteins–Coiled Coils

Troponin C can be regarded as a "fibroglobular" structure constituting a fibrous component, (the alpha-helical "handle"), and two globular halves (N and C domains). The high density of charged residues and ion pairs in the helix "handle" prompted us to analyze the amino acid sequences of the fibrous proteins. The double-helical coiled coil structures are an important class of fibrous proteins occurring in nature. Some examples are alpha-tropomyosin [4] and myosin rod [5] from skeletal muscle, and lamin C [6], a major protein of the nuclear envelope. An intrinsic feature of these fibrous proteins is the presence of a large number of intrahelical ion pairs (Table I). In addition, the intrinsic sequence periodicity of the coiled coil results in the presence of oppositely charged residues at i, $i \pm 5/7$ residues apart which allow the formation of *interhelical* salt bridges between helices [4, 5]. Apparently the need for the numerous *intrahelical* ion pairs arises from the large exposed surface area. It appears that, as in the case of the troponin C "handle," the large concentration of intrahelical ion pairs stabilize the individual alpha helices of the coiled coil from "erosion" [7], while the periodic interhelical salt bridges and hydrophobic interactions stabilize the tertiary structure of the coiled coil [4, 5]. In alpha-tropomyosin there are 68 ion pairs of the type $i \pm 3/4$, constituting 24% of the total residues, while in lamin C and myosin rod these values are 14% and 22%, respectively. The ion pair density found in these fibrous proteins is significantly larger than that found in globular proteins [7].

Globular Proteins

The analysis was also extended to globular proteins [7]. A total of 47 globular proteins from the Brookhaven Protein Data Bank were sampled, displaying different tertiary structures and functions. The proteins were searched for the occurrence of oppositely charged pairs as well as like-charged pairs at i, $i \pm 1, 2, \ldots 8$ residues apart in the alpha-helical and nonhelical regions. The survey of the nonhelical regions and like pairs served as a control for this study. Possible correlations between ion pairs and helix length, order of the charged residues in ion pairs and its relation to helix

TABLE I.   Fibrous proteins.

Proteins	No. of residues	Oppositely charged pairs							
		$i \pm 1$	$i \pm 2$	$i \pm 3$	$i \pm 4$	$i \pm 5$	$i \pm 6$	$i \pm 7$	$i \pm 8$
Alpha tropomyosin	284	26	26	34	34	38	25	39	29
Lamin C	572	29	29	42	41	52	42	48	37
Myosin rod	1094	63	68	102	144	81	84	103	92
		Like charged pairs							
Alpha tropomyosin		31	28	28	28	20	29	41	28
Lamin C		41	48	33	30	33	24	31	31
Myosin rod		94	106	82	62	63	84	103	78

dipole were also investigated. The distances between the charged groups in the ion–ion pairs (opposite and like pairs) were also determined to evaluate the nature and strengths of the interactions between the pairs.

Of a total of 299 alpha helices in the 47 globular proteins considered, 163 helices contain either an ion pair, like pair, or both. The remaining 136 helices do not contain any charged pairs and are generally found buried. Of the 163 helices, 135 contain at least one ion pair while 90 contain at least one like pair. There are 73 helices with ion pairs, while there are only 28 helices with like pairs. The distribution of helix lengths is given in (Fig. 1). The frequency of occurrence of the $i, i \pm 1, 2, 3, 4$ ... 8 types of ion–ion pairs are shown in Table II. It is seen that the $i \pm 3$ and $i \pm 4$ types of ion pairs juxtaposing on the same side of the helix, are the most predominant, followed by $i \pm 1$ and then $i \pm 2$, in which the residues are on the opposite
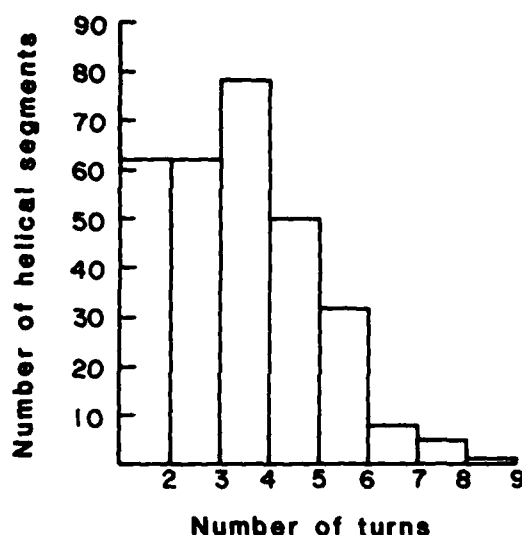
Figure 1.   The distribution of helical segments as a function of their lengths. The mean length of the helices is about 3-4 turns.

TABLE II.   Globular proteins.

Type of pair	Oppositely charged pairs	Like charged pairs
$i, i \pm 1$	95(100)	118(100)
$i, i \pm 2$	70(88)	75(85)
$i, i \pm 3$	120(74)	76(79)
$i, i \pm 4$	128(66)	73(77)
$i, i \pm 5$	56(54)	58(61)
$i, i \pm 6$	62(50)	53(52)
$i, i \pm 7$	59(40)	62(42)
$i, i \pm 8$	55(30)	46(34)

Expected numbers are given in parenthesis.

sides of the helix. There appears to be a slight preference for the $i \pm 4$ type over the $i \pm 3$. The frequencies of occurrence of $i \pm 1, 2, 3 \ldots 8$ types of pairs on beta strands and turns (Fig. 2), and on the proteins as a whole (Fig. 3), were also determined. It is found that the beta strands and turns showed no preference for the $i \pm 3/4$ type of ion pairs. This is to be expected because in beta strands the extended peptide chain places the side chains farther apart. The normalized frequencies for ion pairs on alpha helices were also determined (Fig. 4). It is found that the longer helices contain more than their proportionate number of ion pairs. However, like pairs do not show this property. An important observation is that the observed frequencies of ion pairs are significantly greater than the expected frequencies (Table II).

Some examples of exposed helices (not necessarily amphiphilic) with ion pairs are presented in the form of helical wheels in Figure 5. *Trp* repressor [8] presents a strik-

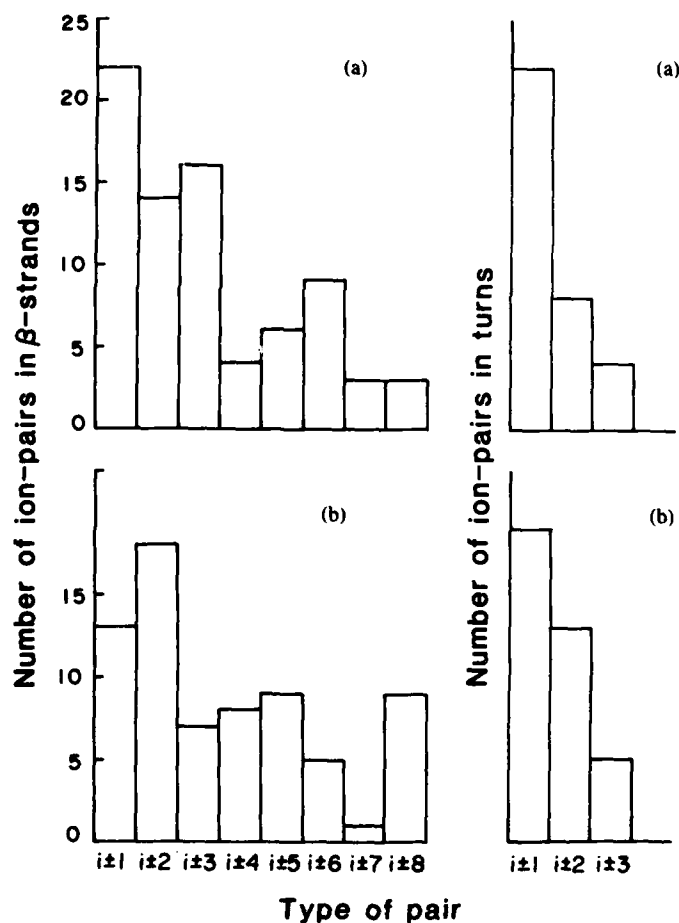Figure 2. Frequency of occurrence of (a) like pairs and (b) ion pairs on beta strands and turns.

ing example of an alphahelix stabilized by ion pairs. The 29 residues at the amino terminal are exposed to solvent. The first 11 residues are uncharged and nonhelical while the remaining stretch (12–29) is helical and endowed with four contiguous $i \pm 3$ type of ion pairs.

Interaction of Ion Pairs With the Helix Macrodipole. It is found that the frequency of occurrence of the $i \pm 1$ type of like-charged pairs is higher than the corresponding ion pairs by 19.5% (Table II). A total of 37% of these like-charged pairs are situated at the ends of the helix. The presence of adjacent negatively charged residues and positively charged residues at the amino and carboxyl ends of the helix, respectively, is favored by the helix dipole. This preference is consistent with an earlier observation that there is a preference for a single appropriately charged residue at the helix termini [9, 10]. When we consider the three terminal residues of the helices, it is found that the termini of 42% of the alpha helices contain the appropriate charged residues, 39% contain neutral residues, and 19% contain the incorrectly charged residues violating the alpha helix dipole. It appears then that the helix dipole is not a major factor contributing to helix stability.

Of the four possible combinations of ion pairs involving the residues Glu^-, Asp^-, Lys^+, Arg^+; the Lys^+-Glu^-/Glu^--Lys^+ combination is the most favored. Similarly, Arg^+ prefers association with Glu^- than with Asp^-. Glu^- shows a stronger preference than Asp^- to occur in association with a positively charged residue. It is also found that in general the negatively charged residue precedes the positively charged residue.
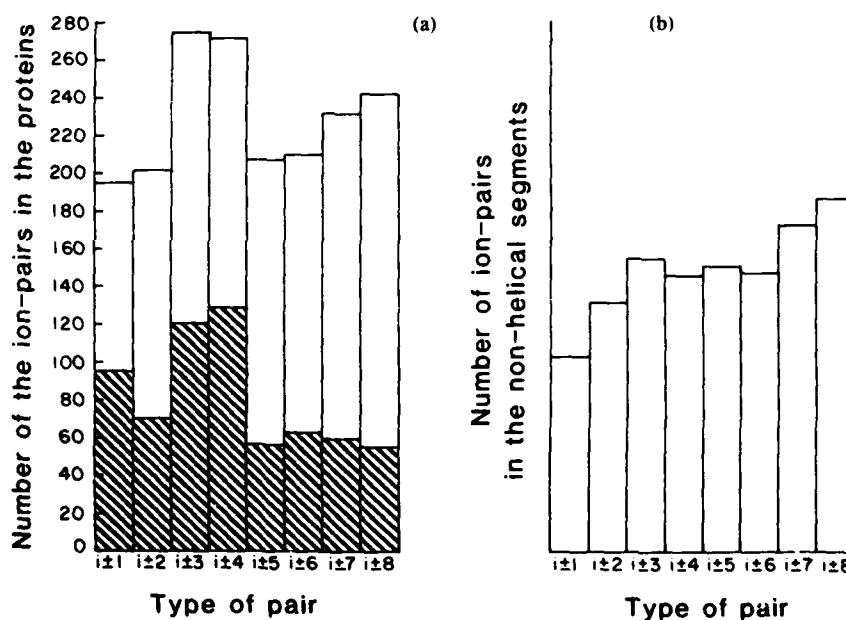
Figure 3. Frequency of occurrence of ion pairs (a) in the proteins as a whole, unshaded part, in the alpha helices, shaded part and (b) in the nonhelical regions which was obtained by the difference between ion pairs in helices and in proteins.
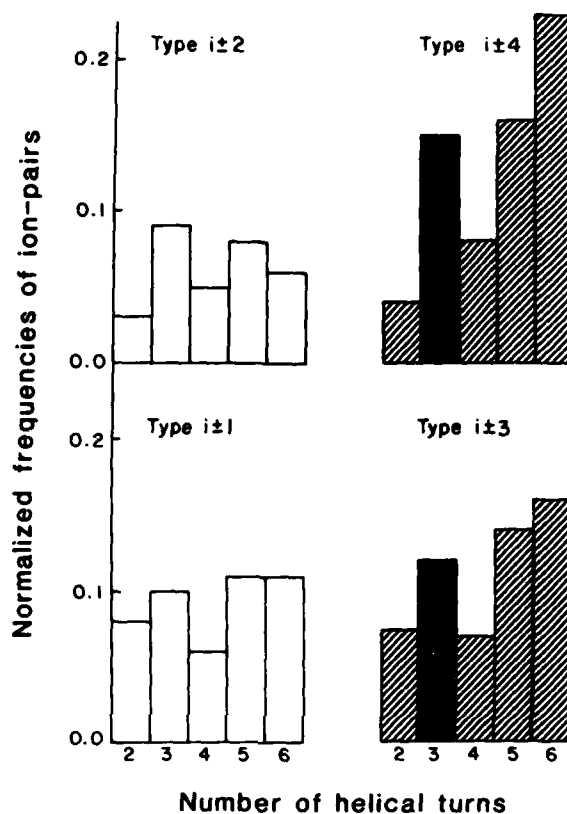
Figure 4. The normalized frequencies of ion pairs in alpha helices. The normalized frequencies represent the density of ion pairs per turn. Notice that the normalized frequencies of the $i \pm 3/4$ ion pairs increase with the helix length. Helices with three turns particularly show a higher density of ion pairs.

Influence of Ion Pairs/Charged Residues on Helix Stability. It is found that about 17% of the ion pair distances are less than 4 Å (hydrogen bonded), 43% at distances ranging from 4 to 7 Å (electrostatic interactions/water bridges), and the remaining 40% are beyond 7 Å (hydrogen bonded to solvent). The like-charged pairs are mainly found at distances greater than 4 Å.

It is known that the solvation energies of charged residues contribute to the stability of a molecular structure [11]. This may be relevant to the observation that shorter helices can "survive" with only charged residues or fewer ion pairs [7] while the longer helices are dependent on the larger concentration of ion pairs. The ion pairs in the exposed alpha helices seem to have increased solvation potentials compared to the randomly distributed charged residues in an alpha helix.
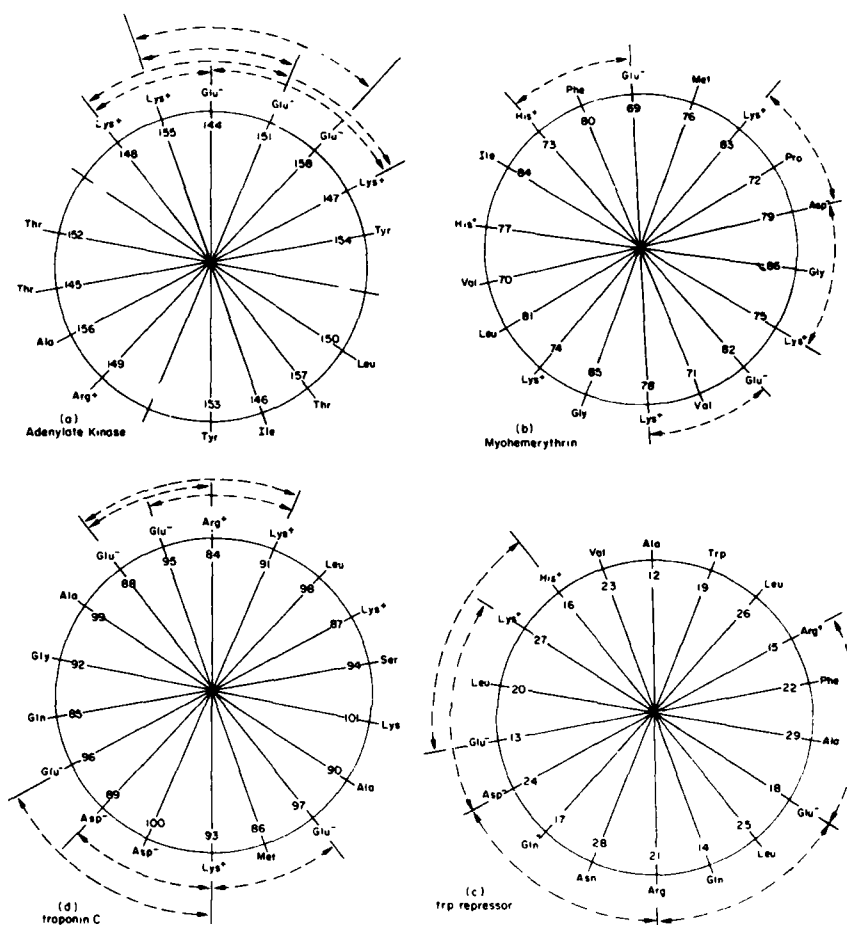
Figure 5. Alpha helical wheels looking from carboxyl to amino end: (a) adenylate kinase (residues 144–158), 6 ion pairs, (b) myohemerythrin (residues 69–87), 4 ion pairs, (c) *trp* repressor (residues 69–87), 5 ion pairs, and (d) troponin C (residues 84–101), 6 ion pairs.

Conclusion

We have shown that in exposed alpha helices ion pairs* of the type i, $i \pm 3$, and i, $i \pm 4$ occur with greater frequency than expected [7]. It appears that the ion pairs contribute to the stability of the exposed alpha helices by protecting the helix backbone hydrogen bonding from "erosion." The ion pairs seem to stabilize the helices by

---

*The ion pairs may be referred to as secondary ion pairs as they stabilize the alpha helices, while the ion pairs that stabilize tertiary structure may be referred to as tertiary ion pairs.

their solvation potentials, hydrogen bonding, and electrostatic interactions between the oppositely charged residues. Although the like pairs are generally fewer in number, they may also have a stabilizing effect on exposed alpha helices because of their solvation potentials. The greater propensity for ion pairs results from the possibility of direct hydrogen bonding and electrostatic interactions between the oppositely charged residues.

It appears that both ion pairs [7] and hydrophobic triplets [12] have a stabilizing role on protein secondary structure; a buried alpha helix contains more hydrophobic triplets while an exposed alpha helix contains more ion pairs. The amphiphilic helices will be stabilized by hydrophobic triplets on the buried side and by ion pairs on the solvent side. Thus, several factors seem to contribute to the stability of alpha helices: ion pairs, hydrophobic interactions, and helix-dipole interactions.

Our work has established a link between the fibrous and globular families of proteins in that, the long alpha helices of fibrous proteins require numerous ion pairs or charged residues, while the shorter alpha helices of the globular proteins require fewer ion pairs. In other words, the greater the exposed area of a protein/helix (elongated molecule) the greater is the ratio of charged to apolar residues, while the lower the exposed area (globular molecule) the lower is the ratio of charged to apolar residues. The "fibroglobular" (troponin C) family of proteins belong to a third class which share some features in common with the globular and fibrous proteins.

Acknowledgments

Bibliography

[1] M. Sundaralingam, R. Bergstrom, G. Strasburg, S. T. Rao, P. Roychowdary, M. Greaser, and B. C. Wang, Science 227, 945–948 (1985).

