APPLICATIONS OF THE QUANTUM THEORY OF ATOMS IN MOLECULES TO CHEMICAL AND BIOCHEMICAL PROBLEMS

by

CHÉRIF F. MATTA

B. Pharm. Sci., Dipl.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

MATTA

Applications of the Quantum Theory of Atoms in Molecules

Doctor of Philosophy (2002)
(Chemistry)

McMaster University
Hamilton, Ontario

TITLE: Applications of the Quantum Theory of Atoms in Molecules to Chemical and Biochemical Problems

AUTHOR: Chérif F. Matta, Bachelor of Pharmaceutical Sciences (*University of Alexandria, Egypt*), Graduate Diploma (*Sadat Academy, Egypt*).

SUPERVISOR: Professor Richard F. W. Bader

NUMBER OF PAGES: xxiii, 300

To the memory of my beloved father, Farid A. Matta

To my beloved mother, Nabila Matta (née Abdel-Nour)

To my beloved Magued, Heba, and Sarah Matta

May the Lord bless them all

Abstract

The quantum theory of atoms in molecules (QT-AIM) is a generalization of quantum mechanics to open subsystems. The theory enables one to study the properties of atoms or functional groups within a molecule, recovering the empirical observation that they exhibit characteristic and additive contributions to every molecular property. This thesis applies theory to explain and predict experiment. In this work, atomic and group properties were shown to provide an excellent basis for the construction of quantitative statistical models which accurately predict the physicochemical properties of the amino acids and uncover regularities in the genetic code. The transferability and additivity of atomic properties were exploited to develop a method of assembling large molecules from buffered fragments, a method which reduces the computational effort and – as a result - brings large biological molecules within theoretical reach. The buffered fragments method was applied to obtain accurate properties of key opioid molecules. The additivity of atomic properties applies to any and all molecular properties including, for example, dielectric polarization. The dielectric polarization of a crystal was shown to be reducible to a sum of the polarization of its composing cells taken together with a sum of the dipoles arising from the charge transfer between neighbouring cells. The problem of aromaticity has been recast in terms of the pair densities in conjunction with QT-AIM, an approach that was shown to explain and predict several chemical and spectroscopic properties of polycyclic aromatic hydrocarbons. The criteria of bonding were revisited and applied to provide a clear-cut answer in a case where distance alone – as a criterion for bonding - provides ambiguous answers.

The thread tying this work together is the philosophy of "reductionism" in the narrow sense that *the whole (e.g. a molecule or a crystal) is nothing but the sum of its spatially bounded non-overlapping parts (atoms)*. By resorting to this philosophy one not only gains a deeper insight into the physical basis of chemistry but is capable of accurately predicting experiment as well.

Acknowledgements

I am most indebted to my supervisor, mentor, and dear friend, **Professor Richard F. W. Bader**. Professor Bader has been the greatest source of encouragement and support both professionally and personally during this endeavor. Not only did he spearhead the effort to help me in joint projects, he presented me with every possible assistance and advice in projects that I have pursued on my own or in collaboration with workers from other research groups. His vast knowledge, and sound advice were invaluable in guiding every aspect of my graduate work and were essential in completing this thesis. Working in his group was the most rewarding learning experience.

Professor Joseph D. Laposa and **Professor Paul H. M. Harrison** were very kind to accept volunteering as members of my supervisory committee. For over six years Professors Laposa and Harrison generously spent invaluable time in assessing my progress, providing advice and suggestions, and correcting the science and the English in my various reports. I am very grateful for their contributions and help. A collaboration with Professor Harrison has been an enriching experience through which I have learned the techniques of x-ray diffraction. My supervisor and supervisory committee are also thanked for having chosen "*the theory of the scanning tunneling microscope*" as my comprehensive examination topic, a topic that I find very interesting and which paved the way for my post-doctoral research in the Polanyi research group at the *University of Toronto*.

Professor John C. Polanyi (*University of Toronto*), my post-doctoral supervisor, mentor, and dear friend is acknowledged for his very strong support, backing, and encouragement during the writing of this thesis. I am very much indebted to him and look forward to an exceptional learning opportunity in his group.

The present and past members of the Bader research group have all been a source of inspiration. In particular, I would like to thank **Professor Jesús Hernández-Trujillo** (*Universidad Nacional Autónoma de México - UNAM*), with whom I had the great pleasure of collaborating in the

work on aromaticity reviewed in Chapter 8. I would also like to thank **Dr. Fernando Martín** (*University of California at Los Angeles - UCLA*) and **Ms. Maggie Austen** (*McMaster University*). Both have kindly shared their knowledge and expertise and their invaluable assistance is sincerely acknowledged. I am much obliged to **Dr. Ting-Hua Tang** (*McMaster University*) who was very patient in teaching me the basics in the early days. I am also very grateful to **Professor Friderich Biegler-König** (*University of Applied Sciences, Bielefeld*) and to **Dr. Todd Keith** (*University of Minnesota*) for having made their programs, respectively AIM2000 and AIMALL97, available.

