Molecular Orbital Calculations for Biological Systems

Edited by Anne-Marie Sapse

Molecular Orbital Calculations for Biological Systems

Topics in Physical Chemistry
A Series of Advanced Textbooks and Monographs

Series Editor, Donald G. Truhlar

F. Iachello and R. E. Levine, *Algebraic Theory of Molecules*
P. Bernath, *Spectra of Atoms and Molecules*
J. Cioslowski, *Electronic Structure Calculations on Fullerenes and Their Derivatives*
E. R. Bernstein, *Chemical Reactions in Clusters*
J. Simmons and J. Nichols, *Quantum Mechanics in Chemistry*
G. A. Jeffrey, *An Introduction to Hydrogen Bonding*
S. Scheiner, *Hydrogen Bonding: A Quantum Chemical Perspective*
T. G. Dewey, *Fractals in Molecular Biophysics*
A.-M. Sapse, ed., *Molecular Orbital Calculations for Biological Systems*

Molecular Orbital Calculations for Biological Systems

EDITED BY

Anne-Marie Sapse

New York Oxford

Oxford University Press

1998

Oxford University Press

Oxford New York
Athens Auckland Bangkok Bogotá Buenos Aires
Calcutta Cape Town Chennai Dar es Salaam Delhi
Florence Hong Kong Istanbul Karachi Kuala Lumpur
Madrid Melbourne Mexico City Mumbai
Nairobi Paris São Paulo Singapore
Taipei Tokyo Toronto Warsaw

and associated companies in
Berlin Ibadan

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

Oxford is a registered trademark of Oxford University Press

Library of Congress Cataloging-in-Publication Data
Molecular orbital calculations for biological systems / edited by
Anne-Marie Sapse.
p. cm.—(Topics in physical chemistry)
Includes bibliographical references and index.
ISBN 0-19-509873-0
1. Molecular orbitals. 2. Biomolecules. 3. Peptides.
4. Amino acids. 5. Antineoplastic agents.
I. Sapse, Anne-Marie.
II. Series: Topics in physical chemistry series.
QP517.M66M65 1998
547'.70448—dc21 97-37834

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

This book is dedicated to my husband, Marcel Sapse,
and to my daughter, Danielle Sapse,
whom I thank for their support.
Anne-Marie Sapse

This page intentionally left blank

Preface

The applications of quantum chemical calculations to biological systems has been made possible by huge advances in computer facilities and the creation of better computer programs, capable of handling large systems. This book describes some of the quantum chemical methods used for such calculations, together with some widely used computer programs.

Chapter 1 gives a short description of ab initio methods, Hartree-Fock and post–Hartree-Fock, focusing on the Gaussian computer programs. Chapter 2 describes semi-empirical calculations and their applications to biological systems. Chapter 3 addresses itself to electrostatic properties of molecules, as determined by quantum-chemical methods. The density functional method is discussed in chapter 4. Chapter 5 compares theoretically obtained parameters to experimental data.

The second part of the book, consisting of chapters 6 and 7, describes the application of ab initio calculations to such biological systems as amino acids and peptides (Chapter 7) and anti-cancer drugs (Chapter 6).

The book addresses itself mainly to biochemists who would like to augment experimental studies with theoretical calculations.

This page intentionally left blank

Contents

This page intentionally left blank

Contributors

John D. Ewbank, Department of Chemistry, University of Arkansas, Fayetteville, Ark. 72701, USA

Xiaoqin Jiang, Department of Chemistry, University of Arkansas, Fayetteville, Ark. 72701, USA

Jane S. Murray, Department of Chemistry, University of New Orleans, New Orleans, La. 70148, USA

Susan Q. Newton, Department of Chemistry, University of Arkansas, Fayetteville, Ark. 72701, USA

Peter Politzer, Department of Chemistry, University of New Orleans, New Orleans, La. 70148, USA

Nigel G. J. Richards, Department of Chemistry, University of Florida, Gainesville, Fl. 32611-7200, USA

Anne-Marie Sapse, City University of New York, Graduate School and John Jay College, New York, N.Y. 10019, USA and Rockefeller University, New York, N.Y. 10021, USA

Lothar Schäfer, Department of Chemistry, University of Arkansas, Fayetteville, Ark. 72701, USA