[2] O. Herzberg, M. N. G. James, Nature (London) 313, 653–659 (1985).

[3] M. Sundaralingam, W. Drendel, and M. Greaser, Proc. Natl. Acad. Sci., USA. 82, 7944–7947 (1985).

[4] A. D. McLachlan M. Stewart. L. B. Smillie, J. Mol. Biol. 98, 281–291 (1975).

[5] A. D. McLachlan, and J. Karn, J. Mol. Biol. 164, 605–626 (1983).

[6] F. D. McKeon, M. W. Kirschner, and D. Caput, Nature 319, 463–468 (1968).

[7] M. Sundaralingam, Y. C. Sekharudu, N. Yathindra, and V. Ravichandran, Proteins: Structure, Function, and Genetics 2, 64–71 (1987).

[8] P. B. Sigler, Personal communication.

[9] D. E. Blagdon and M. Goodman, Biopolymers 14, 241–245 (1975).

[10] P. Y. Chou and G. D. Fassman, Biochemistry 13, 211–222 (1974).

[11] R. Wolfenden, L. Anderson, P. M. Cullis, and C. C. B. Southgate, Biochemistry, 20, 849–855 (1981).

[12] J. Palau and P. Puigdomenech, J. Mol. Biol. 88, 457–469 (1974).

Received May 1, 1987

Thiophilic Adsorption: A New Kind of Molecular Interaction Revealed by Chromatography

JERKER PORATH*

Fogarty Scholar-in-Residence, Fogarty International Center, National Institutes of Health, Bethesda, Maryland, U.S.A. and Institute of Biochemistry, Biomedical Center, University of Uppsala, Box 576, S-75123 Uppsala, Sweden

T. WILLIAM HUTCHENS

Reproductive Research Laboratory, St. Luke's Episcopal Hospital, and Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, Texas 77030, U.S.A.

Abstract

We have recently discovered a new entirely unexpected, and highly selective protein–ligand interaction. This new kind of molecular interaction was recognized by chromatography of proteins on divinylsulfonated agarose gels which had been deactivated using mercaptoethanol. The essential structure of the interacting immobilized ligand is quite simple and nonionic. It can be generally represented by:

$$agarose\text{-}O\text{-}CH_2CH_2\text{-}SO_2\text{-}CH_2\text{-}CH_2\text{-}X\text{-}$$
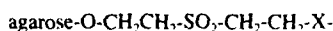
where X was first a thioether but can also be N, O(S > N ≫ O) or possibly any atom with at least one lone electron pair. We have provisionally termed peptides and proteins interacting with this ligand "thiophilic," in recognition of their affinity for the definitive thioether sulfone constituents. The thiophilic adsorption process is promoted by water-structuring or "antichaotropic" salts such a. sulfates or phosphates and would appear to be entropically driven. The thermodynamics of such a process are discussed relative to protein recognition of the immobilized thioether-sulfone ligand. We do not yet know the precise mechanism for the interaction but we believe that salt allows the protein into close contact with the immobilized thioether-sulfone group where short-range forces are likely to be important. Evidence suggests that aromatic side chains on the protein-binding site may be involved and we therefore expect that some kind of electron-donor-acceptor or proton-acceptor mechanism is likely involved. Two important applications of thiophilic adsorption are the selective immobilization of functional antibodies as well as purification of immunoglobulins from serem, ascites fluid, and hybridoma cell culture media. Monoclonal antibodies can be purified in one step under extremely mild (structure-stabilizing) conditions. We therefore consider the further characterization of thiophilic adsorption of major significance in the fields of immunology and biotechnology and hope that this presentation will inspire attempts to explain the interaction in terms of quantum chemistry.

Introduction

In living matter all information transfer, energy-converting systems, and structural components are dependent on molecular interactions which in their simplest form require interplay, involving affinity relationships, between two molecules or even two separate atomic groupings on a single molecule. These interactions can occur entirely

---

* Address all correspondence to Jerker Porath at the University of Uppsala, Sweden.

in free solution but more often, at least in biological systems, molecular interactions involve the constant association and dissociation of solutes with membrane- and/or particle-immobilized ligands or acceptor sites. Simple approximations of these types of solution/solid phase interfacial recognition events may be evaluated and modelled *in vitro* by chromatography using simple, chemically defined ligands. Conversely, as outlined here, when we encounter previously unrecognized interaction phenomena during chromatography of biomolecules in aqueous solutions there is some chance that we will find them also, in one form or another, in biological systems. Our purpose here is to present one specific and intriguing example of a larger concept utilizing chromatography for the development of model interfacial ligand:ligate systems to better understand the individual driving forces responsible for specific solid-phase biomolecular interactions.

During investigations designed to prepare a more desirable, activated solid phase to immobilize ligands useful for protein fractionation it seems as if we have discovered a kind of molecular interaction not previously described or utilized [1]. The discovery was serendipitous. Spherical beads of agarose (6%) were simultaneously crosslinked and activated with divinylsulfone (DVS). While DVS crosslinking confers rigidity upon the gel skeleton and improves the chromatographic performance of the stationary phase at elevated pressures (i.e., high-performance liquid chromatography, "HPLC"), it also activates the gel for further derivatization. However, when the remaining active vinyl groups were eliminated using the simple nucleophile β-mercaptoethanol to prevent nucleophilic components in the mixture to be fractionated from themselves becoming permanently (covalently) attached to the gel, the "deactivated" gel product so obtained revealed a selective adsorption property unlike any previously described. The new gel derivative has tentatively been named the T gel in reference to the Thioether structure of the immobilized ligand. Similarly, the specific adsorption behavior of certain peptides and proteins toward the T gel has been designated "thiophilic" to recognize the affinity of interacting molecules for the sulfur groups in the thioether-sulfone ligand [2,3]. The structure of the T gel is schematically provided in Figure 1. We believe it is important to emphasize the simple and nonionic nature of the interacting ligand. We will describe some

Immobilized T-gel Ligand
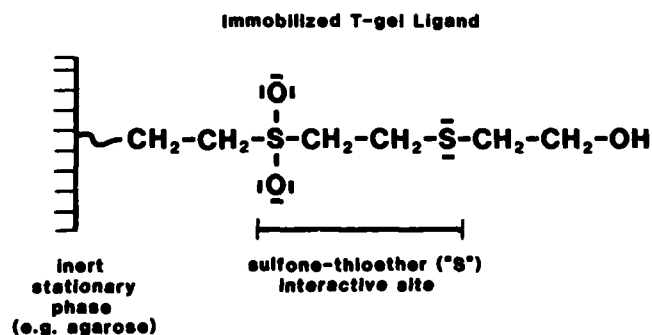
Figure 1.   Structure of T-gel ligand.

unique properties of thiophilic adsorption which, not unlike hydrophobic interaction (e.g., on aliphatic hydrocarbons), is salt-promoted and, in part, an entropy-driven process influenced by ordered water structure at the protein–solvent interface. As a prelude we should like to briefly review the utility of column chromatography for the preliminary experimental identification and characterization of new molecular interactions.

Chromatography and Molecular Adsorption

Molecular adsorption in gels, like the more complicated biological membranes, require considerations belonging to the realm of solid-phase chemistry and therefore is set apart from free solution in several respects. It is important to understand that a solid matrix, especially when it contains an interacting species (i.e., immobilized ligand), restricts diffusion and mobility of the ligand-ligate complex (Fig. 2). However, even though a rough approximation it remains instructive in this context to consider the kinetics and thermodynamics of aqueous solutions starting with the law of mass action:
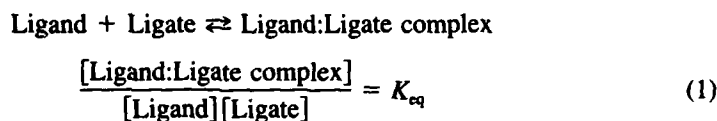
$$\text{Ligand} + \text{Ligate} \rightleftarrows \text{Ligand:Ligate complex}$$

$$\frac{[\text{Ligand:Ligate complex}]}{[\text{Ligand}][\text{Ligate}]} = K_{eq} \qquad (1)$$
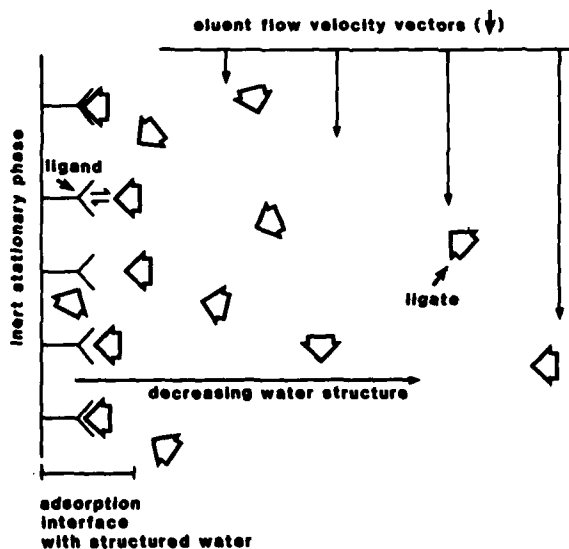
Figure 2. Dynamics of immobilized ligand-ligate interaction during chromatography.

where the ligate or solute, for example, may be a protein with a ligand-binding site. The equilibrium formation or association constant (K_{eq}) as defined above is related to the distribution or partition coefficient (K) of the interacting species in the stationary and mobile phases of a chromatographic bed and consequently also to the retardation of the ligate on its passage through the chromatographic bed of immobilized ligand (see Giddings, Ref. 4):

$$\frac{R}{1 - R} = \frac{[\text{ligate}]_m V_m}{[\text{ligate}]_s V_s} = \frac{V_m}{K V_s} \tag{2}$$

where R is defined as the fraction of ligate in the mobile phase (m) at equilibrium so that 1-R is the fraction of ligate in the stationary phase (s), that is, associated with immobilized ligand. The more familiar equation of Martin and Synge [5] is obtained upon solving the above equation for R:

$$R = \frac{V_m}{V_m + K V_s} \tag{3}$$

Note that for adsorption chromatography, the volume of the stationary phase (V_s) is probably best replaced by solute-accessible or effective surface area, which can depend on the size of the ligate.

Indeed, the chromatographic elution volume (V_e) of any given ligate (e.g., peptide or protein) is related to the physical properties of the adsorption matrix (e.g., agarose gel) such as particle size, geometry, and porosity as well as its essential and definitive chemical properties, namely, ligand type and density. In practice, however, these parameters act collectively to help determine the partition coefficient K and are related to the elution volume according to the following expression:

$$V_e = V_o + K V_s \tag{4}$$

or

$$K = \frac{V_e - V_o}{V_s} \tag{5}$$

Equation (5) clearly shows the necessity to carefully define reference elution volumes to accurately determine the extent and specificity of the molecular interaction between ligand and ligate. It was exactly these considerations that led to the discovery of thiophilic adsorption. To illustrate, consider a chromatographic process in a bed of granular solid support of total volume V_t consisting of a gel made up of a molecular network that is in part permeable to the solutes (ligates) under study (Fig. 3). If a substance cannot permeate the granular particles it travels only in the interstitial space between the grains (voids) and appears in the column eluate after passage of a volume of eluent equivalent to the volume of the voids, V_o. In other words, it travels with the speed of liquid eluent and $V_e = V_o$. Solutes are usually not separated $(K = O)$ under these conditions ("hydrodynamic" separation is to a limited extent possible). If there is no molecular interaction taking place but a solute permeates parts of the internal gel regions it lags behind the moving front of the liquid eluent and appears in the column eluate after some characteristic retention volume V_e, but
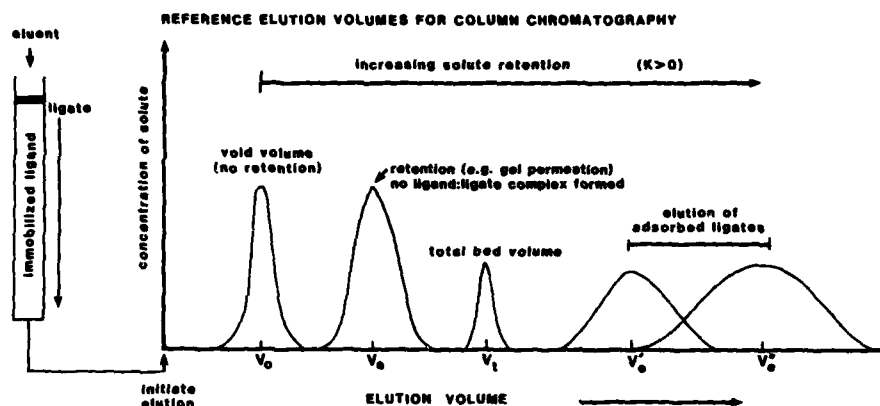
Figure 3.   Reference elution volumes necessary for characterizing ligate elution behavior
during chromatography.

before the total volume (i.e., $V_o < V_e < V_t$). This elution volume depends on the interrelationship in molecular size between the solute molecules and the pores of the matrix. The process is called molecular sieve or size-exclusion chromatography and was discovered by one of us (Porath [6] together with Per Flodin about 30 years ago) using cross-linked dextran (Sephadex). However, it is essential at this point to recognize that the nature of the solvent system or eluent itself can dramatically influence solute elution volume. This was found to be particularly true of thiophilic adsorption, a process which is differentially affected by the type and concentration of salt present in solution. If a water-structuring salt (e.g., sulfate or phosphate) is added to the solvent in relatively high concentration, solute molecules may be forced out of the bulk solution and accumulate at the phase boundary, that is, adjacent to the polymer chains making up the gel network of immobilized ligand. This salt-promoted accumulation delays diffusion of the solute under study and consequently the zone containing it is broadened and further retained compared to the case in salt-free solution. Under such circumstances, the solute now appears after an elution volume, V_e', later than V_e and often in a volume exceeding V_t (Fig. 3). This retardation phenomenon, caused by the salt, can be referred to as salt-promoted adsorption. It is apparently far more common than believed earlier. In fact, salt-promoted adsorption of aromatics to Sephadex was observed [7] already around the time of the discovery of molecular sieving [6]. Salts influencing protein adsorption on unchanged amphiphilic gels can be arranged in a so-called Hofmeister [8] series according to their effects on both protein and water structure:

Phosphates > sulfates > acetates > chlorides > nitrates > thiocyanates

Salts on the right side of this scale promote protein solubility (salting-in) and can, in fact, promote denaturation due in part to their disruptive influence on bound water structure. They are thus referred to as chaotropic [9]. Salts on the left side of this scale promote protein stability and precipitation (salting-out) and are said to have the opposite effect on water structure. To our knowledge, the mechanism(s) of salt-

promoted retardation or adsorption has never been explained. The salt-promoted concentration of ligate at the matrix boundary may sufficiently increase the equilibrium association constant so as to greatly enhance its interaction with the immobilized ligand. A much decreased affinity would be expected in the absence of water-structuring (antichaotropic) salts. In fact, adsorption capacity, as well as adsorption strength, appears to increase in parallel with increasing salt concentration for all kinds of salt-promoted adsorption. If the salt types responsible for promoting ligate adsorption to certain solid matrices (e.g., T gel) are acting by their influence on the organization of water at or near the surface of either ligand or ligate then a consideration of entropy is important.

Thermodynamic Considerations

The distribution coefficient K as expressed in Eqs. (3) and (4) may be related to the Gibbs' free energy of interaction by the equation:

$$\Delta G = -RT \ln K \tag{6}$$

ΔG may also be expressed as:

$$\Delta G = \Delta H - T\Delta S \tag{7}$$

where ΔH, the enthalpy, accounts for the energy of direct interaction between ligands and ligates and ΔS, the entropy term, accounts mainly for overall changes in water structure as a consequence of complex formation. The relative contributions of ΔH and $T\Delta S$ toward the solubility of nonpolar solutes in water as well as the formation (stabilization) of hydrophobic bonds between nonpolar solutes in water can provide insight into the contributions of water structure during salt-promoted ligate interactions with immobilized ligands. It is suggested that the energy required ($\Delta H > O$) to solubilize a nonpolar solute is more than offset by the energy gained ($\Delta H < O$) due to the formation of new hydrogen bonds at the solute-solvent interface [10]. The association of two or more nonpolar groups (or any groups, e.g., ligand:ligate, with bound or ordered water molecules) in an aqueous environment would then be expected to decrease the ordered water structure at their interacting surfaces and result in an increase in entropy. The ΔH term in hydrophobic interactions is small due to the presence of only weak dispersion forces. Thus, generic "hydrophobic interactions" are thought to be entropically "driven" ($\Delta G \approx -T\Delta S$) and favored ($\Delta G < O$) with increased temperature. A simplistic schematic of this process is shown in Figure 4. However, is the favorable entropy change associated with the displacement of bound water a major factor in stabilizing other types of protein-ligand interactions? What contribution does ligand structure contribute to this process?