Professor Ronald J. Gillespie and **Professor Nick H. Werstiuk** (both at *McMaster University*) are acknowledged for stimulating discussions and invaluable suggestions. Professor Gillespie taught me a lot of chemistry during our collaboration which resulted in two published review articles so far. **Dr. Jim F. Britten** (*McMaster University*) introduced me to x-ray diffraction and has always made himself available when I had a question or needed help, I thank him very much.

This work would not have been possible without the support of the **Chemistry Department, McMaster University**, a support for which the current and past Chairs and Associate Chairs are deeply acknowledged. I would like to thank the Department's computer specialist **Mr. Mike Malott** for his invaluable technical support. Many thanks are also due to the Department's administrator **Ms. Carol Dada** and to the wonderful Department staff members: **Barbara, Josie, Julie, Marylyn, Nancy**, and **Tammy** for giving me every possible personal and administrative help over the years.

This section cannot be complete without due acknowledgement to my roots in Egypt where I had the privilege of being educated at the **Collège Saint Marc** and at the **University of Alexandria**. It is with the greatest gratitude that I would like to thank my early mentors: **Professor Mohamed E. El-Raey** (*Institute of Graduate Studies and Research, University of Alexandria*), **Dr. Abdel Aziz Saleh** (*World Health Organization*), **Professor Labiba El-Khordagui**, **Professor Said A. Khalil**, and **Professor Nawal Khalafallah** (all at the *Faculty of Pharmacy, University of Alexandria*).

C. F. Matta

Toronto, April 2002

Preface

The present thesis provides an account of most of the research that was undertaken by the author for the Ph. D. degree. The chapters of the thesis are *based on*, but are *not reproductions of*, papers – authored or co-authored by the author - that have been already published or on manuscripts which are currently under consideration for publication by different journals. This thesis is more of a *review* of the author's research rather than a re-impression of the original text upon which it is based. The original papers/manuscripts were heavily edited, modified, updated, and parts were re-written in order to suit the style of a thesis. Effort was made to keep redundancy across the different chapters to a minimum and to consolidate the material in a coherent whole. Each chapter reviews the material presented in one (or two) paper(s) and/or manuscript(s).

As required by the *School of Graduate Studies of McMaster University*, introductory material entitled "Statement of the Problem" was added as the first section of each chapter, which - together with the "Conclusion" at the end of the previous chapter - show how that particular chapter relates to the rest of the thesis. The thesis falls into ten chapters grouped into five distinct parts. Part I (chapters 1 and 2) provides a general introduction to the thesis by reviewing the theories and types of problems which will be addressed in later chapters. This part also provides most of the equations and the definitions that are referred to in different places in the following chapters. Part II (chapters 3-5) represent a comprehensive study of the genetically-encoded amino acids. How can one exploit the additivity and transferability of the atomic properties, made abundantly evident in part II, is the subject of the two chapters in part III (chapters 6 and 7). A new method for obtaining the properties of large and complex biological molecules from small buffered fragments is described in chapter 6. In the next chapter (7), additivity is applied to a different problem: the dielectric polarization of large systems. The last research part of the thesis, part IV (chapters 7 and 8), attack two important chemical problems,

both of which are investigated in the light of the physics of electron delocalization in conjunction with the quantum theory of atoms in molecules. Chapter 7 recasts the problem of aromaticity in terms of the delocalization of the Fermi hole and provides an introduction to the quantitative description of localization and delocalization used in this part of the thesis. Chapter 8, deals with the criteria of bonding to titanium and emphasizes the consequences of bonding. The last part of the thesis draws general conclusions of the work and provides a critical assessment of what have been accomplished.

Conforming to the requirements of the *School of Graduate Studies of McMaster University,* a full disclosure of collaborative work follows. The disclosure is in the form of an outline of the thesis giving the reference to the complete citation of the work upon which each part is based. In every case, the author of the present thesis is either the single author, the first author, or one of two co-authors of the paper or manuscript.

THESIS OUTLINE WITH DISCLOSURE OF COLLABORATIVE WORK:

PART I **THEORETICAL FOUNDATIONS**

Chapter 1 Methods of Obtaining the Electron Density

Chapter 2 The Quantum Theory of Atoms in Molecules

PART II **COMPREHENSIVE STUDY OF THE FREE AMINO ACIDS**

Chapter 3 Effects of Proton-Transfer and of Conformation on Geometric, Atomic and Bond Properties of the Genetically-Encoded Amino Acids

Based on the paper: (*Matta and Bader, 2000*).

Chapter 4 Molecular Geometries of the Genetically-Encoded Amino Acids

Based on the manuscript: (*Matta and Bader, 2002a*).

Chapter 5 Atomic and Bond Properties of the Genetically-Encoded Amino Acids and their Correlations to Physicochemical and Biological Properties

Table of Contents

List of Tables

List of Figures

List of Schemes

Conventions Adopted Throughout the Thesis

1. Labeling of Computational Levels of Theory and of Basis Sets

Levels of theory are indicated using the convention: Theory // Basis Set (I) / Basis Set (II), where "Theory" refers to the theoretical method used in the geometry optimization as well as in the calculation of the final wavefunction. For example, RHF//6-311++G**/6-31+G* indicates a wavefunction calculated using the restricted Hartree-Fock theory and a 6-311++G** basis set at a geometry optimized using the same theory and the smaller 6-31+G* basis set.