Jacques Weber, Université de Genève, Department of Physical Chemistry, Geneva, Switzerland

Tomasz A. Wesolowski, Université de Genève, Department of Physical Chemistry, Geneva, Switzerland

This page intentionally left blank

Introduction

In its early stages, more than fifty years ago, molecular quantum mechanics began to be used to study the structure of matter, by relating the properties of atoms and molecules to the position and interaction of electrons and nuclei. The basic equation of quantum mechanics, the Schrödinger differential equation, was applied to chemical systems starting with the simplest one, the hydrogen atom. Even for this simple system, solving the Schrödinger equation requires advanced calculus methods using special functions such as Legendre polynomials. However, for any system featuring more than one electron, starting with the helium atom, it has long been recognized that the Schrödinger equation cannot be solved analytically and so approximate methods had to be developed. Accordingly, progress in quantum chemistry is made in two directions: the refining of the methods for exact treatment of small systems such as atoms and small molecules, and the finding of reasonable approximate models in order to be able to treat larger systems. Both paths of research make use of computational chemistry, which simulates chemical structures and reactions numerically. The results of the simulations are tested by comparing them to the experimental data, when such data are available, and consequently are used for predictions. This way, a few hours of computer simulations can provide information which might require months of laboratory work. Short-lived intermediates which can not be isolated experimentally can be characterized via computer modeling. Transition states can be identified and characterized. An example of computer usefulness is the search for a putative intermediate of a reaction. This would normally take many hours of laboratory work. If the theoretical work shows that this intermediate does not represent a minimum on the energy hypersurface and as such could not be isolated, much experimental work is saved.

The information obtained via computational studies on a system concerns its structure and its reactivity.

The structure determination is obtained by geometric optimization (finding the set of geometrical parameters such as bond lengths, bond angles, and dihedral angles which enable the system to adopt the lowest possible energy state). In addition to the geometry, calculations provide both the value of the total energy of the system (the sum of electronic energy and the energy of the nuclei), and binding energies of monomers to form oligomers

(usually expressed as the energy of the oligomer minus the sum of the energies of the monomers). The method also provides a description of the wave function of the system, with occupied and virtual orbitals, for ground state or excited states, net atomic and bond charges, as well as electric fields, spin distributions, and vibrational frequencies resulting from the vibration of the atoms within the molecule.

The chemical reactivity of compounds is studied by transition state location, activation energy calculations, and relative energies of the products versus reactants.

These calculations are now applied to a large variety of systems, but even as early as 1953, the pioneering work of Alberte and Bernard Pullman opened the door for applying quantum chemistry to the investigation of biological systems. Their early papers investigate the possibility that the carcinogenic activity of polynuclear aromatic hydrocarbons is related to their electronic structure. For instance, Alberte Pullman, in the paper "Complements on the Factors Determining the Existence of Carcinogenic Activity in the Aromatic Hydrocarbons," (presented at the Seances de L'Academie Française and published consequently in Comptes Rendus des Seances de L'Academie Française), uses the LCAO (linear combinations of atomic orbitals) method to calculate the energies of ortho and para polarization of some aromatic hydrocarbons and also tries to correlate them to the carcinogenic activity for prediction purposes. Bernard Pullman and Jeanne Baudet in "The Metabolism of Carcinogenic Hydrocarbons," use the structure parameters obtained previously for quinoline systems to calculate the bond index and free valence index in the "M" region of aromatic hydrocarbons, and to describe an epoxy formation.

Although the correlation between structural properties of aromatic hydrocarbons and their carcinogenic properties proved to be much more complicated than was hoped, this type of calculation opened the door to the application of quantum chemistry to biological systems. The calculations are applied not only to cancer-related problems, but also to the study of amino acids, peptides, nucleotides, and other than anti-cancer therapeutic agents.

The size of the systems is still a limiting factor, but huge strides are being made in writing programs which can handle larger systems. Ab initio calculations can be performed now on systems which, not many years ago, could only be treated with semi-empirical or empirical methods. Researchers are striving to find the optimum combination of accuracy and expediency and the ultimate goal is to reduce the computational effort at no accuracy cost.