Thiophilic Adsorption

Distinguishing Structural Requirements of the Immobilized Ligand

Perhaps a reasonable reference point from which to distinguish thiophilic adsorption from the more widely observed but lesser discriminating hydrophobic interac-

HYDROPHOBIC INTERACTIONS:

SOLUTE-BOUND WATER AND CHANGE IN ENTROPY

bonding energy (ΔH) low

nonpolar solutes   hydrophobic bond   less-structured
in aqueous environment   formed   water
with structured water
(<)
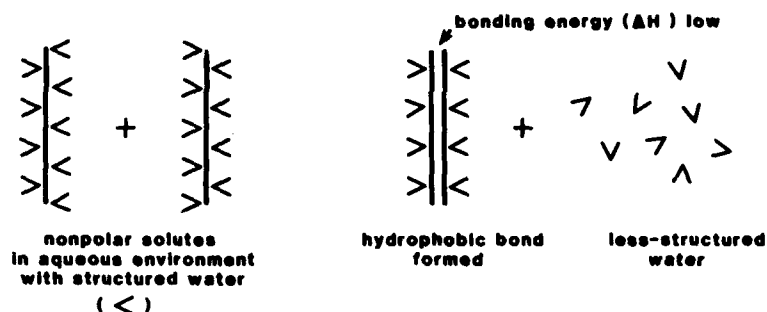
Figure 4. A simple explanation of hydrophobic interaction. The ΔH is small because it is made up of contributions due to weak dispersion forces. The $T\Delta S$ term becomes large due to disordering of the expelled water. These terms therefore give $\Delta G < O$ and constitute a "driving force" for complex action.

tions is to consider the structure of the interacting ligands involved. Although widely varied in structure, representative ligands for the generic hydrophobic adsorption of peptides and proteins to gels with a hydrophilic network structure include $C_4, C_6, C_8 \ldots C_{18}$ linear aliphatic hydrocarbon chains (e.g., Ref. 11). Now let us examine the structure of the T gel and compare it to several other thiophilic as well as nonthiophilic ligands recently synthesized. The following immobilized ligands have all been synthesized and are at various stages of evaluation using chromatography in an effort to determine the precise contribution of individual atoms toward the thiophilic adsorption process:

T gel   Agarose-O$-CH_2-CH_2-SO_2-CH_2-CH_2-S-CH_2-CH_2-OH$
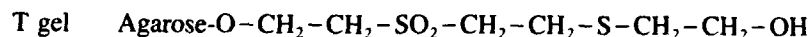
Aliphatic T gel derivatives with thiophilic properties:

Agarose-O$-CH_2-CH_2-SO_2-CH_2-CH_2-S-CH_2-CH_2-SH$

A gel   Agarose-O$-CH_2-CH_2-SO_2-CH_2-CH_2-NH_2$

N gel   Agarose-O$-CH_2-CH_2-SO_2-CH_2-CH_2-NH-CH_2-CH_2-OH$

Agarose-O$-CH_2-CH_2-SO_2-CH_2-CH_2-NH-CH_2-COOH$

AN gel   Agarose-O$-CH_2-CH_2-SO_2-CH_2-CH_2-NH-CH_2-C\equiv N$

PTN gel   Agarose-O$-CH_2-CH_2-SO_2-CH_2-CH_2-NH-C$
$CH_2-C\equiv N$
$C\equiv N$
C
$C\equiv N$

DPA gel Agarose-O$-$CH$_2$-CH$_2$-SO$_2$-CH$_2$-CH$_2$-N$\Big\langle$ $^{\text{CH}_2-\text{C}\equiv\text{CH}}_{\text{CH}_2-\text{C}\equiv\text{CH}}$
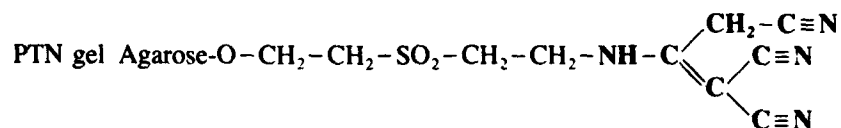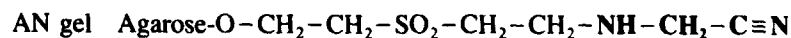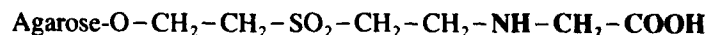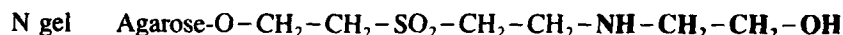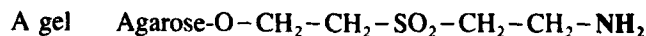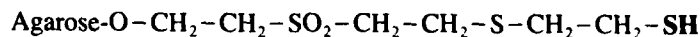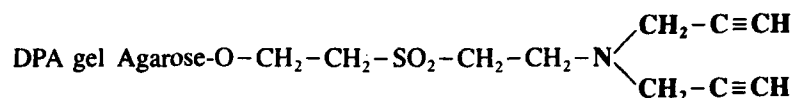
Other T gel derivatives contain aromatic substitutions and are referred to as -S-π adsorbents:
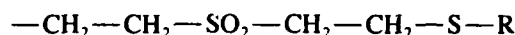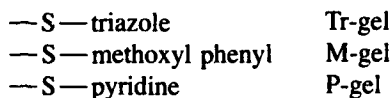
$-$CH$_2$$-CH_2$$-SO_2$$-CH_2$$-CH_2$$-S-$R

e.g.,

$-$S$-$triazole	Tr-gel
$-$S$-$methoxyl phenyl	M-gel
$-$S$-$pyridine	P-gel

These -S-π gels exhibit distinctive adsorption properties which are discussed elsewhere [12, 13].

The T-gel ligand, although uncharged, appears to be the most effective one in terms of strength and capacity. Still another hydroxyl ($-$OH) or even a thiol ($-$SH) can be introduced in the T-gel ligand with characteristic T-gel adsorption capacity essentially retained [1].

$-$O$-$CH$_2$$-CH_2$$-SO_2$$-CH_2$$-CH_2$$-S-CH_2$$-$CHOH$-CH_2$$-$OH

$-$O$-$CH$_2$$-CH_2$$-SO_2$$-CH_2$$-CH_2$$-S-CH_2$$-$CHSH$-CH_2$$-$OH

Further, the ligand:

$-$O$-$CH$_2$$-CH_2$$-SO_2$$-CH_2$$-CH_2$$-S-CH_2$$-CH-CH-CH_2$$-$SH

                                           OH   OH
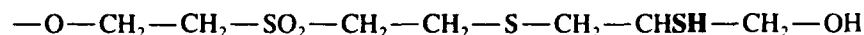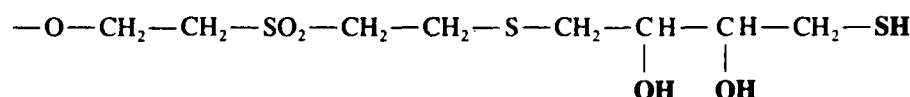
has a similar adsorption capacity [1]. Thus it seems that thiophilic interaction, from the view of the ligand, should in fact be considered to be hydrophilic rather than hydrophobic. Like the T gel, the A gel and DPA gel (J. Porath, in preparation) adsorb immunoglobulins and α_2-macroglobulin from human serum. The A gel is more hydrophilic than hydrophobic and the hydrophobicitiy of the DPA gel is not strong enough to make it an adsorbent for serum albumin.

A ligand with only thioether sulfur, as obtained by coupling mercaptoethanol to oxirane-agarose

$-$CH$_2$$-$CHOH$-CH_2$$-S-CH_2$$-CH_2$$-$OH
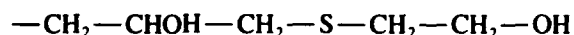
does not convert agarose to a protein adsorbent, neither will a ligand containing only a sulfone group in an aliphatic surrounding be thiophilic. In fact the ligand shown below appeared to be relatively inert as protein adsorbent [1].

$-$O$-$CH$_2$$-CH_2$$-SO_2$$-CH_2$$-CH_2$$-$OH

Similarly, the following structure (derived from oxirane-activated agarose):

$$-CH_2-CHOH-CH_2-O-CH_2-CH_2-OH$$

although similar to the T-gel ligand does not have a sulfone group and was found not to have thiophilic (or any other) protein adsorption property under the conditions evaluated [14]. Further, when the propenetrinitrilo (PTN) group is attached via an oxirane coupler $(-O-CH_2CHOH-CH_2-PTN)$ instead of the divinylsulfone coupler $(-SO_2-CH_2-CH_2-PTN)$, little or no protein adsorption occurs [14]. The thioether sulfur and the sulfone group likely cooperate when interacting with the aromatic side chains in the proteins. It currently appears as though a necessary and probably sufficient structural feature for thiophilicity is contained in the sequence: $-SO_2-CH_2-CH_2-X-$ where, in the absence of aromatic substitutions, X can be any atom with a lone electron pair.

How can this structure so specifically interact with the surface of certain proteins? Is the acceptor site a π-electron system such as the indole nucleus of a tryptophan residue located in an accessible cavity or at the surface of the protein molecule? We believe that some kind of ring structure may be formed by transfer of electrons or protons between the ligand site ("S site") and the corresponding countersite on the protein. This electron-donor-acceptor relationship is schematically depicted in Figure 5.

Figure 5. Possible electron-donor-acceptor thiophilic interaction mechanism.

Properties that Distinguish Thiophilic Adsorption from Generic
Hydrophobic Interactions

We must recognize that as we create ligands increasingly specific for a certain architecture or binding site(s) on the surface of proteins it necessarily becomes increasingly difficult to categorize these proteins into groups according to their specific adsorption behavior (as is the case, for example, in anion- or cation-exchange chromatography). However, because new types of protein adsorption phenomena will inevitably be discovered and exploited, attempts should be made to formulate these observations into mechanisms which recognize and successfully challenge existing boundaries of group fractionation theory. We find that proteins often used as model standards to characterize the adsorption behavior of hydrophobic gels (e.g., human and bovine serum albumins, ovalbumin, myoglobin, ribonuclease A, and cytochrome c) do not interact with the T gel under conditions where immunoglobulins are effectively adsorbed [2]. Similarly, under identical conditions, alcohol dehydrogenases 1 and 2 (from *Zymomonas mobilis*) passed through the T gel unadsorbed but were strongly adsorbed to either C_4 or C_8 hydrophobic gels attached tandemly [15].
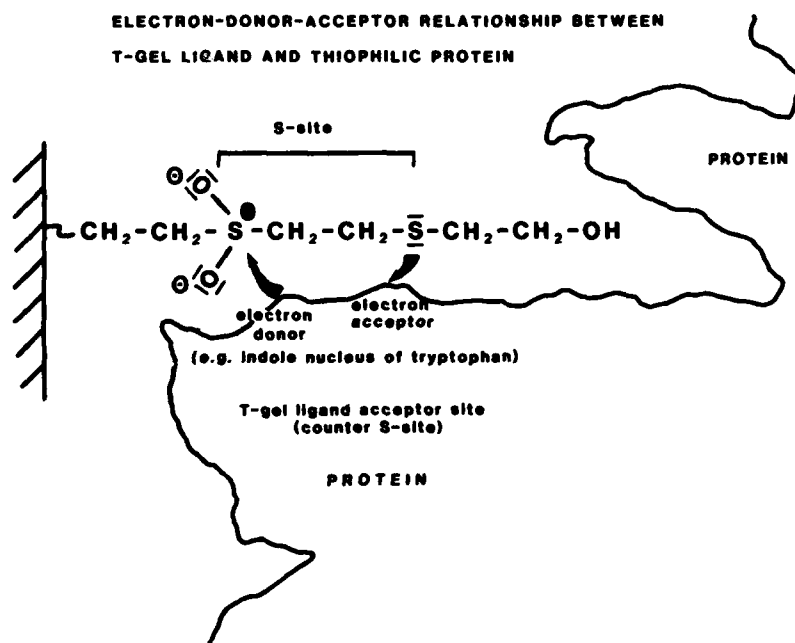
So, we now review arguments in favor of the view that thiophilic interaction is different in nature from hydrophobic interaction as is exerted between two alkyl chains in aqueous solution. The hydrophobic gel (H gel) referred to in the following, contains an octyl (C_8) ligand linked via 1,4-butanediol diglycidyl ether to the matrix in a gel consisting of 6% agarose [16].

Consider two chromatographic beds connected in tandem, the upper one containing T gel, the lower one H gel [1]. The gels were equilibrated with a 0.05 M Tris-HCl buffer of pH 7.5 containing 0.5 M potassium sulfate, the latter being the adsorption-promoting, water-structuring salt. A sample of human serum, likewise equilibrated with the sulfate-containing buffer, was introduced at the top of the T-gel column followed by washing with sulfate-containing buffer. Some proteins were adsorbed in the T-gel column, other proteins in the H-gel column. The remaining proteins were eluted unretained by either column. After disconnecting the composite T → H column the proteins were desorbed separately from each bed simply by eluting with sulfate-free buffer. Gel electrophoretic analysis of the two separate fractions revealed that the H gel, as expected, had depleted the serum with respect to albumin. The T gel, surprisingly, had selectively removed the immunoglobulins and α_2-macroglobulin. Reversing the tandem order of the beds (i.e., H → T) gave identical results thus proving the difference in kind of interaction. The H gel and T gel also respond differently in protein adsorption behavior as a function of salt type used.

Potassium sulfate is just one of several simple salts that can be used to increase adsorption on both the T gel and H gel. Shortly after its first description, hydrophobic adsorption to an uncharged gel was shown to be promoted by chlorides [17] and, later, by phosphates and sulfates. In contrast, thiophilic adsorption is effectively prevented by sodium chloride. In fact, 0.5 M sodium chloride has been used to further increase the specificity of the T gel for immunoglobulins by preventing adsorption of α_2-macroglobulin [2]. But even immunoglobulins are not adsorbed well if high enough concentrations of sodium chloride are present [18]. Conversely, sodium chloride is quite effective at promoting adsorption of serum albumins to hydrophobic ad-

sorbents [11, 18]. Thus, the division line between those salts promoting adsorption and those promoting desorption is located differently in the Hofmeister series for H and T gels, and we have another argument in favor of our proposal that H and T interactions are distinctly different in nature.

We can next compare the effects of temperature on ligate adsorption by the H and T gels to help distinguish the two adsorption mechanisms involved. We have indicated in our discussion of Eq. (7) why hydrophobic interactions increase with temperature. This appears to have been verified experimentally [e.g., Refs. 10, 11]. However, immunoglobulins are more strongly adsorbed to the T gel with lower temperature [1]. What about the simpler model substances? The tripeptide Leu Leu Leu is rather hydrophobic and was not very strongly adsorbed to the T gel. Nevertheless, adsorption on the T gel became even weaker and approached zero with increasing temperature. The opposite temperature effect was observed (as expected) for the H gel and may provide a basis for our third argument in favor of the difference in interaction. An interesting and surprising set of results were obtained with the dipeptide TrpTrp. TrpTrp showed very strong thiophilic and hydrophobic adsorption properties, and its interaction with both the T gel and H gel was quite significantly *decreased* with *increasing* temperature. The temperature-dependent behavior of TrpTrp on the H gel was quite contrary to that expected based upon information often presented in the literature. More work is underway to better evaluate temperature effects on the adsorption behavior of model (structurally defined) ligates.

Finally, by constructing a ligand composed of two segments or interactive sites, one common to the H ligand and the other to the T ligand, a superposition of both interaction effects is obtained. Thus, coupling of octanethiol to a divinylsulfone-activated matrix produces a gel which adsorbs albumin as well as immunoglobulin and in both cases adsorption is very much stronger than on gels with the simpler composed ligands. This observation also argues against H and T interactions being identical in nature. The superposition principle shows that the H and T interactions do not exclude, but rather reinforce each other although both may involve related target or countersites on the protein surface. In passing, it may be worth mentioning that the use of double- or triple-affinity principles seems to offer a new unexplored road to immobilize, on solid support, proteins and protein-containing aggregates or biological systems of higher order such as viruses and cells.

Proposed Thiophilic Interaction Mechanisms

Let us now turn to the problem of how to identify the thiophilic interaction site(s) on proteins or other ligates. Amino acids are too weakly adsorbed to give conclusive results. To intensify the interaction we tested some very simple homopeptide structures. The involvement of aromatic amino acids during peptide adsorption was hereby suggested. When we compared the results of homopeptide adsorption evaluations (relative partition coefficient) on both the H gel and T gel, we found the following series of increasing adsorption strength:

H: immunoglobulins ≪-Tyr- < -Phe- < -Trp-≪ serum albumin

T: serum albumin ≪-Tyr- < -Phe- < -Trp-≪ immunoglobulins

The relative elution order among the homopeptides analyzed was different from that found earlier for Sephadex (and agarose):

-Phe- < -Tyr- < -Trp-

where in this latter case, the interaction is much weaker. We are continuing to investigate the thiophilic behavior of synthetic model peptides as well as naturally occurring peptides of various sizes. While it may continue to be instructive to evaluate structurally defined peptide interactions with the T gel under varied experimental conditions, interpretation of the results may not be entirely relevant to the mechanism(s) by which thiophilic proteins are adsorbed.

The identification of specific T-gel ligand acceptor sites on the surface of a thiophilic protein is much more complicated. To promote further discussions, Figure 6 illustrates, schematically, the interaction of ligand and counterligand in a hydrophobic pocket at the protein surface. Although thiophilicity may be exerted without intervening antichaotropic salts, the latter effectively forces the interacting species (T-gel ligand and its acceptor site) into close proximity which, in addition to the lower dielectric constant within the pocket, accounts for a considerable reinforcement in the interaction. This situation may resemble steroid hormone-receptor association and certain antigen–antibody complex formations as well as other forms of biologically significant complexation phenomena.

In an effort to better understand the thiophilic adsorption process, the ligate or protein-acceptor site needs to be quantitatively characterized in terms of its interaction with the T-gel ligand. Toward this end, we have selected as models a variety of

Figure 6.   T-gel ligand-binding site on protein.

physically well-characterized proteins. The three-dimensional structures of some of these proteins have been solved by crystallization and X-ray diffraction.

The relative thiophilicity of these proteins was first evaluated using the T-gel and buffer conditions (specifically 10% (0.76 M) ammonium sulfate in 20 mM HEPES buffer pH 7.5) consistent with the selective yet maximal adsorption of immunoglobulins from unfractionated serum [2, 3] or cell culture medium. Under these experimental conditions, the model proteins could be arbitrarily classified as either thiophilic or nonthiophilic based upon percent of total protein adsorbed by a given bed volume of T gel (Table I).

Adsorption of the thiophilic proteins was then found to be influenced by the type and concentration of water structure-forming salt as well as pH, with an adsorption minima at pH 5–6 [2, 3].

We have since evaluated means of experimentally altering (increasing) the apparent thiophilicity of certain model proteins to better understand the relationship between column capacity and selectivity. Accordingly, the affinity of several model proteins for the T gel was evaluated under equilibrium conditions (batch incubations) with increasing concentrations (up to 20%) of water structure-forming salt (ammonium sulfate). The data were evaluated according to the method of Scatchard [19]

$$\text{Bound/Free} = \frac{-1}{Kd}\,(B) + \frac{B\max}{Kd}$$

TABLE I.   Comparative adsorption of model proteins to the T gel.