The "G" in the basis set designation indicates that gaussian basis functions are used to construct the basis set. Ideally one would use a basis set consisting of Slater type orbitals (STOs) each of which exhibit the correct cusp at the nuclear positions. The exponential dependence of an STO have the form $\exp\left[-\left(\frac{\xi r}{a_0}\right)\right]$ where a_0 is the Bohr radius, r the distance from the nucleus and ξ is the orbital exponent. The STOs are computationally prohibitive, this is why one approximates the STOs by linear combinations of gaussian-type orbitals (GTOs). The exponential dependence of GTOs is of the form $\exp\left[-\left(\frac{\alpha r^2}{a_0^2}\right)\right]$ where α is called the gaussian exponent. GTOs have the computational advantage that the product of any two gaussian functions is another gaussian function, greatly simplifying many-centre integrals such as two-centre overlap integrals.

Basis set designations such as n_1-$n_2n_3n_4$++G** indicate that the basis set is a "split valence" basis set (indicated by the dash), to which diffuse functions (plus signs) and polarization functions (stars) have been added. The digit before the dash indicates the number of gaussian primitive functions used in linear combination to approximate the core Slater-type orbitals. Thus in the 3-21G basis set, three gaussians are used to approximate the 1*s* orbital of a hydrogen or a carbon atom and we say that the three primitive gaussians have been contracted into the basis function 1*s*. The number of digits after the dash indicate the split in the valence shell: two digits indicate a double split valence

shell, three digits a triple split etc. A double split valence basis set uses two sizes of contracted functions for each orbital-type in the valence shell. The digits after the dash indicates the number of gaussian primitives used in the contractions of the split valence basis functions. Thus in the 6-31G basis set, the core is represented by six gaussian primitive functions and the valence shell by two sizes of basis functions: one is contracted from 3 primitive gaussians and the other is just one gaussian primitive. In the case of hydrogen and carbon, for example, these are:

H: $\underbrace{1s}_{3\,GTO}\ ,\ \underbrace{1s'}_{1\,GTO}$ C: $\underbrace{1s}_{6\,GTO}\ ,\ \underbrace{2s, 2px, 2py, 2pz}_{3\,GTO},\ \underbrace{2s', 2px', 2py', 2pz'}_{1\,GTO}$.

Split valence basis set introduces a flexibility in the spatial extent of a given type of basis functions but does not allows for a corresponding angular flexibility of the basis set. The addition of basis functions with angular momentum beyond what is required for the ground state atom removes this limitation. These additional basis functions are termed "polarization functions" and indicated by the stars after the G, the first star referring to heavy atoms and the second to the hydrogens. For example, the basis set 6-31G** indicates that the 6-31G basis set have been supplemented with *d* functions on the heavy atoms (first star) and with *p* functions on the hydrogens (second star). The 6-31G** basis set can be equally designated as 6-31G(*d,p*). The accurate description of anions, hydrogen-bonded systems, and systems with significant electron density far from the nucleus requires the addition of *s* and *p* basis functions with much larger spatial extension than the standard ones. These are called "diffuse functions" and are denoted by the plus signs before the G. Thus, the 6-31+G* basis set is the 6-31G(*d*) basis set supplemented with diffuse function to the heavy atoms, while the 6-31++G(*d*) adds diffuse functions to the hydrogens as well.

2. Conventions Applicable to Contour Maps

(Figures: 2.1, 6.5, 7.1, 7.2, 7.3, 7.4, 8.5, and 9.6)

For the electron density, the virial field density, and the kinetic energy density the outermost contour equals 0.001 au. The next contours have the values of: 0.002, 0.004, 0.008, 0.02, 0.04, 0.08, 0.2, 0.4, 0.8, 2.0, 4.0, 8.0, 20.0, 40.0, and 80.0 au respectively. In the case of the Laplacian plots, solid contours

denote regions of charge concentration where $\nabla^2\rho$ <0 and dashed contours denote regions of charge depletion $\nabla^2\rho$ >0, at increase and decrease from a zero contour in steps of ±0.001, ±0.002, ±0.004, ±0.008, ±0.02, ±0.04, ±0.08, ±0.2, ±0.4, ±0.8, ±2.0, ±4.0, ±8.0, ±20.0, ±40.0, ±80.0 au respectively.

3. Integrated Atomic and Molecular Volumes (v(Ω))

Except when stated otherwise, all integrated volumes are up to the external isodensity envelope of ρ = 0.001 au, which represents the van der Walls surface in the gas phase and which usually enclose more than 99% of the electron population of the system.