The subsequent chapters will describe various quantum-chemical methods, compare them to experimental results and discuss their applications to such biological systems as amino acids, peptides, carcinostatic drugs, and DNA fragments. Proteins and large DNA fragments cannot be treated as yet with quantum-chemical methods, due to their size, but progress is being made continuously.

Molecular Orbital Calculations for Biological Systems

This page intentionally left blank

1

Ab Initio Calculations

Anne-Marie Sapse

Various difficulties of classical physics, including inadequate description of atoms and molecules, led to new ways of visualizing physical realities, ways which are embodied in the methods of quantum mechanics. Quantum mechanics is based on the description of particle motion by a wave function, satisfying the Schrodinger equation, which in its "time-independent" form is:

$$((-h^2/8m\pi^2)\nabla^2 + V)\Psi = E\Psi$$

or, for short: $H\Psi = E\Psi$

In this equation, H, the Hamiltonian operator, is defined by $H = -(h^2/8m\pi^2)\nabla^2 + V$, where h is Planck's constant (6.6 10^{-34} Joules), m is the particle's mass, ∇^2 is the sum of the partial second derivative with x,y, and z, and V is the potential energy of the system. As such, the Hamiltonian operator is the sum of the kinetic energy operator and the potential energy operator. (Recall that an operator is a mathematical expression which manipulates the function that follows it in a certain way. For example, the operator d/dx placed before a function differentiates that function with respect to x.) E represents the total energy of the system and is a number, not an operator. It contains all the information within the limits of the Heisenberg uncertainty principle, which states that the exact position and velocity of a microscopic particle cannot be determined simultaneously. Therefore, the information provided by $\Psi(n_i t)$ is in terms of probability: $\Psi^2_{(x,t)}$ is the probability of finding the particle between x and $x + dx$, at time t.

The Schrödinger equation applied to atoms will thus describe the motion of each electron in the electrostatic field created by the positive nucleus and by the other electrons. When the equation is applied to molecules, due to the much larger mass of nuclei, their relative motion is considered negligible as compared to that of the electrons (Born-Oppenheimer approximation). Accordingly, the electronic distribution in a molecule depends on the position of the nuclei and not on their motion. The kinetic energy operator for the nuclei is considered to be zero.

For a many-electron molecule, the Hamiltonian operator can thus be written as the sum of the electrons' kinetic energy term, which in turn is the sum of individual electrons'

kinetic energies and the electronic and nuclear potential energy terms. The first term can be expressed as:

$$-\frac{\hbar^2}{2m}\sum_i^n\left(\frac{\delta^2}{\delta_i^2}+\frac{\delta^2}{\delta y_i^2}+\frac{\delta^2}{\delta Z_i^2}\right)$$

where the sum is taken over the number of electrons.

The electronic potential energy is due to the attraction between the positive nuclei and the negative electrons, which can be expressed as:

$$-\frac{\Psi^2}{2Zm}\sum_i^n\left(\frac{\delta^2}{\delta_i^2}+\right.$$

where i represents as before the summation over electrons and I is the summation over nuclei. Z is the charge of the I nucleus and $R_I - r_j$ is the distance between the Ith nucleus and the ith electron. To this term, one must add the term representing the repulsion between electrons:

$$\sum_i\sum_{j\neq i}\frac{e^2}{|\vec{r}_i-\vec{r}_j|}$$

where $\vec{r}_i - \vec{r}_j$ represents the distance between electron i and electron j. The nuclear potential energy, that is the repulsion between nuclei, is given by

$$\sum_I\sum_{J\neq I}\frac{Z_I Z_J e^2}{|\vec{\mathrm{R}}_I-\vec{\mathrm{R}}_J|}$$

where $\vec{R}_I - \vec{R}_J$ is the distance between nucleus I and nucleus J, and must also be added.

In this way, the Schrödinger equation $\mathrm{H}\Psi = \mathrm{E}\Psi$ describes the motion of electrons in the electrostatic field of fixed nuclei.

As mentioned in the introduction, this equation cannot be solved analytically for systems larger than the hydrogen atom. Therefore, a number of approximations have to be introduced.