Model proteins	Percent adsorption to T gel
I. Thiophilic	
immunoglobulins (human serum)	98
lentil lectin (lens culinaris)	81
carboxypeptidase A (bovine pancreas)	80
trypsin (bovine pancreas)	50
trypsin inhibitor (Kunitz, soybean)	19
trypsin-soybean trypsin inhibitor	97
chymotrypsin (bovine pancreas)	18
II. Nonthiophilic	
ovalbumin (chicken egg white)	0
ribonuclease A and S (bovine pancreas)	0–1
cytochrome c (horse and tuna heart)	0
myoglobin (sperm whale)	0–4
carbonic anhydrase I (human)	2–3
pancreatic trypsin inhibitor (bovine pancreas)	4–6
albumins (human and bovine serum)	4–5
α chymotrypsinogen A (bovine pancreas)	6–7

Individual proteins (1 mg/2 mL) were applied to identical T-gel columns (0.5 mL) in column equilibration buffer consisting of 10% ammonium sulfate in 20 mM HEPES, pH 7.5. Unbound proteins were washed through with 4–6 mL of column equilibration buffer. Bound proteins were eluted with 20 mM HEPES buffer, pH 7.5. Data are presented as percent of total protein adsorbed. An arbitrary cut off at 10% was chosen to distinguish thiophilic from nonthiophilic proteins under the conditions defined.

to determine the equilibrium dissociation constant (Kd) and the protein-binding capacity $(Bmax)$ per unit T-gel ligand. Representative data are shown in Figure 7. The Scatchard analyses suggest that certain proteins (e.g., immunoglobulins and lysozyme) may be interacting with the T-gel ligand via more than one ligand-binding site with varying but relatively high affinities $(Kd = 1-8 \times 10^{-6})$. The higher affinity interaction (slope shown) has been corrected for the influence of an apparent second, lower affinity $(8 \times 10^{-6}M)$ interaction site by the method of Rosenthal [20]. The additional possibility of multiple interaction sites cannot be excluded at this time. Other proteins (e.g., α chymotrypsin) appear to be interacting through one or a single class of lower affinity binding sites $(Kd = 400 \times 10^{-6}M)$. Interestingly, some proteins (e.g., cytochrome c and myoglobin) show little or no thiophilic behavior $(K = 0)$ in even 20% ammonium sulfate. These and similar data were compared to the observed partition coefficient [calculated from V_e using Eq. (5)] of these same
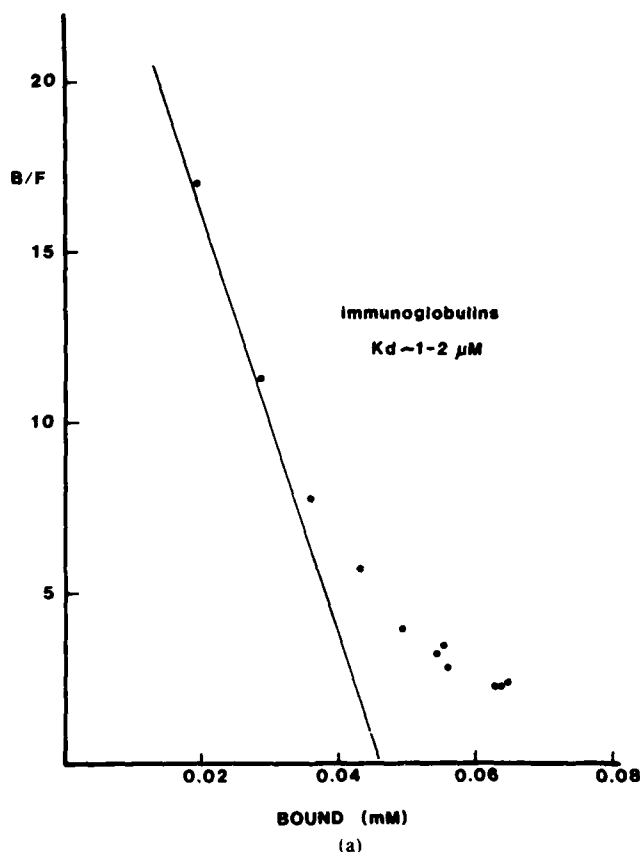
Figure 7. Scatchard analyses of protein interaction(s) with the T-gel thioether-sulfone ligand. Increasing concentrations (10 points) of purified protein was incubated at constant volume and temperature (22°C) with a known quantity of T-gel until equilibrium was attained (<30 min). Gel-bound protein was calculated from the difference of total protein added and free protein at equilibrium as determined by absorbance.

proteins during (1) zonal elution chromatography under identical, isocratic conditions and (2) gradient elution chromatography using inverse (e.g., 20% → 0%) salt gradients. The relative elution order of the model proteins evaluated is shown in Table II. It would appear from preliminary data that the equilibrium Kd values for the thiophilic proteins in this series can be correlated with their measured partition coefficients. However, more data need to be obtained under a variety of experimental conditions before a satisfactory quantitative relationship is derived. So, an increase in salt concentration or "driving force" can promote the binding of additional proteins to the T gel, however such an apparent relaxation in selectivity may be a function of perturbed water/protein structure and/or an additional (new), but extremely weak, affinity property of these proteins for the ligand. Details of this investigation are to be presented elsewhere [21].

Finally, the equilibrium affinity constant (Kd) and/or partition coefficient (K) of ..iese proteins for the T gel is being examined for possible correlation with type and accessibility (available surface area) of individual amino acids at the protein's surface. Factors including the overall density and relative proximity of both charged as

BOUND (mM)

(b)
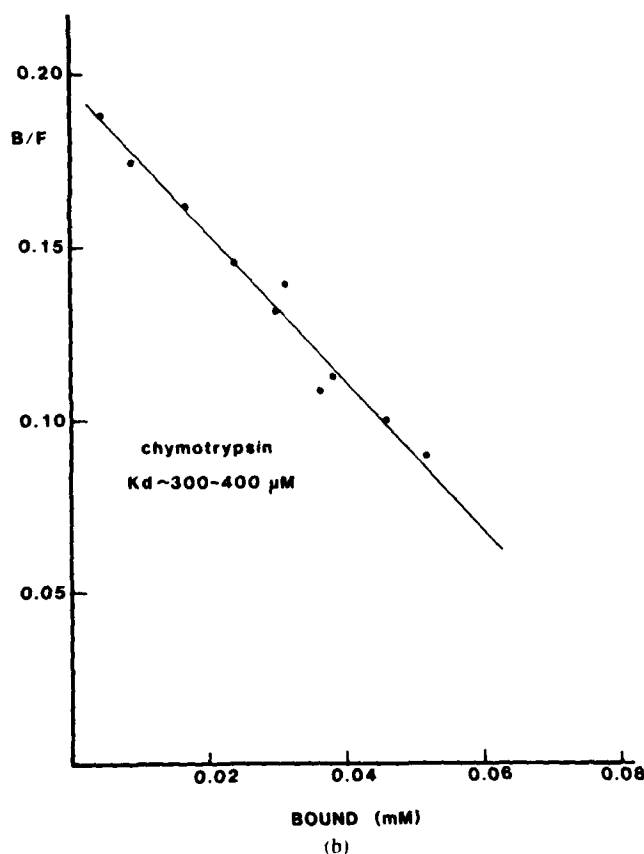
Figure 7.   (Continued.)

TABLE II.   Thiophlic adsorption: relative elution behavior of model proteins.

Protein and gradient elution order	Calculated relative elution volume (mL)[a]
1. Cytochrome c	1
2. Myoglobin	2.1
3. Ribonuclease S	3.2
4. Bovine serum albumin	3.3
5. Pancreatic trypsin inhibitor	3.5
6. Ribonuclease A	3.8
7. Human serum albumin	4.0
8. Carbonic anhydrase I	5.0
9. $\propto$ Chymotrypsinogen A	6.7
10. Ribonuclease S-protein	7.5
11. Soybean trypsin inhibitor	8.3
12. $\propto$ Chymotrypsinogen A	9.0
13. Lysozyme	9.7
14. Immunoglobulins (gamma)	11.0
15. Insulin	11.2
16. Carboxypeptidase A	11.5

[a]Relative elution volume $= \dfrac{\text{elution volume of protein}}{\text{elution volume of cytochrome c}}$.

Individual proteins were loaded onto a 2 mL (1 × 2.5 cm) T-gel column equilibrated with 20% (1.5 M) ammonium sulfate, 0.5 M sodium chloride and 20 mM HEPES pH 7.5. Elution of bound proteins was initiated during development of a reverse ammonium sulfate gradient (20% → 0%). Relative elution order was monitored by absorbance at 280 nm.

well as nonpolar and aromatic surface patches are being considered. There does not appear, as yet, to be any one overall structural feature (e.g. net surface charge or isoelectric point, molecular mass, subunit structure, disulfide linkages, carbohydrate, etc.) or specific surface feature which is common to the thiophilic or, conversely, the apparently nonthiophilic proteins investigated.

Thiophilic Adsorption in Immunochemistry and Biotechnology

Finally, it is important to illustrate just how selective thiophilic interaction can be. The T gel can be utilized alone or in a tandem column arrangement with other gels to accomplish highly selective fractionations of complex protein mixtures when the aim is the isolation of precious bioactive materials. We have already reported the selective immobilization of human serum immunoglobulins [2]. It is also possible to isolate monoclonal antibodies from ascites fluid and hybridoma cell culture media in one rapid step using mild conditions consistent with preserved antibody function.

The first example (Fig. 8) illustrates fractionation of hybridoma cell culture media containing a monoclonal antibody. Only a relatively small fraction of the proteins were adsorbed to the T gel. Upon removal of the water-structuring salt from the column eluent solution the major portion of the proteins adsorbed came out in a sharp elution zone. Upon further washing (or upon inclusion of NaCl), a second elution re-
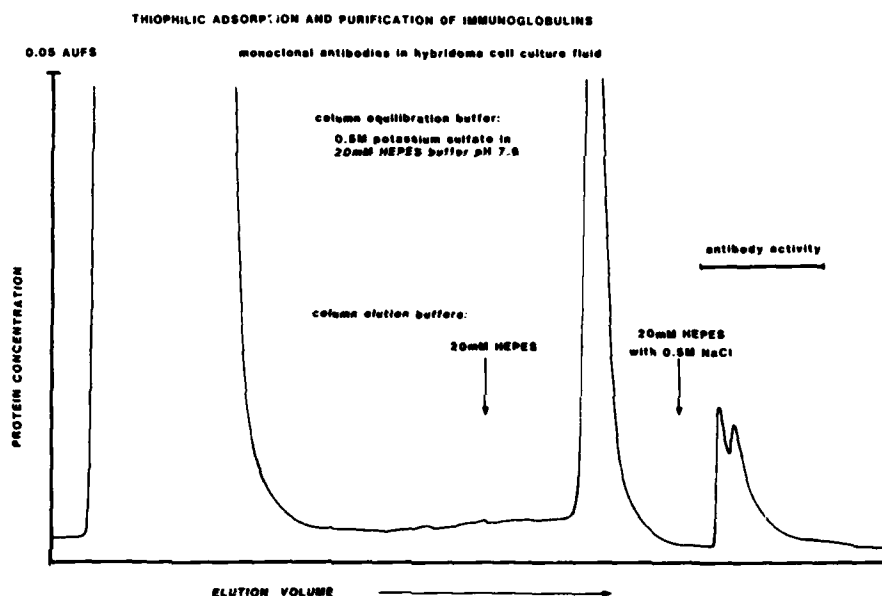
Figure 8.   T-gel elution profile of monoclonal antibodies produced during hybridoma cell
culture. Pooled fractions were analyzed for purity as shown in Figure 9.

gion well separated from the first one, appeared in the eluate. This zone contained
most of the functional antibody present in the original sample. It was an almost pure
antibody as determined by analytical gel electrophoresis under native and denaturing
conditions (Fig. 9). In this way, by a single operation, nearly complete purification
can routinely be achieved under mild conditions with typical recoveries of over 90%.
This technique, simple as it is, represents a drastic improvement over existing meth-
ods for isolation of monoclonal antibodies and is applicable on any scale, at least in
principle. The important point of this illustration, however, is not simply one of im-
munoglobulin purification, but rather, selective recognition and immobilization of a
protein. The unanswered question remains the molecular mechanism by which a sim-
ple, chemically defined ligand, such as that operative during thiophilic adsorption,
can recognize a protein surface with such specificity.

A significant problem associated with monoclonal antibody purification from
serum-dependent hybridoma cell culture media has been removal of nonspecific or
background (typically bovine) immunoglobulins. In efforts to circumvent this prob-
lem and to further address immunoglobulin-related difficulties with serum-dependent
cell cultures of all types, we have evaluated the use of thiophilic adsorption to selec-
tively eliminate immunoglobulins from bovine and equine sera. Selective removal of
immunoglobulins present in bovine serum has been accomplished. Bovine sera with
immunoglobulin levels of 10–12 mg/mL prior to thiophilic adsorption had residual
levels at or below 1–5 ng/mL after a single thiophilic adsorption step. The calf
serum thus purified was used in hybridoma cell culture media to evaluate cell growth
rates and production of monoclonal antibodies. These properties appeared unaltered
using culture media prepared with the purified versus untreated fetal calf serum [22].
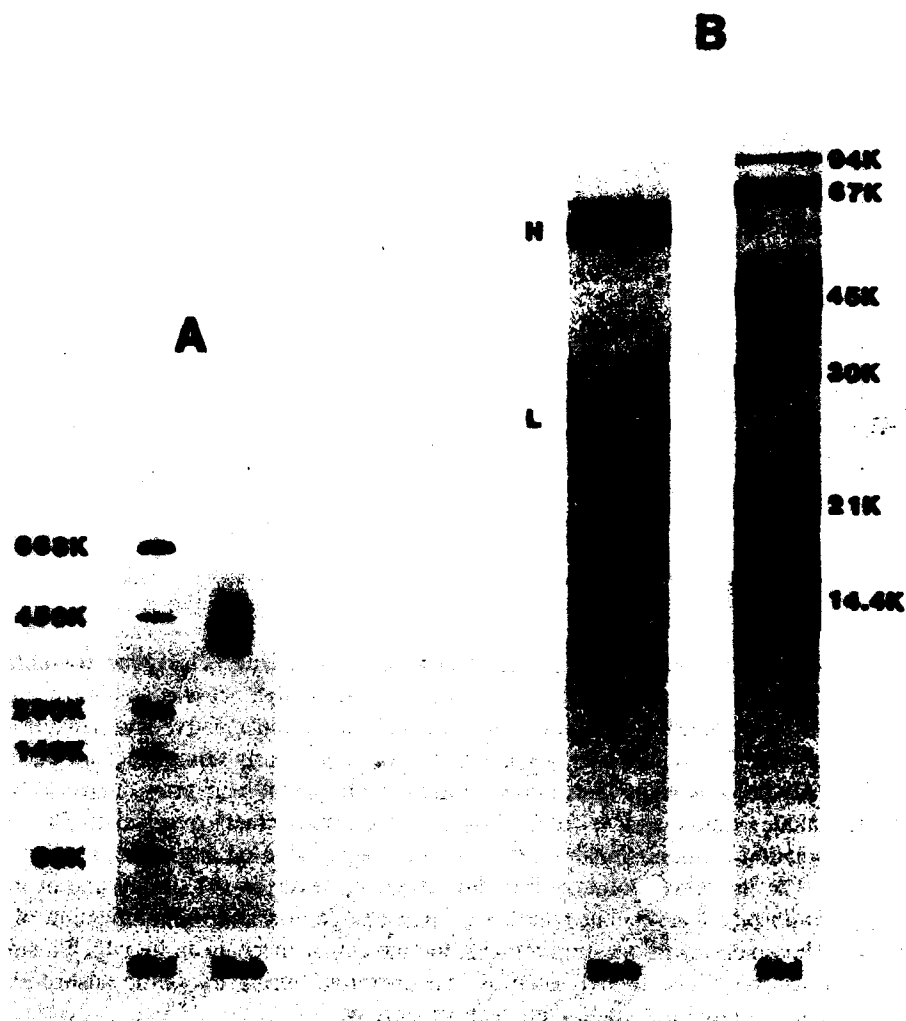
Figure 9. Electrophoretic purity of monoclonal antibodies purified by thiophilic adsorption as described for Figure 8. Panel A represents pooled fractions analyzed by gradient gel electrophoresis under nondenaturing conditions and panel B represents the same fractions analyzed by sodium dodecylsulfate polyacrylamide gel electrophoresis. Proteins were localized by Coomassie blue dye and silver staining. Experimental details are provided elsewhere [2].

Monoclonal antibodies present in the culture media were subsequently purified to near homogeneity in one step using thiophilic adsorption chromatography as described for Figure 8. We are also investigating the use of T-gel-purified bovine sera for the more efficient growth of certain viruses in culture. We therefore consider thiophilic adsorption to be a major advancement in the field of biotechnology and believe that more detailed analyses of the precise interaction mechanisms involved may lead to a more direct approach in our search for group-selective protein adsorbents.

Acknowledgments

Bibliography

[1] J. Porath, F. Maisano, and M. Belew, FEBS Lett. **185**, 306 (1985).
[2] T. W. Hutchens and J. Porath, Anal. Biochem. **159**, 217 (1986).
[3] T. W. Hutchens and J. Porath, Biochemistry (1987) in press.
[4] J. C. Giddings, *Dynamics of Chromatography. Principles and Theory* (Marcel Dekker, New York, 1965).
[5] A. J. P. Martin and R. L. M. Synge, Biochem. J. **35**, 91 (1941).
[6] J. Porath and P. Flodin, Nature **183**, 1657 (1959).
[7] J. Porath, Biochim. Biophys. Acta **39**, 193 (1960).
[8] F. Hofmeister, Arch. Exp. Path. Pharm. **17**, 247 (1888).
[9] W. H. Sawyer and J. Puckridge, J. Biol. Chem. **248**, 8429 (1973).
[10] H. S. Frank and M. W. Evans, J. Chem. Phys. **13**, 507 (1945).
[11] S. Shaltiel, in *Affinity Chromatography and Biological Recognition*, I. M. Chaiken, M. Wilchek and I. Parikh, Eds. (Academic Press, New York, 1983) pp. 229–239.
[12] J. Porath, M. Belew, F. Maisano, and B. Olin, IUPAC Symp. Vol. (1985).
[13] J. Porath, Biotechnol. Progr. **3**, 14–21 (1987).
[14] J. Porath, Biopolymers **26**, S193–S204 (1987).
[15] R. Scopes and J. Porath, (in preparation).
[16] F. Maisano, M. Belew, and J. Porath, J. Chromatogr. **321**, 301 (1985).
[17] J. Porath, L. Sundberg, N. Fornstedt, and I. Olsson, Nature **245**, 465 (1973).
[18] J. Porath, J. Chromatogr. **376**, 331 (1986).
[19] G. Scatchard, Ann. N.Y. Acad. Sci. **51**, 660 (1949).
[20] H. E. Rosenthal, Anal. Biochem. **20**, 525 (1967).
[21] T. W. Hutchens, T. T. Yip, and J. Porath, (in preparation).
[22] T. W. Hutchens, M. -R. Ruan, J. Andersson, and J. Porath, (in preparation).

Photoelectron Spectroscopy of Biologically Active Molecules. 14. Some Analgesic-Antipyretic and Anti-Inflammatory Agents

L. KLASINC, B. KOVAČ, A. SABLJIĆ, AND S.P. MCGLYNN

The Rudjer Bošković Institute, Zagreb, Croatia, Yugoslavia; and Department of Chemistry, Louisiana State University, Baton Rouge, Louisiana, U.S.A.

Abstract

The HeI photoelectron (PE) spectra of the drugs antipyrine (phenazone, 1), aminopyrine (amidopyrine, 2), phenylbutazone (3), and indomethacin (4) are reported. These drugs possess analgesic-antipyretic and anti-inflammatory activity, and have been in clinical use for many years. Their electronic structure, which is important to their biological activity, is determined by the application of composite molecule methods to an analysis of the PE spectra.

Introduction

The relief of pain has always been a medical priority. The chemical (pharmacological) approach to this objective was (and is) to discover substances that provide such relief. As a result, numerous agents have been tested and introduced into medicine. Unfortunately, most of them are addictive, that is, most of the older agents are narcotics. During the last century, however, numerous non-narcotic mild analgesics have been discovered, the analgesic property often being accompanied by antipyretic and anti-inflammatory activity. The most prominent representatives of these latter agents belong to two classes of compounds: the salicylates and the *p*-aminophenols.