4. Usage of the Words "Tautomerization" and "Tautomer" in Chapter 3

In chapter 3, the word "tautomer" is used to designate the ionization state of the main-chain group: whether it is in the zwitter ionic (**z**)-form [-CαHα(NH_3^+)COO^-] or in the non-zwitter ionic (**n**)-form [-CαHα(NH_2)COOH]. Thus the word "tautomer" in chapter 3 refers to proton-transfer tautomerism and strictly follows the definition: "Two isomeric structures which differ significantly in the relative positions of their atoms are not resonance forms but are in fact separate isomeric compounds, called **tautomers**". *(Hendrickson JB et al., 1970, p.160)* However, the usage of the word – while correct - is atypical, since "in most cases tautomers are of similar energy and are interconvertible ... (**tautomeric change** or **tautomerism**)", *(Hendrickson JB et al., 1970, p.160)* which is not the case since the **n**-form of an amino is *ca.* 25-28 kcal/mol more stable than the corresponding **z**-form in the gas phase (at the RHF/6-311++G**/6-31+G* level of theory). Thus one cannot draw the familiar equilibrium double arrow, and:

$$\mathbf{n}-tautomer \not\rightleftharpoons \mathbf{z}-tautomer\,.$$

List of Abbreviations

Abbreviation	Meaning
AA	amino acid molecule (free molecule)
Aa	amino acid residue (amino acid within a polypeptide)
AIM	atoms in molecules
AO	atomic orbital
au	atomic unit
bcp	bond critical point
BL	bond length
CC	charge concentration
cc	complex conjugate
ccp	cage critical point
CNDO	complete neglect of differential overlap
cp	critical point
CP	critical point
Cp	cyclopentadienyl ring
CSI	charge separation index
CysS	cystine
DFT	density functional theory
emu	electromagnetic unit per mol
GGA	generalized gradient approximation
GTO	gaussian-type orbital
HF	Hartree-Fock
HK	Hohenberg and Kohn
IAM	independent atom model
IAS	inter-atomic surface
IUPAC	International Union of Pure and Applied Chemistry
KS	Kohn-Sham
L.H.S	left-hand side of an equation
LDA	local density approximation
LOCC	ligand-opposed charge concentration
MCG	main-chain group (of an amino acid)
MO	molecular orbital
MSI	molecular similarity index
ncp	nuclear critical point
NMR	nuclear magnetic resonance spectroscopy
n-MR	n-membered ring, e.g. 5-membered ring
PEO	7-α-[1-(R)-hydroxy-1-methyl-3-phenylpropyl]-6,14-endo-etheno-tetrahydrooripavine
PET	7-α-[1-(R)-hydroxy-1-methyl-3-phenylpropyl]-6,14-endo-ethenotetrahydrothebaine
PH	Poincaré-Hopf relationship
QSAR	quantitative structure-to-activity relationships
QT-AIM	quantum theory of atoms in molecules
R.H.S.	right-hand side of an equation
rcp	ring critical point
RHF	restricted Hartree-Fock
SCF	self-consistent field
STO	Slater-type orbital
VSCC	valence shell charge concentration
VSEPR	valence-shell electron-pair repulsion model of molecular geometry

Contents

PART I:

THEORETICAL FOUNDATIONS

Chapter 1

Methods of Obtaining the Electron Density

> "Thus on this view we have in the cathode rays matter in a new state, a state in which the subdivision of matter is carried very much further than in the ordinary gaseous state: a state in which all matter - that is, matter derived from different sources such as hydrogen, oxygen &c. - is one and the same kind; this matter being the substance from which all the chemical elements are built up."
>
> J. J. Thomson, October 1897 (*Thomson, 1897*).

1.1 Introduction

Over 104 years have passed already since the first discovery of an elementary particle, the electron, by J. J. Thomson. The world has never been the same since that day, as this discovery has had an impact on almost every aspect of human activities from electric appliances, to computers, material science, chemistry, physics, and biology. Perhaps more importantly, the electron accelerated the development of the most fundamental physical theory to date: quantum mechanics. The partnership between the electron and quantum mechanics is deeply rooted in modern scientific history, a story that stretches for a little over a century. A review of the development of the theory of the electron and / or quantum mechanics warrants a thesis by itself, and the interested reader may consult (*Springford, 1997*) for a history of the electron and (*Tarasov, 1980*) for an outline of the historical development of early quantum theory.

The theory underlying much of the work presented in this thesis, the quantum theory of atoms in molecules (QT-AIM), is a generalisation of quantum mechanics to open systems (i.e. parts of a

quantum system) as opposed to closed whole systems. In doing so, QT-AIM provides a framework that bridges the gap separating physics on one side and experimental chemistry on the other by recovering - from quantum mechanics - the concepts and observations of experimental chemistry.

QT-AIM defines the open systems in terms of the topology of the electron density, a field that we can observe, that we can calculate, and that determines chemical behaviour and reactivity. This chapter reviews the principal methods that are employed to obtain the electron density, with an emphasis on the theoretical methods utilized in the present thesis. The following chapter, which concludes this part of the thesis, reviews the basic concepts of QT-AIM which is used in this thesis to solve a number of selected problems in chemistry and biochemistry.

We begin by asking the question: what exactly is the electron density ($\rho(\mathbf{r})$)? Simply, $\rho(\mathbf{r})$ is the average number of electrons per unit volume at the point $\mathbf{r}\equiv(x,y,z)$. The following section (1.2) reviews the important connection between this experimentally accessible quantity, $\rho(\mathbf{r})$, and the theoretically accessible wavefunction $\Psi(\mathbf{r})$, the carrier of all the information that can be known about the system.