When the Schrödinger equation is applied to atoms, the wave function Ψ is made up of a set of functions called atomic orbitals, which correspond to a given energy state containing a number of electrons determined by Pauli's exclusion principle. If the exact form of these functions is known, the exact energy of the system can be computed. If the exact function is not known, an educated guess can be used. The Variation Principle states that the expectation value of the energy based on the choice of an appropriate Ψ will always be higher than the exact energy of the system. Accordingly, minimizing the energy as a function of parameters characterizing a wave function establishes equations whose solutions represent the set of parameters corresponding to the energy closest to the exact energy of the system, obtainable for a wave function of that particular form.

In order to find a good approximate wave function, one uses the Hartree-Fock procedure. Indeed, the main reason the Schrödinger equation is not solvable analytically is the presence of interelectronic repulsion of the form $e^2/\vec{r}_i - \vec{r}_j$. In the absence of this term, the equation for an atom with n electrons could be separated into n hydrogen-like equations. The Hartree-Fock method, also called the Self-Consistent-Field method, regards all electrons except one (called, for instance, electron 1), as forming a cloud of electric charge

through which electron 1 moves. The electronic cloud is characterized by its charge density which is, in turn, a function of the atomic orbitals describing the electrons. Once the interaction between electron 1 and the cloud is calculated, the Schrödinger equation is solved and improved atomic orbitals are obtained, which replace the initial guess. The new set of orbitals is used to calculate a new charge density which leads to an even better set of atomic orbitals. This iteration procedure is used until a certain threshold is reached.

When the Hartree-Fock method is applied to molecules, molecular orbitals are used instead of atomic orbitals. To construct the molecular orbitals, one widely used approximation is LCAO (linear combinations of atomic orbitals). According to molecular orbital theory, the total wave function of the system is written as a combination of molecular orbitals, φ_i's, which are complemented by spin functions describing electrons in terms of spin $\frac{1}{2}(\alpha)$ or $-\frac{1}{2}(\beta)$.

The Hartree-Fock method involves the calculation of integrals of atomic functions. The computation time required is approximately N^4, where N is the number of atoms. For a large system, this makes the calculations a formidable task. Indeed, integrals carried out for atomic orbitals which are exponential functions of the e^{-ar} form, where r is the distance from the electron to origin, are very cumbersome. To facilitate the task, these functions were replaced by Gaussian functions of the form e^{-r^2}, which greatly shorten the computation time. However, since atomic orbitals *are not* of Gaussian form, they had to be expanded in a series of Gaussians, with the general form:

$$G(x,y,z) = CX^nY^mZ^le^{-\alpha r^2}$$

where α is a constant determining the radial extent of the function and C is also a constant.

These "primitive" Gaussians form the actual basis functions which are called "contracted" Gaussians. The atomic orbitals are then expressed as:

$$\Psi_i = \sum_{\mu} c_{\mu} x_{\mu}$$

where c_{μ} are coefficients and x_{μ} are the contracted Gaussians $X_{\mu} = \sum_{\mu} d_{\mu j} g_j$

A number of computer programs were devised in order to perform calculations on different systems. One of the most widely used series of programs is the series of Gaussian programs written at the Carnegie-Mellon Institute in Pittsburgh. These programs establish sets of basis functions and use them for Hartree-Fock and post-Hartree-Fock calculations. The smallest basis system used is called STO-3G, where STO stands for Slater-type orbitals, which are the s, p, and d orbitals of atoms. In the STO-3G case, s orbitals are used for hydrogen atoms and s and p for the other atoms. Each Slater-type orbital is expanded into three Gaussian functions. In general, this set is used for systems too big to allow the use of larger basis sets. Other minimal basis sets are STO-4G, STO-5G, and STO-6G. The value of the computed energy of the system depends on the number of Gaussians used. However, energy differences, optimum geometries, and atomic charges are fairly insensitive to the size of the expansion. STO-3G-calculated optimum bond distances are very accurate when compared with experimental ones, in most cases. For instance, the C-F bond length in CH_3F is calculated to be 1.384A and is found to be 1.385A experimentally. The C-C bond lengths are usually reliably computed, as in ethane, 1.538 v. 1.531, for example. The C-N bond length is found to be 1.153 v. 1.154 in HCN, showing excellent agreement. The C-O bond is slightly overestimated, with 1.217 v. 1.203, and the C-N bond length in peptide bonds is also predicted to be longer than experimental values. The complete set of A-B lengths for A and B being C, N, O, and P show a mean absolute dissociation from experiment of 0.03A.