In a recent paper [1], we used the experimental technique of photoelectron (PE) spectroscopy and the theoretical technique of the composite molecule method (CMM) [2] to investigate the electronic structure of several compounds belonging to, or related to these two classes. The accent in the investigation was placed on elucidation of the interactions of an amidic group with an attached phenyl, the phenyl substituent being appended either at the amidic carbon or nitrogen centers. It was found that a substantial difference of electronic structure is associated with the two amidic sites: the benzamides resemble the benzoic acid derivatives [3] whereas th anilides resemble the anilines. The purpose of the present work is to extend these investigations to some more complicated analgesic-antipyretic and anti-inflammatory agents, namely antipyrine (1), aminopyrine (2), phenylbutazone (3), and indomethacin (4), which last belongs formally to both groups [4]. Chemically speaking, 1–3 are pyrazolones and indomethacin is an indole. However, the chemical structures reveal the existence of a common part 5 which, with X = H, Y = Me, and Z = OH or OEt, is representative of the well-known analgesics paracetamol or phenacetine, respectively. Since
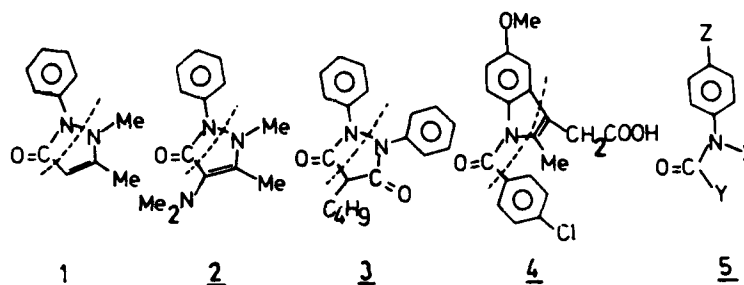
we had already investigated the electronic structure of paracetamol and phenacetine [1] as well as that of several pyrazolones [5] and indoles [6], it seemed appropriate to ascertain if different X, Y, and Z substituents introduced substantial changes of electronic structure in the compounds 1–4, that might affect biological activities and modes of action. Namely, the electronic structure of drugs as manifested either through their ionization potentials and/or calculated orbital energies has been shown repeatedly to correlate with biological activity [7–9].

Of the above compounds, 1 and 2 are the oldest, having been introduced into medicine at the end of the 1800s. They found wide use as antipyretics, analgesics, and anti-inflammatory agents. However, since compound 2 produces agranulocytosis (bone marrow toxicity) both drugs have disappeared from the market.

Compound 3, introduced in 1949 for treatment of rheumatoid arthritis and related disorders, is an effective anti-inflammatory agent whose use is limited by toxicity. Its therapeutic effects are similar to the salicylates but, because of its toxicity, it should not be used routinely as either analgesic or antipyretic. It undergoes extensive metabolic transformation in humans, the most significant primary reactions involving glucuronidation and phenyl-ring hydroxylation. Thus, oxyphenbutazone, the p-hydroxylated metabolite ($Z = $ OH in 5) with activity similar to that of the parent drug, is also offered on the market. Interestingly, compound 4 evolved from an organized laboratory search for anti-inflammatory drugs and has been in use since 1963. The anti-inflammatory effects are evident in the treatment of rheumatoid and other types of arthritis, including acute gout. Indomethacin exhibits central and peripheral analgesic properties which are distinct from its anti-inflammatory effects; its antipyretic effect is demonstrable in patients with fever. More about 1–4 and related agents can be found in Ref. 10.

Experimental

Compounds. Compounds 1–4 were of high purity and were supplied by the Bureau of Drug Control, Zagreb.

Spectra. The HeI PE spectra were recorded (FWHM ~ 20 meV) on a Vacuum Generators UV-G3 spectrometer [8]. Temperatures of 120, 100, 170, and 190°C at the inlet system were used in studies of 1, 2, 3, and 4, respectively. The energy scale was calibrated using mixtures of Ar and Xe.

Results

The low-resolution PE spectra of the pyrazolones 1–3 are shown in Figure 1 and that of compound 4 is given in Figure 2. The expanded low energy part of the spectra of compounds 1 and 2 is shown in Figure 3. Numbers at the top of PE bands correspond to vertical ionization energies, $E_{i,v}$/eV. A correlation of the electronic energies of aniline [16], acetanilide [1], 3, 1,2-diphenylpyrazoline [13], hydrazobenzene [13], and 1 and 2 are given in Figure 3; those of indole-3-acetic acid [6], indole [6], 2-methylindole [6], 2-methyl-5-methoxyindole [6], and 4 are given in Figure 5.
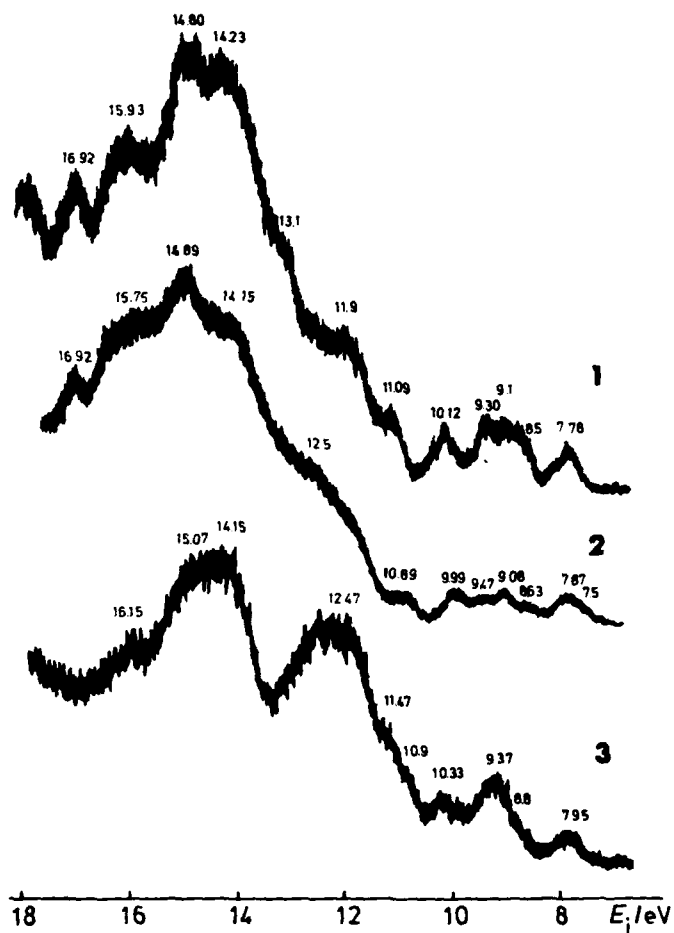
Figure 1.   HeI photoelectron spectra of (top to bottom): antipyrine (1), aminopyrine (2), and phenylbutazone (3).
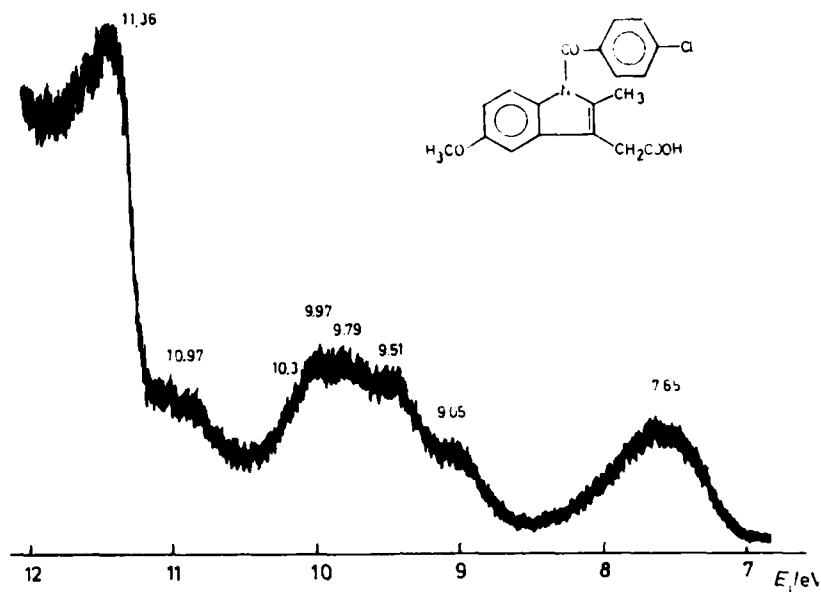
Figure 2.   HeI photoelectron spectrum of indomethacin (4).
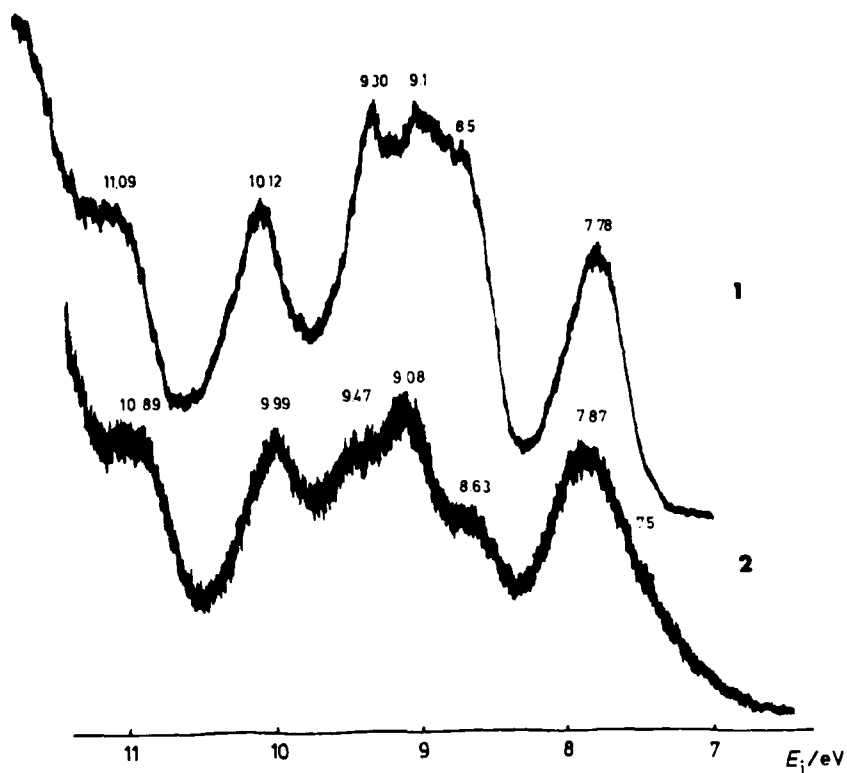
Figure 3.   Expanded low energy part of the HeI photoelectron spectra of antipyrine (1) and aminopyrine (2).

Discussion

Besides the previous work on pyrazolones [5] and indoles [6], the substituted hydrazine studies by Kimura and co-workers [12], Rademacher and co-workers [13], and Nelsen and Buschek [14] are of great relevance for the present analysis. These authors found a strong dependence of the two lone-pair splittings on dihedral angle (i.e., on the chemical conformation). These findings, later confirmed by nuclear magnetic resonance (NMR) [14] and quantum chemical calculations [12, 13], led to deduction of the gas phase conformations. A comprehensive overview of n-n conjugation in hydrazines and hydrazobenzenes has been given by Brown and Jorgensen [15].

A conjunction of our own results [1] for aniline, acetanilide, hydrazobenzene, and 1,2-diphenylpyrazoline with those of Rademacher, Bass and Wildemann [15] leads immediately to an electronic structure for $\underline{3}$ (Fig. 4). The relevant orbitals are: (i) The antibonding (π_1) and bonding (π_3) molecular orbitals of aniline that arise from an interaction of the amine substituent with the b_1 component of the degenerate e_{1g} π orbital of bezene; (ii) The π_2 orbital, largely the a_2 component of e_{1g} which is unaffected by substitution at a nodal ring center; and (iii) The amidic carbonyl lone-pair orbitals, n_0. The π_2 and n_0 orbitals remain essentially unaltered from one compound to another. However, the addition of two anilines to form hydrazobenzene splits each of the π_1 and π_3 orbitals into two components, in accord with the degree of n_N^+ and n_N^- contributions contained in the resultant orbitals. Furthermore, this splitting is very sensitive to conformation. Finally, given that the environment of the 5-membered ring is much

Figure 4. Correlation diagram for the observed vertical ionization energies, $E_{i,v}$/eV, of aniline [16], acetanilide [1], phenylbutazone ($\underline{3}$), 1,2-diphenylpyrazoline [13], and hydrazobenzene [13].

the same in 1,2-diphenylpyrazoline and in $\underline{3}$, we can assume the electronic structure to be very similar in both cases. A synthesis of these attitudes, together with the fact that the introduction of a $-CO.R$ group to an aromatic amine generally causes an increase of ionization energies and produces a new n_0 ionization at about 9.5 eV, leads to the following assignments for $\underline{3}$: E_i/eV: $7.95(\pi_1^+) < {\sim}8.8(\pi_1^-) < {\sim}9.3(2\pi_2)$ $< 9.5(2n_0) < 10.33(\pi_3^-) < 10.9(\pi_3^+)$.

Assignments for $\underline{1}$ and $\underline{2}$ do not come readily for a number of reasons: First, very little PE data exist for mixed hydrazines that contain one aromatic and one aliphatic substituent. This, of course, complicates all extrapolary arguments. Second, several conformations of the same molecule can exist in an equilibrium mixture. For example, our own studies [3] of the tautomerism of the CH-, OH-, and NH- forms of pyrazolones in the gas phase suggest a preference for the CH form when all other forms can coexist and that, in general, the NH form is the least stable. Compounds $\underline{1}$ and $\underline{2}$ are forced into an NH form as a result of appropriate substitution. From the expanded low energy parts of the PE spectra of $\underline{1}$ and $\underline{2}$, tentatively, we may assign:

for $\underline{1}$: E_i/eV: 7.78 (π_1^+); 8.5 (π_1^-); 9.1 (π_2); 9.3 (n_0); 10.12 and 11.09 $(\pi_3$ and $\pi_{CC})$
for $\underline{2}$: E_i/eV: 7.5 (n_{NMe_2}); 7.87 (π_1^+); 8.63 (π_1^-); 9.08 (π_2); 9.47 (n_0); 9.99 and 10.89
              $(\pi_3$ and $\pi_{CC})$

It would be remiss not to point out the deficiencies in this assignment for compound $\underline{2}$. For example, the n_{NMe_2} ionization is expected at 8.5 eV. Its appearance at 7.5 eV seems to be a result of interaction of the dimethylamino group with the neighboring carbonyl. The corresponding n_0 ionization becomes broad in shape and is shifted to higher energy. On the other hand, an assignment of the 7.5 and 7.87 eV events to π_7^+ and π_7^-, respectively, would because of the negligible splitting, indicate a completely different conformation of $\underline{2}$ versus $\underline{1}$ and $\underline{3}$. Obviously, additional work is mandated.

The assignment of the PE spectrum of $\underline{4}$ follows from that of indole, 2-methylindole, 2-methyl-5-methoxy-indole, indole-3-acetic acid, and the general effects of N-acylation on PE spectra (Fig. 5). The shape of relevant indole orbitals is shown in Figure 5: in indole-3-acetic acid and 2-methyl-indole, the ionization energies decrease relative to the parent indole; and an additional ionization event, one associable with the lone pair of the carbonyl group (i.e., $I(n_{OAc})$), should appear at ${\sim}10.3$ eV in the acid. Furthermore, methoxy substitution of the N-methyl derivative in the 5-position should destabilize π_1 slightly, π_2 and π_3 not at all (because of a node at the

CH-form   OH-form   NH-form

Figure 5. Correlation diagram for the observed vertical ionization energies, $E_{i,v}/eV$, of indole-3-acetic acid [6], indole [6], 2-methylindole [6], 2-methyl-5-methoxyindole [6], and indomethacin (4).
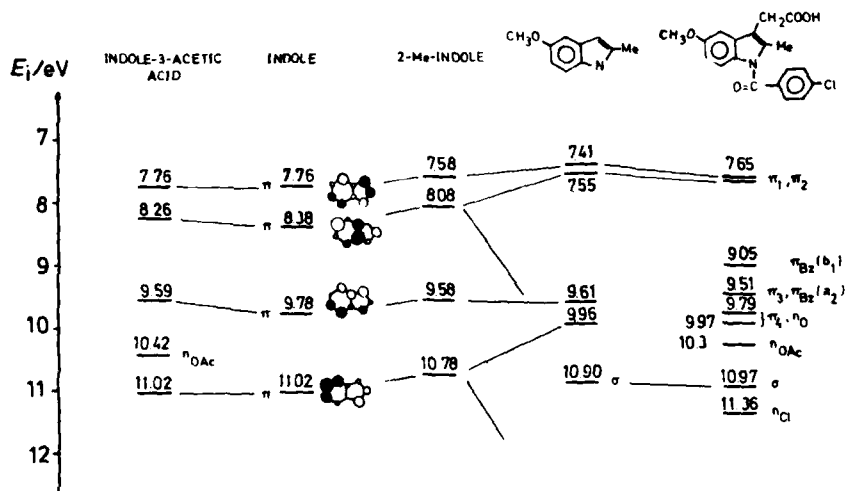
point of substitution in π_3). Finally, the fact that benzoylation of the nitrogen, with a p-chlorobenzoyl entity, is known [1] to increase the ionization energies, leads to a facile assignment: the three ionization events that lie below 10 eV (two associable with the benzene ring, namely π_{Bz} (b_1) and π_{Bz} (a_2) at 9.04 and ~9.6 eV, respectively; and one associable with the CO group of benzoyl, namely $I(n_0)$, are immediately assignable. The other $I(n_{OAc})$ event is observed as a shoulder at ~10.3 eV, and, of course, the $I(n_{Cl})$ events occur at 11.4 eV.

Summary

The application of a composite molecule method to an analysis of the PE spectra of some rather complex drugs has been shown to yield some definitive insights into the electronic structure of those drugs, despite the fact some important intermediaries (i.e., component molecules) had neither been studied nor analyzed. In part, this success is attributable to the fact that all four drugs possess a common molecular fragment, 5, and that this fragment is mainly responsible for the comparatively low observed ionization potentials (<8 eV).

In the absence of detailed knowledge of the specific steps in the drug mechanism of these compounds, it is therefore reasonable to suppose that all these drugs function as electron donors.

Acknowledgment

Bibliography

[1] L. Klasinc, I. Novak, " ‛ljić, and S. P. McGlynn, Int. J. Quantum Chem. **QBS13**, 251 (1986).

[2] S. P. McGlynn, L. Vanquickenborne, D. Carrol, and M. Kinoshita, *Introduction to Applied Quantum Chemistry* (Holt, Rinehart and Winston, New York, 1971).

[3] T. Meeks, A. Wahlborg, and S. P. McGlynn, J. Electron Spectrosc. **22**, 43 (1981).