1.2 The Electron Density

A time-independent wavefunction is a function of the position in space ($\mathbf{r}$) and the spin degrees of freedom which can be either up or down. The physical interpretation of the wavefunction is due to Max Born (*Born, 1926a; Born, 1926b*) who was the first to interpret the square of its magnitude, $|\Psi(\mathbf{r})|^2$, as a probability density function, or probability distribution function. This probability distribution specifies the probability of finding the particle (here, the electron) at any chosen location in space ($\mathbf{r}$) in an infinitesimal volume $dV = dxdydz$ around $\mathbf{r}$. The probability of finding the electron at $\mathbf{r}$ in dV is given by $|\Psi(\mathbf{r})|^2 dV$ which is required to integrate to unity over all space (normalization condition). A many-electron system, such as a molecule, is described by a many-electron wavefunction $\Psi(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\ldots,\mathbf{r}_N)$ which when squared gives the probability density of finding electron

(1) at $\mathbf{r}_1$, electron (2) at $\mathbf{r}_2$, ... and electron (N) at $\mathbf{r}_N$ simultaneously, i.e. the probability of a particular instantaneous configuration of all electrons in the system. The probability of finding say electron (1) at $\mathbf{r}_1$ without specifying the location of the N-1 remaining electrons is found by integrating the many-electron wavefunction over the coordinates of all electrons except electron (1). In other words, the probability of finding electron (1) at $\mathbf{r}_1$ - irrespective of the positions of the remaining electrons in the molecule - is given by:

$$\sum_{spins} \int |\Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)|^2 \, d\mathbf{r}_2 ... d\mathbf{r}_N . \tag{1.1}$$

However, since electrons are indistinguishable, and thus cannot be labeled, what is true for electron (1) is true for any electron in the system, and if we multiply the latter integral by the number of electrons in the system we obtain a one-electron density function, commonly known as the electron density:

$$\rho(\mathbf{r}) = N \sum_{spins} \int |\Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)|^2 \, d\mathbf{r}_2 ... d\mathbf{r}_N \equiv N \int |\Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)|^2 \, d\tau' \tag{1.2}$$

which is the probability density of finding a single electron, no matter which, at $\mathbf{r}$ weighted by the total number of electrons in the system. (The integration over the spatial coordinates of all electrons but one followed by the summation over all spins will be denoted by $\int d\tau'$ in this thesis). Integrating the electron density (Eq. 1.2) over all space with respect to the coordinates of electron (1) yields the total number of electrons (N) in the molecule. The electron density is accessible theoretically (*Smith and Absar, 1977*) as well as experimentally (*Coppens, 1997; Koritsanszky and Coppens, 2001*) from low-temperature x-ray diffraction experiments using multipolar refinement, (*Hansen and Coppens, 1978*) and its importance lies in the chemical insight its topological properties provide.

Figure 1.1 displays two isodensity envelopes of the electron density function obtained from theoretical calculations (DFT, B3LYP/6-31++G**) of the guanine-cytosine DNA base pair. The 0.001 au isodensity envelope corresponds to the van der Waal's envelope in the gas phase and usually contains over 99% of the total electron population of the molecule, while the 0.002 au envelope represents the molecular volume in condensed phase. (*Bader, 1990*)

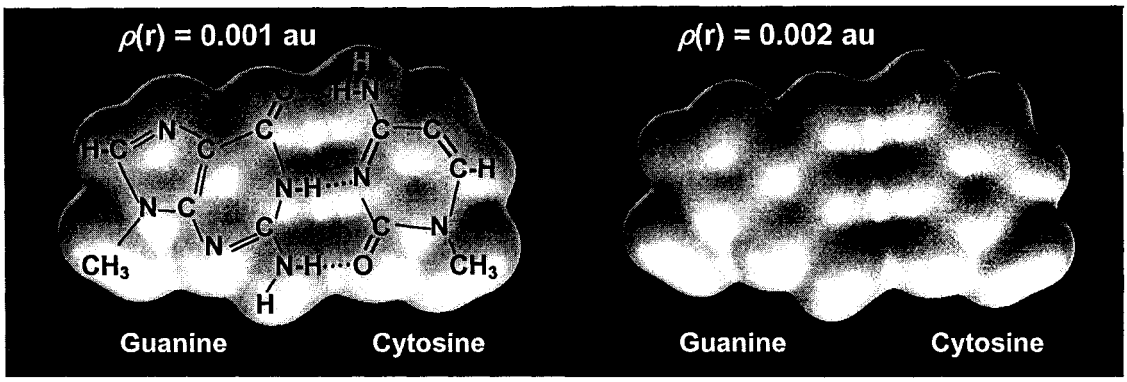


Figure 1.1. Displays of two isodensity $\rho(\mathbf{r})$ envelopes of a guanine-cytosine triply hydrogen bonded DNA base pair. The 0.001 au envelope corresponds to the van der Waal's molecular size in the gas phase while the 0.002 au is a better measure of molecular size in the condensed phase.

1.3 The Electron Density from Experiment (X-Ray Diffraction)

The electron density in a crystal is a three dimensional periodic function since the density within the unit cell is repeated through spatial translation along the three basis vectors defining the crystal lattice. Thus, the electron density in a crystal can be expressed: (*Sands, 1975*)

$$\rho(x+p, y+q, z+r) = \rho(x, y, z) \tag{1.3}$$

where p, q, and r are the unit cell dimensions. In other words, the density in a crystal has three periodicities (h,k, and l), along the x-, y-, and z-axes respectively. (*Rhodes, 1993*) The integers h, k, and l range from -∞ to ∞ and are the coordinates of a reflection in the reciprocal space. As for any periodic function, the density function can also be expanded as a Fourier series:

$$\rho(xyz) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} F(hkl)\ \exp\left[-2\pi i(hx+ky+lz)\right] \tag{1.4}$$

where V is the volume of the unit cell, x, y, and z the fractional coordinates at the point at which ρ is specified in the unit cell, and the expansion coefficients $F(hkl)$, known as structure factors, are the unknowns to be determined from the diffraction experiment. The experimental techniques known to

date can only measure a reflection's intensity, which is proportional to the square of the magnitude of the structure factor ($I(hkl) \propto |F(hkl)|^2$). Since the structure factors are generally complex numbers, phase information is lost in the process leading to the well known *phase problem*, a problem that has several elegant solutions which fall out of the scope of our discussion. (See (*Stout and Jensen, 1989*) for a review). For simplicity, thermal vibrations are ignored in the following discussion, since the physical ideas remain essentially the same.

The electron density in Eq. 1.4 is expressed as a triple sum rather than a triple integral over the reciprocal space because the structure factors, the reflections of the diffraction patterns, constitute a discrete set of values. In effect, the equation is a discrete Fourier transform of $F(hkl)$. Since each function of the variable χ can be transformed to a function of the reciprocal of the variable (χ^{-1}) through the Fourier transform, the transform is a reversible operation. The inverse transform of Eq. 1.4 is a function of the reciprocal space, and it is given by: (*Sands, 1975*)

$$F(hkl) = \int_0^1 \int_0^1 \int_0^1 \rho(xyz) \exp\left[2\pi i(hx + ky + lz) \right] dx\ dy\ dz \tag{1.5}$$

Eq.1.5 is an integral over the volume of the crystallographic unit cell since now the expansion "coefficients" are the values of the electron density field which is continuous within the unit cell. (Note the change in sign of the exponential). The expression 1.5 implies that the electron density at every volume element in the unit cell contributes to each and every structure factor $F(hkl)$.

In crystallography, "solving a structure" is achieved by iterative improvement of the agreement between intensities calculated from the structure factors of a model density (computed using Eq. 1.5) and those observed in the diffraction experiment. The overall agreement between the calculated and observed structure factors is measured by the so called residual factor or *R-factor*, defined:

$$R - factor = \frac{\sum \left| \, |F_{observed}| - |F_{calculated}| \, \right|}{\sum |F_{observed}|} \tag{1.6}$$

The solution of structure is considered to have converged when the *R-factor* reaches its smallest possible value, so that further Fourier recycling does not result in any change in its value to within a preset numerical accuracy. A typical value of the *R-factor* for an accurate determination is 0.002.

Thus, obtaining an experimental density starts with an educated guess for the density that the crystallographer builds using independent chemical information such as the number and types of atoms in the unit cell. This is usually accomplished through an atomic / molecular model of the unit cell which considerably simplifies the task of building the density in this cell. The spherical atoms are positioned in assumed positions inside an otherwise empty unit cell and the atoms are assumed to retain their ground-state electron densities, their electronic populations (no transfer of charge between atoms is allowed), and their spherical symmetry even after they combine to form a molecule. A molecule constructed by overlapping spherical atomic densities is known as a *promolecule* or as the *independent-atom-model* (IAM). The promolecule is an approximation capable of yielding excellent geometries and thus is routinely used to solve crystal structures, however, this approach does not provide an adequate representation of the subtle features induced by chemical bonding. Within this model, Eq. 1.5 can be re-written in terms of atomic densities as: (*Sands, 1975*)

$$F(hkl) = \sum_j \iiint \rho(xyz)\, \exp\left[\, 2\pi i(hx + ky + lz)\, \right] dx\, dy\, dz \tag{1.7}$$

where the sum is over all the atoms in the unit cell, and the triple integral is over the volume of a single atom. Introducing a coordinate transformation for atom j, centered at x_j, y_j, z_j, such that:

$$\left.\begin{aligned} x' &= x - x_j \\ y' &= y - y_j \\ z' &= z - z_j \end{aligned}\right\} \tag{1.8}$$

we get:

$$F(hkl) = \sum_j \left\{ \iiint \rho_j(x'y'z') \exp\left[2\pi i(hx' + ky' + lz')\right] dx'dy'dz' \right\} \left\{ \exp\left[2\pi i(hx_j + ky_j + lz_j) \right] \right\} \tag{1.9}$$

which can be re-written:

$$F(hkl) = \sum_j f_i \exp\left[2\pi i(hx_j + ky_j + lz_j)\right]. \tag{1.10}$$

where

$$f_j = \iiint \rho_j(x'y'z') \exp\left[2\pi i(hx' + ky' + lz')\right] dx'dy'dz'. \tag{1.11}$$

The pre-exponential term in Eq. 1.10, known as the atomic scattering factor or as the atomic form factor (f_i) and defined in Eq. 1.11, is the same for a given atom or ion despite its bonding pattern in the crystal under consideration. The value of f_i for a given species depends only on the Bragg angle θ and is scaled so that it equals the number of electrons in this species when $\theta = 0$. It is calculated quantum mechanically and tabulated in the *International Tables for X-Ray Crystallography*.