[4] D. M. Woodbury and E. Finge, in *The Pharmacological Basis of Therapeutics*, L. S. Goodman and A. Gilman, Eds. (Macmillan, New York, 1975), 35th Edition pp. 325–358.

[5] G. Kluge, G. Kania, F. Achenbach, H. Wilde, I. Novak, and L. Klasinc, Int. J. Quantum Chem. **QBS11**, 237 (1984).

[6] H. Güsten, L. Klasinc, J. V. Knop, and N. Trinajstić, in *Excited States of Biological Molecules*, J. Birks, Ed. (Wiley, Chichester, 1975), p. 45; H. Güsten, L. Klasinc, and B. Ruščić, Z. Naturforsch., **31a**, 1051 (1976).

[7] S. H. Snyder and C. R. Merrie, Proc. Natl. Acad. Sci. U.S.A. **54**, 258 (1965); see also, *Molecular Orbital Studies in Chemical Pharmacology*, L. B. Kier, Ed. (Springer, W. Berlin, 1970).

[8] L. N. Domelsmith, L. L. Munchausen, and K. N. Houk, J. Am. Chem. Soc. **99**, 4311. 6506 (1977); L. N. Domelsmith and K. N. Houk, Int. J. Quantum Chem. **QBS5**, 257 (1978); L. N. Domelsmith, L. L. Munchausen, and K. N. Houk. J. Med. Chem. **20**, 1346 (1977); L. N. Domelsmith, T. A. Eaton, K. N. Houk, G. M. Anderson III, R. A. Glenonn, A. T. Shulgin, N. Castagnoli, and P. A. Kollman, J. Med. Chem. **24**. 1414 (1981).

[9] V. Butković, B. Kovač, I. Novak, B. Ruščić, A. Sabljić, L. Klasinc, and S. P. McGlynn, in *Modelling of Structure and Properties of Molecules*, Z. B. Maksić, Ed., (Ellis Horwood Ltd., Chichester), in press; L. Klasinc, B. Kovač, E. Polla, and S. Mutak, Acta Pharm. Jugosl. **37**, 67 (1987) (part 13 of the series).

[10] *Symposium on "Anti-Rheumatic Drugs,"* E. C. Huskisson, Ed. (Praeger Publishers, New York, 1983).

[11] L. Klasinc, B. Kovač, and B. Ruščić, Kem. Ind. (Zagreb) **23**, 569 (1974).

[12] K. Kimura and K. Osafune, Bull. Chem. Soc. Japan, **48**, 2421 (1975); K. Kimura, S. Katsumata, and K. Osafune, Bull. Chem. Soc. Japan, **48**, 2736 (1975).

[13] P. Rademacher, V. M. Bass, and M. Wildemann, Chem. Ber. **110**, 1939 (1977).

[14] S. F. Nelson and J. M. Buschek, J. Am. Chem. Soc. **95**, 2011, 2013 (1973).

[15] R. S. Brown and F. S. Jorgensen, in *Electron Spectroscopy: Theory, Techniques and Applications*, Vol. 5, C. R. Brundle and A. D. Baker, Eds. (Academic Press, London, 1984) pp. 2–120.

[16] L. Klasinc, B. Kovač, and H. Güsten, Pure Appl. Chem. **55**, 289 (1983).

On the Use of the Weighted Identification Numbers in the QSAR Study of the Toxicity of Aliphatic Ethers

B. BOGDANOV,* S. NIKOLIĆ, A. SABLJIĆ, AND N. TRINAJSTIĆ

The Rugjer Bošković Institute, P.O.B. 1016, 41001 Zagreb, Croatia, Yugoslavia

S. CARTER

Department of Chemistry, The University of Reading, Reading RG6 2AD, England, United Kingdom

Abstract

We have examined an application of the weighted identification number in the QSAR study of the toxicity of aliphatic ethers on mice. The results obtained are superior to those achieved by the connectivity index.

Introduction

Recently a novel graph-theoretical index, known as the weighted identification (WID) number, has been introduced [1], which appears to be a highly selective structural descriptor. It is also worth noting that the WID number can be computed straightforwardly for any structure [1, 2].

We decided to test the applicability of the WID number in quantitative structure-activity relationships (QSAR) studies. The toxicity of aliphatic ethers was selected for this purpose because the problem has already been treated with the connectivity index of Randić [3] with some success [4]. The connectivity index is so far the most successful graph-theoretical descriptor used in QSAR work [5, 6] and thus we will be able to investigate how the WID number performs in comparison with the connectivity index on the same sample.

People have been interested in the anesthetic activity of aliphatic ethers since the discovery of diethyl ether in 1542 [7]. Interest has been particularly focused on the toxicities of ethers [8]. We will consider the correlation between the WID number and the toxicities (pC) of a set of 21 aliphatic ethers on mice [4, 8] in an attempt to produce a QSAR model of predictive power. We will also carry out calculations with the connectivity index using exactly the same types of regression analyses as those employed for the WID number.

---

* Permanent address: Department of Chemistry, University of Skopje, S' opje, Macedonia, Yugoslavia.

The WID Number

In presenting a brief derivation of the WID number we will use graph-theoretical language for convenience [9, 10]. The incentive for the development of the WID number was Randić's molecular identification (ID) number [11] and its successful application in QSAR studies [12, 13]. However, isomeric trees were found with the same ID number [1]. This fact stimulated us to look for a number with much greater selectivity than Randić's ID number, and if such a number could be found, to investigate whether it could be used in QSAR studies. Let $G = (V, E)$ be a graph with the vertex-set $V = V(G)$ and the edge-set $E = E(G)$. Let $V = (v_1, v_2, \ldots, v_N)$ be a labelling of V. The distance between the vertices v_i and v_j is denoted by $d(i, j)$. Note that $d(i, j) = d(j, i)$, and $d(i, i) = 0$. Distances $d(i, j)$ are elements of the distance matrix of G, $\mathbf{D} = \mathbf{D}(G)$ [14–16]. The distance-sum D_i in $\mathbf{D}$ is defined by [17–19]:

$$D_i = \sum_{i=1}^{N} d(i, j); \qquad 1 \le i \le N \tag{1}$$

The distance sum has been used by Seybold [20] as a measure of the compactness or centrality of a particular site in a molecule. The distance sums may be easily obtained with any of several available computer-oriented algorithms for constructing the distance matrix for any structure [21, 22] and they simply represent the sums of the elements in the rows (or columns) of the distance matrix. The weights of edges w_{ij} in G are defined as [3, 18]:

$$w_{ij} = \begin{cases} (D_i D_j)^{-1/2} & \text{if } d(i, j) = 1; 1 \le i \le N, 1 \le j \le N \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

They represent the elements of the matrix of weights, $\mathbf{W} = \mathbf{W}(G)$. Let $w = (v_{i_1}, v_{i_2}, \ldots, v_{i_N})$ be a walk [9] of length k. The weight of this walk is defined by:

$$\prod_{j=1}^{k} (D_{i_j} D_{i_{j+1}})^{-1/2} = \prod_{j=1}^{k} w_{i_j i_{j+1}} \tag{3}$$

The weight of all walks of length k between vertices v_{i_1} and $v_{i_{k+1}}$ is given by:

$$\sum_{i} \prod_{j=1}^{k} w_{i_j i_{j+1}} \tag{4}$$

The entry $w_{ij}^{(k)}$ of $\mathbf{W}^k$ is the sum of all weighted walks of length k. The WID number of G is then defined as follows:

$$\text{WID}(G) = N - (1/N) + (1/N)^2 \cdot \text{ID}^*(G) \tag{5}$$

where:

$$\text{ID}^*(G) = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}^* \tag{6}$$

Note that:

$$(w_{ij}^*)_{1 \le i, j \le N} = W^* \tag{7}$$

and:

$$W^* = \sum_{k=0}^{N-1} W^k. \tag{8}$$

Note that because of the way in which the WID number is constructed, its limits are:

$$N \leq \text{WID}(G) \leq N + 1 \tag{9}$$

for each graph G with N vertices.

We have devised a computer program for calculating the WID numbers which starts with the distance matrix and proceeds via D_i, W, W^k, ID* and finally ends with the WID number [23, 24].

Results and Discussion

The toxicity of 21 aliphatic ethers together with their WID numbers and connectivity indices are given in Table I. The WID numbers are calculated as shown in the previous section. The connectivity index χ is calculated by the following formula [3]:

$$\chi = \sum_{\text{bonds}} (m_i \cdot n_j)^{-1/2} \tag{10}$$

where m_i and n_j are the valencies of the endpoints of the bond i-j.

TABLE I. Toxicities of aliphatic ethers ($R_1 - 0 - R_2$) on mice pC and the corresponding WID numbers and connectivity indices.

Ether	pC[a]	WID	χ
Dimethyl	1.43	3.29255	1.414
Methyl ethyl	1.74	4.12444	1.914
Methyl propyl	2.45	5.05815	2.414
Methyl isopropyl	2.26	5.07386	2.270
Methyl butyl	2.70	6.03157	2.914
Methyl isobutyl	2.79	6.03985	2.770
Methyl secbutyl	2.79	6.03723	2.808
Methyl terbutyl	2.79	6.05011	2.561
Methyl pentyl	2.88	7.01917	3.414
Diethyl	2.22	5.05815	2.414
Ethyl propyl	2.60	6.03157	2.914
Ethyl isopropyl	2.60	6.03723	2.700
Ethyl butyl	2.82	7.01917	3.414
Ethyl isobutyl	2.82	7.02315	3.270
Ethyl secbutyl	2.85	7.02154	3.308
Ethyl terbutyl	2.92	7.02712	3.061
Ethyl pentyl	3.00	8.01257	3.914
Ethyl terpentyl	3.15	8.01811	3.621
Dipropyl	2.79	7.01917	3.414
Propyl isopropyl	2.82	7.02154	3.270
Di-isopropyl	2.82	7.02449	3.126

[a] Refs. 4 and 8.

We examined two types of correlations: (a) linear least-squares fit

$$pC = a + b \cdot I \tag{11}$$

and (b) quadratic least-squares fit

$$pC = a + b \cdot I + c \cdot I^2 \tag{12}$$

where I = WID or χ. The results of the above regression analyses are given in Tables II and III.

From the statistical data in Tables II and III, we see that the WID number is superior to the connectivity index. It appears, both from earlier comparative studies [25], and from this work, that of all single graph-theoretical indices used for the correlation with the toxicities of aliphatic ethers, the most promising QSAR model is achieved with the WID number. Even when a polyparametric regression equation with several kinds of graph-theoretical indices is employed, the quality of the QSAR model with only the WID number is unsurpassed.

One possible reason for this good performance of the WID number is discussed here. If we carry out the regression analyses by using only the number of atoms in the ether, the following statistical equations are obtained for the two cases considered above:

$$pC = 1.002 + 0.314 \cdot N; \quad r = 0.947 \quad s = 0.133$$

$$F^{1,19} = 166 \quad r^2 \text{ (adjusted)} = 0.892 \tag{13}$$

TABLE II.   Statistical characteristics of a linear relationship between the aliphatic ether toxicities on mice and the WID numbers and connectivity indices.

| I | Statistical data | | | | | |
	a	b	r	s	$F^{1,19}$	r^2(adjusted)
WID	0.602	0.325	0.942	0.139	149	0.881
χ	0.792	0.633	0.909	0.172	91	0.818

TABLE III.   Statistical characteristics of a quadratic relationship between the aliphatic ether toxicities and the WID numbers and connectivity indices.

| I | Statistical data | | | | | | |
	a	b	c	r	s	$F^{2,18}$	r^2(adjusted)
WID	−1.321	1.020	−0.060	0.976	0.090	181	0.947
χ	−1.019	2.046	−0.261	0.955	0.123	93	0.902

$$pC = 0.027 + 0.767 \cdot N - 0.048 \cdot N^2; r = 0.975 \qquad s = 0.092$$

$$F^{2, 18} = 174 \quad r^2 \text{ (adjusted)} = 0.945 \tag{14}$$

These results are rather nice and one may be tempted to recommend the number of atoms to be used in many QSAR models. However, the number of atoms is a descriptor of low discriminatory power, it cannot, of course, differentiate isomeric molecules.

From the structure of the formula for the WID number [see Eq.(5)], we see that this number is rather closely related to the number of atoms in a molecule. Hence, the WID number could be simply presented as:

$$\text{WID} = N + \text{corr.} \tag{15}$$

where:

$$\text{corr.} = -(1/N) + (1/N)^2 \cdot \text{ID}^* \, (G) \tag{16}$$

The correction (16) for a great number of chemical graphs is rather small. It will increase with the increasing complexity of a graph. The superiority of the WID over N is clear in the case of isomers. For example, all 366,319 isomers of $C_{20}H_{42}$ alkane are differentiated by the WID number while they all have the same $N = 20$. This sensitivity of the WID number is related to the small correction given in expression (16).

In the past, very discriminative graph-theoretical indices have been found not to be particularly useful in QSAR studies. The large amount of structural information contained in such graph-theoretical indices may obscure those factors that are significant for a particular property that is to be modelled via QSAR technology. A good example to illustrate this point is Balaban's index [18], which is a highly discriminative descriptor which has so far shown little use in QSAR studies [26].

Concluding Remarks

We wish to point out that the WID has many good features for applications to QSAR studies. It is the most discriminative graph-theoretical index that has been found to date. The WID is designed to avoid large structural information which may obliterate its use in constructing QSAR models. Therefore, the WID number shows potential for use in QSAR work. However, more work is needed before the range of its applicability is established. Some research in this direction is already in progress [27].

Acknowledgment

Bibliography

[1] K. Szymanski, W. R. Müller, J. V. Knop, and N. Trinajstić, Int. J. Quantum Chem.: Quantum Chem. Symp. 20, 173 (1986).

[2] N. Trinajstić, S. Nikolić, and D. Horvat, Kem. Ind. (Zagreb), in press.

[3] M. Randić, J. Am. Chem. Soc. **97**, 6609 (1975).

[4] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic, New York, 1976), p. 202.

[5] D. H. Rouvray, Sci. Am. **254**, 40 (1986).

[6] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analyses* (Research Studies Press, Letchworth, 1986).

[7] V. Cordus, Pharmacorum Conficiendorum Ratio, Vulgo Vocant Dispensatorium, Norimbergae, 1546.

[8] D. P. Marsh and C. D. Leake, Anaesthesiology **11**, 455 (1950).

[9] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1972), second printing.

[10] N. Trinajstić, Chemical Graph Theory (CRC Press, Boca Raton, FL, 1983).

[11] M. Randić, J. Chem. Inf. Comput. Sci. **24**, 164 (1984).

[12] M. Randić, Int. J. Quantum Chem.: Quantum Biol. Symp. **11**, 137 (1984).

[13] M. Randić, in *Molecular Basis of Cancer, Part A: Macromolecular Structure, Carcinogens and Oncogens*, R. Rein, Ed. (Liss, New York, 1985), p. 309.

[14] D. H. Rouvray, in *Chemical Applications of Graph Theory* (Academic, London, 1976), p. 175.

[15] Ref. 10, p. 44.

[16] D. H. Rouvray, in *Mathematics and Computational Concepts in Chemistry*, N. Trinajstić, Ed. (Horwood, Chichester, 1986), p. 295.

[17] M. Randić, Math. Chem. (Mülheim/Ruhr) **7**, 5 (1979).

[18] A. T. Balaban, Chem. Phys. Lett. **89**, 399 (1982).

[19] N. Trinajstić, M. Randić, and D. J. Klein, Acta Pharm. Jugosl. **36**, 267 (1986).

[20] P. G. Seybold, Int. J. Quantum Chem.: Quantum Biol. Symp. **10**, 95 (1983); ibid. **10**, 103 (1983).

[21] M. Bersohn, J. Comput. Chem. **4**, 110 (1983).

[22] W. R. Müller, K. Szymanski, J. V. Knop and N. Trinajstić, J. Comput. Chem. **8**, 170 (1987).

[23] S. Carter, N. Trinajstić and S. Nikolić, Acta Pharm. Jugosl., **37**, 37 (1987).

[24] S. Carter, S. Nikolić, and N. Trinajstić, in preparation.

[25] O. Mekenyan, D. Bonchev, A. Sabljić, and N. Trinajstić, Acta Pharm. Jugosl., **37**, 75 (1987); M. Randić, A. Sabljić, S. Nikolić and N. Trinajstić, Quant. Struct.-Act. Rel., submitted.

[26] A. Sabljić, J. Chromatogr. **314**, 1 (1984); ibid. **319**, 1 (1985).

[27] S. Carter, S. Nikolić, and N. Trinajstić, work in progress.

Human Images by Nuclear Magnetic Resonance

E. RAYMOND ANDREW

Departments of Physics, Radiology and Nuclear Engineering Sciences,
University of Florida, Gainesville, Florida 32611, U.S.A.

Abstract

Proton nuclear magnetic resonance (NMR) signals may be obtained from the human body in such a way as to produce images of anatomical slices. The NMR signal is mapped in the selected slice by application of field gradients which provide spatial encoding. Excellent NMR images are now obtained which are useful in clinical practice. Introduction of NMR imaging into major hospitals is proceeding rapidly, with 10 commercial companies supplying the market; over 500 whole body systems have been installed worldwide. As a method of medical imaging NMR has the advantage that it uses no ionizing radiation and is therefore inherently safe; moreover, it gives sections in transverse, coronal, and sagittal planes with equal ease, and has good tissue contrast and pathological contrast arising from relaxation time differences. Contrast may be improved by use of contrast agents. Current trends suggest that before very long, whole-body NMR systems will be found in all major hospitals.

It is of great importance to the physician to be able to view the inside of the human body in order to effectively deal with diseases by medical or surgical treatment. One of the best ways to accomplish this is by use of x-rays, and now, with the development of computed tomography (CT) x-ray scanning, extremely good images can be obtained. However, x-rays do have a serious disadvantage: they are an ionizing radiation and therefore they can do us some harm in the course of doing us some good.

On the other hand, obtaining human images by magnetic resonance (MRI) has the great advantage of not using ionizing radiation and is therefore inherently safer. Moreover, in addition to being nonhazardous, magnetic resonance images offer improved diagnostic information through tissue discrimination and pathological discrimination arising from differences in relaxation time. Furthermore, one can obtain images in any desired plane: transverse, sagittal, or coronal with equal ease. With CT x-ray scanners, one can obtain direct images only in the transverse plane.

When obtaining human images by nuclear magnetic resonance (NMR), we place the body in a magnetic field and obtain an image of the distribution of the hydrogen nuclei in a slice of the body. Therefore, it is necessary to first define the slice to be imaged. This is done by the selective excitation method [1, 2]. The body is placed in a magnetic field with a field gradient along the Z direction. The body is then irradiated with a 90° resonant NMR pulse of narrow spectral width so that nuclei in a narrow range of Z values only are excited. In this way a slice in the body is defined and only the protons in this slice are excited.

The field gradient is then immediately switched into the plane of the slice, along say the X direction, and the NMR-free induction decay evolves during a time t_x. The

NMR signal from each volume element depends on the proton density there and its frequency depends on its x coordinate. After time t_x the gradient is suddenly switched along the Y direction and the NMR signal continues to evolve during the subsequent time t_y. The NMR signal from each volume element depends now not only on its x coordinate (which determined its phase at time t_x) but also on its y coordinate which determines its frequency in t_y. Every volume element in the defined slice is thus labelled or encoded according to its x and y coordinates. The NMR-free induction decay is recorded during t_y to n data points.

The procedure is now repeated for n different values of t_x, and an array of $n \times n$ data values is recorded. To this array of data, a two-dimensional Fourier transform is applied to obtain an image with $n \times n$ picture elements (pixels). This is the 2D Fourier imaging method [3] which is most commonly used in clinical magnetic resonance imaging at present. Further details of the procedures are given in Ref. 4. Instead of applying a fixed gradient during an incremented time t_x, it is often more convenient to apply an incremented gradient during a fixed time t_x: this variation is called the spin-warp method [5]. Nowadays, n is typically 256 and allows a resolution rather less than a millimeter in the head and rather more than this in the body. Needless to say, the instrument must have its own dedicated computer which instructs the system, collects and stores the data, and performs the calculations.

Figure 1 shows a proton magnetic resonance image of a thin transverse slice through the author's head obtained on the 0.15 Tesla NMR scanner which has been in operation in the University of Florida hospital for over three years. This image shows the eyes, nose, scalp, the marrow in the skull, and the two hemispheres of the brain. This image is one of a set of 10 parallel slices obtained simultaneously by the multi-slice technique [4] in about 5 minutes.

As mentioned earlier, NMR images may be obtained readily in any orientation, and Figure 2 shows a sagittal image of the author's head, also from a set of 10 coplanar images obtained simultaneously. The image in Figure 2 is a midline section showing the scalp, the marrow of the skull, the corrugations of the cortex, the spinal cord, and the cerebellum.

Figure 3 shows one of a set of 10 sagittal images through the author's chest, showing the vertebrae and the spinal cord. Images of the heart are blurred by its motion, but this motional artifact can be removed by synchronizing the successive 90° pulses with impulses from an electrocardiograph probe. Figure 4 shows one of a set of 10 transverse images through the author's abdomen showing the liver, spine in section, aorta, intersections with the rib cage, and the two arms on either side of the body.

These four figures are examples of the quality of magnetic resonance images from what is hopefully a normal human body using a low-field, first-generation instrument. In July 1987 we expect to take delivery of a 1.5 Tesla superconducting NMR imaging system in the University of Florida hospital which should provide images with improved resolution and diagnostic capability.

It will be noticed that the images shown have all been obtained from the NMR response of protons in the body. In fact, almost all clinical imaging is done with the hydrogen nuclei. Hydrogen is the most abundant chemical element in the body. perience of an MRI examination, and magnetic resonance will then become part of the everyday language of the man-in-the-street.

Figure 1. Proton NMR image of a thin transverse slice through the author's head.

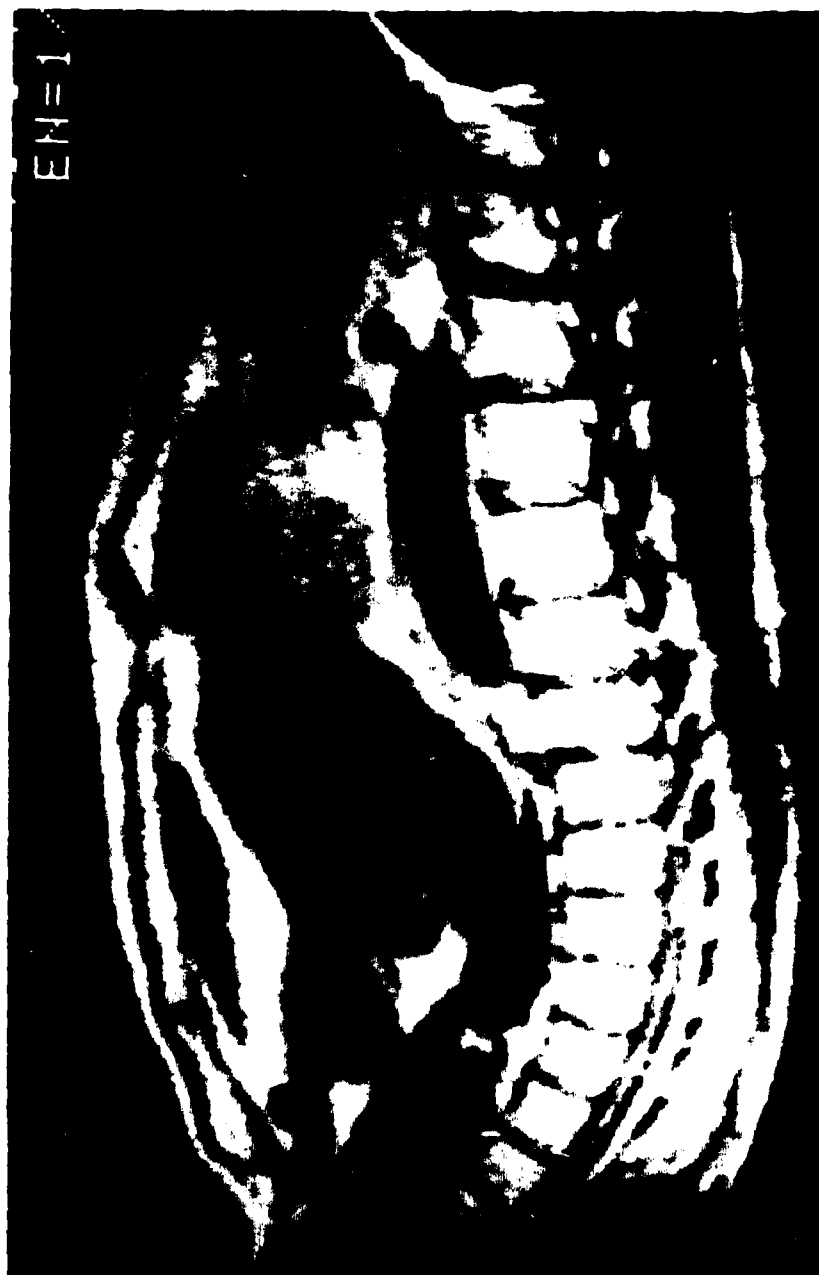Figure 2. Proton NMR image of a midline sagittal section through the author's head.

Figure 3. Proton NMR image of a sagittal section through the author's chest.

Figure 4. Proton NMR image of a transverse section through the author's abdomen.

Moreover, the proton is isotopically almost 100% abundant and has the best magnetic properties of all stable nuclei. Oxygen has no suitable isotope. For carbon only the isotope ^{13}C is possible and it is only 1% abundant. Phosphorus is of considerable interest because of its role in metabolism and ^{31}P is a good NMR nucleus; however, its concentration in the body is low, leading to a spatial resolution an order of magnitude worse than hydrogen. Consequently, protons are by far the favorite nucleus for magnetic resonance imaging. Some imaging work has been done with ^{7}Li, ^{13}C, ^{19}F, ^{23}Na, ^{31}P, and also with unpaired electrons, but this is all at the research level and not in regular clinical practice.

The use of magnetic resonance imaging in hospitals depends on the extent to which the physicians find it useful in their clinical work. One example in our University hospital is provided by the pioneering work of Dr. W. F. Enneking on musculoskeletal tumors. In the past, treatment usually consisted of amputation of the limb, but in recent years the excision of the tumor is a much more common practice. However, this demands very precise delineation of the tumor. The longer NMR relaxation times exhibited by tumors [6] enable the tumors to be demonstrated with superior contrast to x-ray images. Some 350 cases of tumors of the musculoskeletal system have now been examined with our magnetic resonance imaging (MRI) system over the past three years and NMR images have played a major role in limb salvage surgery in these cases [7].

Sometimes lesions of interest do not show a significant difference of relaxation times T_1 and T_2 from normal tissue and are not well discriminated. In such cases, it can be useful to administer a contrast agent, which enhances the relaxation rate differentially. A popular agent is a solution of a gadolinium chelate DTPA. The gadolinium atom, a rare earth, has a strong electronic magnetic moment from its inner unpaired 4f electrons, while the chelate group grasps the outer electrons making the molecules unreactive and nontoxic. Moreover, the large chelate molecule is less able to penetrate the blood-brain barrier, but can penetrate tumors and other lesions, relaxing them strongly and giving them good constrast in a relaxation-weighted image.

In the United States, pregnant women are not normally examined by MRI, but in Britain national guidelines now allow this after the first trimester, which often proves useful in cases of abnormality. Figure 5 shows an example kindly provided by Professor Brian Worthington of Nottingham University of a 38-week-old fetus in a breech presentation in the womb. The baby was subsequently safely delivered by Caesarian section.

The subject of magnetic resonance imaging has made great strides since Lauterbur [8] published the first NMR images of two tubes of water in 1973. Now, in 1987, it can be said that MRI has become an accepted modality of medical imaging. Approximately a million patients have now been examined. Just as you will all no doubt have had many x-ray examinations, it will not be long before you will all have had the experience of an MRI examination, and magnetic resonance will then become part of the everyday language of the man-in-the-street.

Figure 5. Proton NMR image of a coronal section through the womb of a pregnant woman showing the fetus in breech presentation. This image was kindly provided by Professor Brian S. Worthington.

Acknowledgments

Bibliography

[1] A. N. Garroway, P. K. Grannell, and P. Mansfield, J. Phys. C 7, L457 (1974).

[2] P. C. Lauterbur, C. S. Dulcey, C. M. Lai, M. A. Feiler, W. V. House, D. Kramer, C. N. Chen, and R. Dias, Proc. 18th Ampere Congress, Nottingham 27 (1974).

[3] A. Kumar, D. Welti, and R. R. Ernst, J. Mag. Res. 18, 69 (1975).

[4] E. R. Andrew, Proc., Roy. Soc. B 255, 399 (1985).

[5] W. Edelstein, J. M. S. Hutchinson, G. Johnson, and T. Redpath, Phys. Med. Biol. 25, 751 (1980).

[6] R. Damadian, Science 171, 1151 (1971).

[7] E. R. Andrew, and H. T. A. Pettersson, in Magnetic Resonance in Cancer, P. S. Allen, D. P. J. Boisvert, and B. C. Lentle Eds. (Pergamon Press, New York, 1986).

[8] P. C. Lauterbur, Nature 262, 190 (1973).

Parameters and Mechanisms of Calcium Binding to Peptides and Proteins

KENZI HORI, JOSEPH N. KUSHICK*, ALAN FACTOR,
AND HAREL WEINSTEIN

Department of Physiology and Biophysics, Mount Sinai School of Medicine, City University of New York, New York, New York 10029

In order to understand the molecular mechanisms underlying the function of Ca^{2+} channels and Ca^{2+}-dependent modulatory proteins, we explore the characteristics of Ca^{2+} binding sites in molecules for which the structures are known at the atomic level. On such a structural basis, theoretical approaches formulated in the computational methods of quantum chemistry and molecular mechanics [e.g., see Refs. 1 and 2] can be used to study the structural and electronic factors that determine the mechanisms of Ca^{2+} binding and its consequences. One such approach is coded in the CHARMM package of computer simulation software, which can be used to perform both energy minimization and simulations of molecular dynamics for very large molecular systems such as the calcium binding systems [3].

To make theoretical analysis of the Ca-binding mechanism possible, a first step in the present study was the development of suitable parameters for the calculation of interactions of Ca^{2+} with peptides and proteins with the CHARMM programs. Quantum mechanical methods were used to calculate the energies of the complexes $[Ca(OH_2)_4]^{2+}$ and $[Ca(OCH_2)_4]^{2+}$, chosen as model systems for the interaction of Ca^{2+} with coordinating groups in peptides, at a variety of Ca-O distances. The parameters for the CHARMM program were then obtained by fitting the results to the analytical expression for "nonbonded interactions" in CHARMM [3].

The resulting parametrization was tested by calculations of the structure and of the Ca^{2+}-binding properties of two molecules that had been studied with experimental methods including x-ray crystallography: the hexapeptide cyclo-(Pro-Gly)$_3$ and its 2:1 complex with Ca^{2+} [4], and the 75-residue long intestinal calcium-binding protein (ICaBP) [5]. The geometrical parameters calculated for the hexapeptide (Fig. 1a) and its Ca^{2+}-containing complex (Fig. 1b, c) were found to be in good agreement with the experimental data from crystallography and from nuclear magnetic resonance (NMR) in nonpolar media [4], both with respect to the conformation of the peptides and the coordination of the calcium. Similarly, calculations of ICaBP with CHARMM using the new Ca^{2+} parameters yielded good agreement with data from x-ray crystallography (Fig. 2), when the appropriate constraints were applied. Briefly, the results show that minimization of the structure without the inclusion of
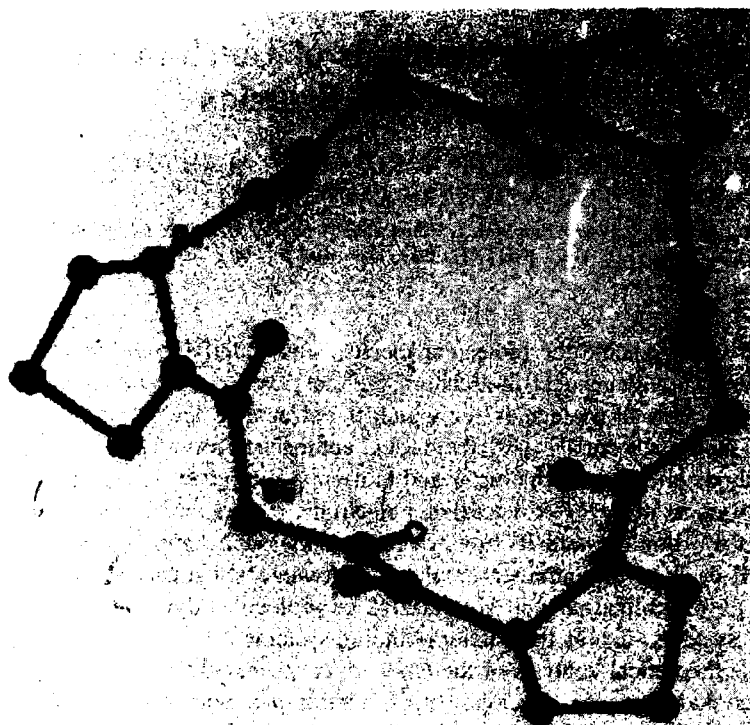

(a)

Figure 1. (a) Energy-optimized structure of cyclo-(Pro-Gly)₃. Note the hydrogen-bonding
arrangement of (Gly)N-H to O-C(Pro) which supports the C₃ symmetry of the structure ob-
served also in nonpolar media [4]. (b) Energy-optimized structure of the complex of Ca²⁺
with two molecules of cyclo-(Pro-Gly)₃ positioned above and below, viewed from above.
The residues are numbered 01-06 in one ring, and 07-12 in the second. Ca²⁺ is numbered
13, and the oxygens nearest to the calcium are identified by connecting lines. (c) Side view
of the energy-optimized structure of the complex shown in Fig. 1b. See legend to Fig-
ure 1b. for details.

the Ca²⁺ maintains the general tertiary structure of the ICaBP, which is composed of
two groups of helix-loop-helix arrangements linked by a "linker segment." However,
the calculations of the structure in the absence of calcium reveal significant changes
in the secondary structures of the two Ca-binding loops. The second loop (Loop II),
which is part of a classical "EF hand" defined by Kretzinger [6, 7], is somewhat less
affected than Loop I which does not seem to belong to an EF hand arrangement [7].
When energy minimization is performed with the Ca²⁺ in each of the two loops, the
Ca coordination pattern in the crystal is generally well reproduced and the structures
of the loops are restored, but the linker segment is distorted in comparison to the
crystal structure. To reproduce the geometry observed in the crystal it is necessary
to include in the calculations those water molecules that are observed in the crystal to
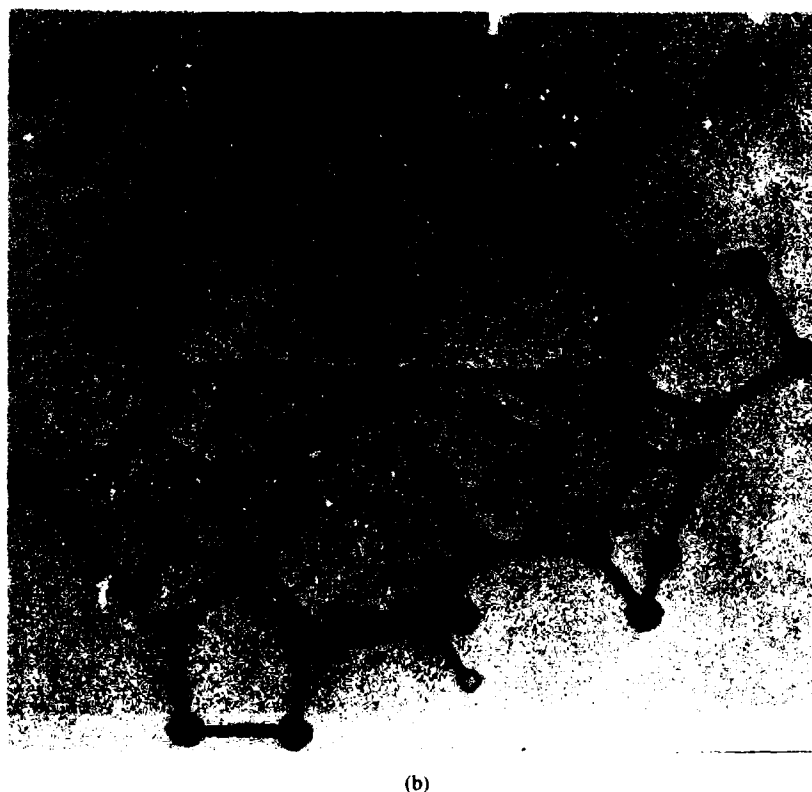
(b)

Figure 1. (Continued)

be nearest to atoms in the ICaBP, and thus likely to be tightly bound to the protein. Inclusion of the water molecules identified to have close contacts with atoms belonging to residues 42, 44, and 60 of ICaBP into the energy minimization calculation (see Fig. 2), restores the calculated structure to the observed geometry. Taken together, the results of energy minimization calculations of ICaBP emphasize the importance of the tertiary structure in maintaining the Ca-coordinating secondary structure that is required for calcium binding.