The promolecule approximation is less severe for heavier atoms, since most of the density of a heavy atom is spherical and remains undistorted by bonding. However, in the case of light atoms, this approximation can introduce significant errors manifested for example in the well known shortening of *X*-H bonds, where *X* is a second-row (or higher) atom. (*Coppens, 1997*) Generally, the promolecule refinement model is satisfactory, and commonly used, to obtain molecular geometries (except for hydrogen atom positions). The author of this thesis, in collaboration with others, has used this approach to solve several crystallographic structures of glycoluril molecules and has found very good agreement between calculated and experimental geometries. (*Duspara et al., 2001; Matta et al., 2000; Matta et al., 2002; Sun et al., 1998*) However, the promolecule approach is inadequate to study the subtleties of chemical bonding, and thus a more flexible refinement model is needed for such work.

Several improvements on the promolecule approach based on aspherical atomic modeling of the density are described in the literature. (*Carven et al., 1987; Figgis et al., 1980a; Figgis et al., 1980b; Hansen and Coppens, 1978; Stewart, 1976*) The most widely used aspherical model is probably Hansen and Coppens' MOLLY multipole density formalism. (*Coppens, 1997; Hansen and Coppens, 1978; Koritsanszky and Coppens, 2001*) Aspherical multipolar refinement is frequently used in experimental determination of accurate electron density maps at low temperatures, densities

that are routinely subjected to a topological analysis based on the quantum theory of atoms in molecules. (*Coppens, 1997; Koritsanszky and Coppens, 2001*) In the MOLLY formalism, the atomic density functions are given in terms of the polar coordinates **r**, θ, and φ: (*Hansen and Coppens, 1978*)

$$\rho_{atom} = P_{core}\rho_{core} + P_{valence}\kappa^3\rho_{valence}(\kappa r) + \sum_{l=0}^{l_{max}} \kappa'^3 R_l(\kappa' r)\sum_{m=0}^{l} P_{lm\pm} d_{lm\pm}(\theta,\varphi) \qquad (1.12)$$

where the P's refer to population coefficients satisfying the condition that $P_{core} + P_{valence} + \sum_{m=0}^{l} P_{lm\pm} = total$ *number of electrons* associated with the atom or ion in question, the ρ_{core} and $\rho_{valence}$ are chosen as the Hartree-Fock densities of the free atom or ion normalized to one electron, and R_l is an exponential radial function. The first two terms are the spherical core and valence densities, and the third term allows for valence density deformations. In the density modeling given in Eq. 1.12, the core density is unperturbed, the parameter κ scales the coordinates to provide radial contraction or expansion flexibility to the valence shell, the first summation in the third term is an additional monopole necessary to accurately describe the outer *s*-electron shell of transition metal atoms which can be more diffuse than the outermost *d*-shell, and $d_{lm\pm}$ are spherical harmonic functions which describe the aspherical features of the density. Substituting the expression for the multipolar aspherical atomic density (Eq.1.12) into Eq. 1.11 we obtain an expression for the aspherical multipolar atomic form factor which when used in the structure factor expression (Eq.1.9) gives the corresponding structure factor expression: (*Coppens, 1997; Hansen and Coppens, 1978; Koritsanszky and Coppens, 2001*)

$$F(hkl) = \sum_j \Big[P_{j,core} f_{j,core}(H) + P_{j,valence} f_{j,valence}(H/\kappa) + 4\pi \sum_{l=0}^{l_{max}} \sum_{m=0}^{l} P_{lm\pm} i^l \langle j_l \rangle d_{lm\pm}(\beta,\gamma) \Big] \exp[2\pi i(hx_j + ky_j + lz_j)]\, T_j(hkl) \qquad (1.13)$$

where $f_{j,core}$ and $f_{j,valence}$ are the Fourier transforms of $\rho_{j,core}$ and $\rho_{j,valence}$ respectively, $\langle j_l \rangle$ is the l^{th}-order Fourier-Bessel transform of the radial part of the multipole functions R_l, $d_{lm\pm}$ are spherical harmonics expressed in reciprocal space polar coordinates, and $T_j(hkl)$ are the temperature factors which measure

the extent to which every atom j oscillates around the position specified in the model due to thermal vibrations.

In the reciprocal space least-squares refinement, the parameters $P_{j,valence}$, P_{lm}, κ, and κ' are optimized in addition to the positions of the atoms in the unit cell (x_j,y_j,z_j) and the temperature factors $(T_j(hkl))$. The use of very low temperatures and this model of the density yields densities in excellent agreement with accurate theoretically determined densities. (*Coppens, 1997; Koritsanszky and Coppens, 2001*)

Now attention is turned to the main theoretical methods employed to obtain the electron density. A number of theoretical methods of increasing complexity exist but the following review will briefly outline the theoretical methods used in this thesis, the Hartree-Fock theory and the density functional theory, which are the most commonly used methods at the time of writing. This review is not and cannot be exhaustive.