The conclusions regarding the role of the tertiary structure in the binding of Ca^{2+}, were probed by an analysis of the Ca^{2+} binding ability of some proteins predicted to bind the ion on the basis of sequence homology with calmodulin [8]. Using the "mutation" procedure available in CHARMM, the sequence of Loop II in ICaBP was consecutively mutated to that of each of the four loops identified in a yeast gene product shown to have a certain degree of sequence homology with the four loops of calmodulin [8]. The stabilization of Ca^{2+} by these loops was calculated from the energy differences between optimized structures of the Ca-containing loop within the rigid frame of ICaBP (which was kept fixed during the optimization of the loop structure) and the energy of the same structure optimized in a similar fashion in the
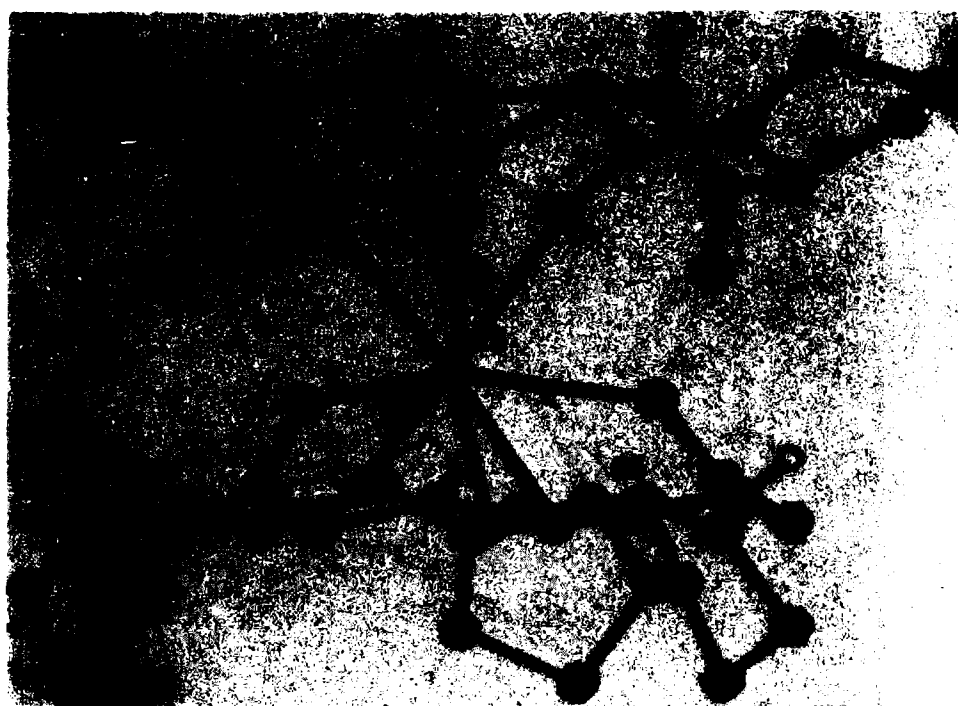
(c)

Figure 1. (Continued)

absence of Ca^{2+}. The stabilization of Ca^{2+} in two of the loops is found to be suffi-
cient to overcome the hydration energy of the ion, whereas for the other two loops
the difference between protein-binding energy and hydration is too small to permit a
definitive conclusion. For comparison purposes, mutations of Loop II into a series of
sequences constructed randomly were also analyzed for Ca^{2+}-binding abilities. The
stabilization of Ca^{2+} in the random loops is calculated to be insufficient to overcome
the favorable hydration energy of the ion; these loops are predicted not to bind Ca^{2+}.

Molecular dynamics simulations of the loop movements at high temperatures were
used to explore the restrictions in the conformational space of the loop surrounded by
the fixed protein structure of ICaBP. Results from these studies revealed the role of
sequence in the stabilization of Ca^{2+} in the binding loops, as well as the degree to
which the conformational space available to the loop is constrained by the rest of the
protein region that yields a structure prepared for Ca^{2+} binding.

Preliminary results from molecular dynamics simulations of the behavior of ICaBP
at very low and very high temperatures, support the conclusion that the tertiary struc-
ture of the helix-loop-helix arrangement in the EF hand of the Ca^{2+}-binding Loop II
restricts the conformational space of the loop and plays a major role in the binding of
the ion.

Figure 2. Energy-optimized structure of ICaBP. Only the alpha-carbon chain is shown. The sequence is numbered at every fifth residue. Lettering indicates the positions of the two Ca^{2+} (Ca) ions bound in the loops, and the three tightly bound waters (W1, W2, W3) that were included in the energy minimization calculations.

Bibliography

[1] H. Weinstein and J. P. Green, Quantum Chemistry in Biomedical Sciences. Ann. N.Y. Acad. Sci. 367 (1981).
[2] B. Venkataragharan and R. J. Feldmann, Macromolecular structure and specificity: computer-assisted modeling and applications. Ann. N.Y. Acad. Sci., 439 (1985).
[3] B. R. Brooks, R. R. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. 4, 187–217 (1983).
[4] G. Kartha, K. I. Varughese, and S. Aimoto, Proc. Natl. Acad. Sci. (USA) 79, 4519–4522 (1982).
[5] D. M. E. Szebeny, S. K. Obendorf, and K. Moffat. Nature 294:327–332 (1981).
[6] R. H. Kretsinger, 1980. CRC Crit. Rev. Biochem. 8:119–174 (1980).
[7] O. Hertzberg and M. N. G. James, Biochemistry 24:5298–5302 (1985).
[8] P. Baum, C. Furlong, and B. Byers, Proc. Natl. Acad. Sci. (USA) 83:5512–5516. (1986).

Pharmacological Activities in Thermal Proteins: Relationships in Molecular Evolution

SIDNEY W. FOX, FRANZ HEFTI, JUKKA HARTIKKA, EMMANUEL JUNARD, ALEXANDER T. PRZYBYLSKI, AND GRAHAM VAUGHAN

Institute for Molecular and Cellular Evolution and the Department of Neurology, University of Miami, Miami, Florida, U.S.A.

Introduction

The model of protobiological events that has been presented in these pages [1] has increasing relevance to pharmacological research. The thermal proteins* that function as key substances in the proteinoid theory have recently been found to prolong the survival of rat forebrain neurons in culture [2] and to stimulate the growth of neurites.

A search for such activity in thermal proteins added to cultures of modern neurons was suggested by the fact that some of the microspheres assembled from proteinoids rich in hydrophobic amino acids themselves generate fibrous outgrowths [3, 4].

Experimental

Thermal polycondensation of α-amino acids has been described [5], as has culturing of rat neurons to which the polymers were added [6].

Results

The results are presented in Table I. A comparison of these results with those from a set of thermal proteins for another biological activity is presented in Table II.

Discussion

The model of protobiological development was made possible by the finding that thermal polymerization of aspartic acid [7] could be extended to *co*polymerization. The laboratory conditions are geological in nature. This process yielded heteropolymers [8, 9] that could even include all α-amino acids, nearly all of which would however singly fail to polymerize or would decompose when heated.

The dicarboxylic amino acids play a special role in the polycondensation. Recent results by Luque-Romero et. al. [10] reconfirm the interpretation that aspartic acid fa-

---

*The term thermal protein was selected by Chemical Abstracts and has been used by them since 1972. Other terms for such polymers are proteinoids, thermal proteinoids, and thermal copolyamino acids.

TABLE I.   Thermal Proteins Active in Neuronal Survival in Culture

Test Material	Cells/Dish (Relative survival[a])
Controls	0
Poly(asp,glu,trp;1:1:1)	60,000 ± 5500
Poly(asp,glu,trp;1:1:2)	58,900 ± 6700
Tryptophan	0
Asp,glu,trp molar mixture	0
Poly(asp,glu, leu;1:1:1)	19,900 ± 1700
Poly(glu,trp;1:1)	0
Poly(trp)	0
Poly(asp,glu;1:1)	0

[a] 4 days from 200,000 cells

TABLE II.   Some Thermal Proteins Exhibiting Specific Biological Activity.

Polymer	Antiglyoxalase I	Neurotrophic
Glu,trp mixture(control)	0	0
Copoly(asp,glu,trp)	+	+
Copoly(asp,glu,leu)	0	+
Copoly(asp,glu,val)	0	+
Copoly(glu,trp)	+	0

cilitates thermal polymerization whereas glutamic acid is involved in the ordering process, starting as an N → C polymerization initiator [11]. The self-ordering is strong enough [12, 13] to yield thermal proteins limited in heterogeneity to a degree comparable to that in unfractionated proteins in modern organisms [14]. Limited heterogeneity of such polymers permits the practical aspect of repeatable biological function [15].

In discussion of the evolutionary position of thermal copolymerization of amino acids, Calvin [16] has pointed out that the proposal would be supported if evolutionary relics of the primordial self-ordering could be found. Working from the whole protein data base as well as parts of it, Ivanov and Förtsch have recently reported that in the majority of the cases they have studied modern proteins are found to contain such evolutionary relics [17].

As reported here, some of the thermal proteins rich in hydrophobic amino acids such as tryptophan or leucine, when added to cultures of rat forebrain neurons, prolong the survival. They also stimulate outgrowths of dendrites and axons [2]. The fact that model polymers for primordial proteins have such effects on modern neurons further supports the interpretation of an evolutionary sequence integrated by the common chemistry of precellular and cellular proteins [11] and by the common biofunctions [15, 18].

Pharmacological potentialities have been found earlier in thermal proteins [5, 19, 20]. Examination of these various functions reveal specificities for thermal proteins as a model for specificities in modern proteins [Table II].

Acknowledgments

Bibliography

[1] S. W. Fox, Intl. J. Quantum Chem. QBS8, 441–454 (1981); QBS11, 17–29 (1984).
[2] F. Hefti, J. Hartikka, A. Przybylski, G. Vaughan, and S. W. Fox, submitted for publication.
[3] L. L. Hsu, S. Brooke, and S. W. Fox, Curr. Mod. Biol. (BioSystems) 4, 12–25 (1971).
[4] S. W. Fox, A. T. Przybylski, T. Nakashima, and G. Vaughan, Intl. Symp. on Neural Regeneration, Abstracts P45, Asilomar CA, Dec 8–12 (1985).
[5] R. M. Syren, C. R. Windsor, and S. W. Fox, Intl. J. Quantum Chem. QBS6, 283–288 (1979).
[6] F. Hefti, J. Hartikka, F. Eckenstein, R. Heumann, and M. Schwab, Neuroscience 14, 55–62 (1985).
[7] H. Schiff, Chem. Ber. 30, 2449 (1897).
[8] K. Harada and S. W. Fox, Arch. Biochem. Biophys. 86, 274–280 (1960).
[9] S. W. Fox and K. Harada, J. Am. Chem. Soc. 82, 3745–3751 (1960).
[10] M. M. Luque-Romero, L. S. de Medina, and J. M. Blanco, BioSystems, 19, 267–272 (1986).
[11] S. W. Fox, Naturwissenschaften 67, 576–581 (1980).
[12] S. W. Fox and T. Nakashima, Biochim. Biophys. Acta 140, 155–167 (1967).
[13] T. Nakashima, J. R. Jungck, S. W. Fox, E. Lederer, and B. C. Das, Intl. J. Quantum Chem. QBS4, 65–72 (1977).
[14] S. W. Fox and T. Nakashima, in Individuality and Determinism, Chemical and Biological Bases, S. W. Fox, Ed. (Plenum Press, New York, 1984) pp. 185–201.
[15] S. W. Fox and C. R. Windsor, Intl. J. Quantum Chem. QBS11, 103–108 (1984).
[16] M. Calvin, Chemical Evolution (Oxford University Press, 1969) pp. 152–183.
[17] O. C. Ivanov and B. Förtsch, Origins Life 17, 35–49 (1986).
[18] S. W. Fox, Naturwissenschaften 67, 378–383 (1980); S. W. Fox and T. Nakashima, BioSystems 12, 155–166 (1980).
[19] S. W. Fox and C. -T. Wang, Science 160, 547–548 (1968).
[20] S. W. Fox, R. M. Syren, and C. R. Windsor, in Submolecular Biology and Cancer, G. Wolstenhome, Ed. Ciba Foundation Series (new 67, 175–193 (1979).

Author Index

Andre, J. M., 85
Andrew, E. R., 331
Arteca, G. A., 133

Bogdanov, B., 325

Carter, S., 325
Chantranupong, L., 75
Chen, K.-X., 15
Cieplak, P., 65
Collins, J., 75
Conrad, M., 167

Davies, R. H., 221
Del Bene, J. E., 27
Delhalle, J., 85
Dijkman, J. P., 211
Dory, M., 85

Factor, A., 341
Ford, G. P., 57
Fox, S. W., 1, 347
Fripiat, J. G., 85

Gresh, N., 15
Grossman, S. C., 245

Hariharan, P. C., 37, 111
Hartikka, J., 347
Hefti, F., 347
Hodgkin, E. E., 105
Hori, K., 341
Hutchens, T. W., 297

Jerman-Blažič, B., 245
Jhon, M. S., 9, 189
Junard, E., 347

Kaufman, J. J., 37, 111
Keegstra, P. B., 37
Klasinc, L., 317
Kollman, P. A., 65
Kong, Y. S., 189
Kovač, B., 317
Kurnig, I. J., 47
Kushick, J. N., 341

Ladik, J. J., 3
Leow, G., 75

Löwdin, P. O., v, 9, 189

McGlynn, S. P., 317
Mezey, P. G., 127, 133

Nikolić, S., 325

Öhrn, N. Y., v
Osman, R., 211

Porath, J., 297
Przybylski, A. T., 347
Pullman, B., 15

Randić, M., 245
Ravichandran, V., 289
Rein, R., 281
Richards, W. G., 105
Roszak, S., 37
Rouvray, D. H., 245
Roychoudhury, M., 281

Sabin, J. R., v
Sabljić, A., 317, 325
Scheiner, S. 47
Sekharudu, Y. C., 289
Seybold, P. G., 245
Shibata, M., 281
Singh, U. C., 65
Smith, C. T., 57
Sokalski, W. A. 111
Srinivasan, S., 281
Sundaralingam, M., 289

Trinajstić, N., 325

Urry, D. W., 261

Vasilescu, D., 149
Vaughan, G., 347
Viani, R., 149

Waleh, A., 75
Weinstein, H., 211, 341

Yathindra, N., 289

Zerner, M. C., v

Published Symposia of the
International Journal of Quantum Chemistry

1967   QUANTUM CHEMISTRY SYMPOSIUM NO. 1
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory)

1968   QUANTUM CHEMISTRY SYMPOSIUM NO. 2
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1969   QUANTUM CHEMISTRY SYMPOSIUM NO. 3 PART 1
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1970   QUANTUM CHEMISTRY SYMPOSIUM NO. 3 PART 2
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1971   QUANTUM CHEMISTRY SYMPOSIUM NO. 4
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory and Quantum Biology)

1971   QUANTUM CHEMISTRY SYMPOSIUM NO. 5
(Proceedings of the International Symposium on Atomic, Molecular and Solid-State Theory and Quantum Biology)

1972   QUANTUM CHEMISTRY SYMPOSIUM NO. 6
(Proceedings of the International Symposium on Atomic, Molecular and Solid-State Theory and Quantum Biology)

1973   QUANTUM CHEMISTRY SYMPOSIUM NO. 7
(Proceedings of the International Symposium on Atomic, Molecular and Solid-State Theory and Quantum Biology)

1974   QUANTUM CHEMISTRY SYMPOSIUM NO. 8
(Proceedings of the International Symposium on Atomic, Molecular and Solid-State Theory and Quantum Statistics)
QUANTUM BIOLOGY SYMPOSIUM NO. 1
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1975   QUANTUM CHEMISTRY SYMPOSIUM NO. 9
(Proceedings of the International Symposium on Atomic, Molecular and Solid-State Theory and Quantum Statistics)
QUANTUM BIOLOGY SYMPOSIUM NO. 2
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1976 QUANTUM CHEMISTRY SYMPOSIUM NO. 10
(Proceedings of the International Symposium on Atomic, Molecular and Solid-State Theory and Quantum Statistics)
QUANTUM BIOLOGY SYMPOSIUM NO. 3
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1977 QUANTUM CHEMISTRY SYMPOSIUM NO. 11
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory, Collision Phenomena, and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 4
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1978 QUANTUM CHEMISTRY SYMPOSIUM NO. 12
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory, Collision Phenomena, and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 5
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1979 QUANTUM CHEMISTRY SYMPOSIUM NO. 13
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory, Collision Phenomena, Quantum Statistics, and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 6
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1980 QUANTUM CHEMISTRY SYMPOSIUM NO. 14
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory, Collision Phenomena, Quantum Statistics, and Computational Methods).
QUANTUM BIOLOGY SYMPOSIUM NO. 7
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1981 QUANTUM CHEMISTRY SYMPOSIUM NO. 15
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-State Theory, Collision Phenomena, and Computational Quantum Chemistry)
QUANTUM BIOLOGY SYMPOSIUM NO. 8
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1982 QUANTUM CHEMISTRY SYMPOSIUM NO. 16
(Proceedings of the International Symposium on Quantum Chemistry, Theory of Condensed Matter, and Propagator Methods in the Quantum Theory of Matter)
QUANTUM BIOLOGY SYMPOSIUM NO. 9
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharmacology)

1983 QUANTUM CHEMISTRY SYMPOSIUM NO. 17
(Proceedings of the International Symposium on Atomic, Molecular and Solid-State Theory, Collision Phenomena, and Computational Quantum Chemistry)

QUANTUM BIOLOGY SYMPOSIUM NO. 10
(Proceedings of the International Symposium on Quantum Biology and Quantum
Pharmacology)

1984   QUANTUM CHEMISTRY SYMPOSIUM NO. 18
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-
State Theory, and Computational Quantum Chemistry)
QUANTUM BIOLOGY SYMPOSIUM NO. 11
(Proceedings of the International Symposium on Quantum Biology and Quantum
Pharmacology)

1985   QUANTUM CHEMISTRY SYMPOSIUM NO. 19
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-
State Theory, Scattering Problems, Many Body Phenomena, and Computational
Quantum Chemistry)
QUANTUM BIOLOGY SYMPOSIUM NO. 12
(Proceedings of the International Symposium on Quantum Biology and Quantum
Pharmacology)

1986   QUANTUM CHEMISTRY SYMPOSIUM NO. 20
(Proceedings of the International Symposium on Atomic, Molecular, and Solid-
State Theory, Scattering Problems, Many Body Phenomena, and Computational
Quantum Chemistry)
QUANTUM BIOLOGY SYMPOSIUM NO. 13
(Proceedings of the International Symposium on Quantum Biology and Quantum
Pharmacology)

1987   QUANTUM CHEMISTRY SYMPOSIUM NO. 21
(Proceedings of the International Symposium on Quantum Chemistry, Solid-State Theory,
and Computational Methods)
QUANTUM BIOLOGY SYMPOSIUM NO. 14
(Proceedings of the International Symposium on Quantum Biology and Quantum Pharma-
cology)

All of the above symposia can be individually purchased from the Subscription Department,
John Wiley & Sons.