1.4 The Electron Density from Theory

In this section, the theoretical methods used in this thesis are summarized. In the review presented here, emphasis will be placed on the conceptual and physical significance of the equations rather than on repeating detailed mathematical derivations which can be found in the literature. See for example (*Koch and Holthausen, 2001; Levine, 2000; Parr and Yang, 1989; Pilar, 1990; Szabo and Ostlund, 1989*)

1.4.1 Atomic Units (au)

The equations and the results presented in this thesis are frequently reported in atomic units (au). The system of atomic units is meant to simplify the equations of molecular and atomic quantum mechanics. The units in this system are combinations of the fundamental units of mass (mass of the electron), and charge (charge of the electron), and of Planck's constant. By setting these three quantities to unity one

can write the equations in a more compact form. For instance, in the *Système Internationale* (SI) units system, the electronic Schrödinger equation for the hydrogen atom is:

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 - \frac{e^2}{4\pi\varepsilon_0 r}\right]\Psi = E\,\Psi \tag{1.14}$$

This equation can be cast into a dimensionless form in which all the quantities are expressed in au:

$$\left[-\frac{1}{2}\nabla^2 - \frac{1}{r'}\right]\Psi' = E'\,\Psi' \tag{1.15}$$

The au system has a simple relation to other systems of units. For example: 1 au of length = a_0 (Bohr radius) = 5.29×10^{-9} cm = 0.529Å; 1 au of charge = e = 1.602×10^{-19}C = 4.803×10^{-10} esu; 1 au of charge density = e/a_0^3 = 6.748 eÅ^{-3} = 1.081×10^{12} Cm^{-3}; 1 au of Laplacian density = e/a_0^5 = 24.099 eÅ^{-5}; 1 au of dipole moment = ea_0 = 2.5418 debyes; 1 au of energy = e^2/a_0 = 627.51 kcal/mol = 2.6255×10^3 kJ/mol = 27.212 eV. For a formal discussion of how the au system of units naturally arises in quantum chemistry the reader is referred to (*Szabo and Ostlund, 1989*).

1.4.2 The Basis of Electronic Structure Calculations

The state of any quantum mechanical system which is stationary in time is described by the time-independent Schrödinger equation:

$$\hat{H}\Psi(\mathbf{x}_1,...,\mathbf{x}_N,\mathbf{R}_1,...,\mathbf{R}_M) = E\,\Psi(\mathbf{x}_1,...,\mathbf{x}_N,\mathbf{R}_1,...,\mathbf{R}_M) \tag{1.16}$$

where $\hat{H}$ is the Hamiltonian operator, $\mathbf{x}_i \equiv (\mathbf{r}_i,\sigma_i)$ the spatial and spin coordinates of the i^{th} electron, $\mathbf{R}_i$ the coordinates of the i^{th} nucleus, Ψ the total wavefunction, and E the energy of the system. In atomic units, and in absence of external electric or magnetic fields, the Hamiltonian operator is written:

$$\hat{H} \equiv -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{A=1}^{M}\frac{1}{M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}} \tag{1.17}$$

where A and B run over M nuclei and i and j refer to the N electrons, M_A the mass of nucleus A, Z_A refers to the nuclear charge of A, r_{iA} is the distance between electron i and nucleus A, r_{ij} the distance

between electron i and j, and R_{AB} the distance between nuclei A and B. In Cartesian coordinates, the Laplacian operator, defined by the identity (∇^2= div grad), is:

$$\nabla_q^2 \equiv \frac{\partial^2}{\partial x_q^2} + \frac{\partial^2}{\partial y_q^2} + \frac{\partial^2}{\partial z_q^2}. \tag{1.18}$$

The first two terms in the Hamiltonian (Eq.1.17) account for the kinetic energy of the electrons and nuclei respectively. The remaining terms describe the potential energy, specifically, the third term represents the attractive electrostatic interaction between the nuclei and the electrons, the fourth represents the electron-electron repulsive potential, and the last term represents the nuclear-nuclear repulsive potential. A Schrödinger equation (Eq.1.16) constructed with this Hamiltonian is extremely cumbersome to solve. The solution can be considerably simplified by taking advantage of the significant difference in the masses of the nuclei and of the electrons. This difference allows one to consider the motion of the electrons as much faster and independent of the motion of the nuclei. In other words, if the nuclei change their spatial coordinates, the electrons readjust their distribution instantaneously in response. This is the central physical idea behind the *Born-Oppenheimer (clamped-nuclei) approximation*. This approximation allows one to separate Eq. 1.16 into an electronic equation and a nuclear equation. The electronic equation is the one used in electronic structure calculation, and solutions of the nuclear equation lie at the heart of chemical dynamics. Thus, the Born-Oppenheimer approximation allows one to isolate an electronic Schrödinger equation:

$$\hat{H}_{electronic}\Psi_{electronic} \equiv \left[-\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} \right]\Psi_{electronic} = E_{electronic}\Psi_{electronic}. \tag{1.19}$$

The total energy of the system is then obtained by adding the nuclear-nuclear repulsion energy to the electronic energy obtained from Eq. 1.19:

$$E_{total} = E_{electronic} + E_{nuclear} = E_{electronic} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{R_{AB}}. \tag{1.20}$$

Since the coordinates of the nuclei do not appear explicitly in Eq. 1.19 the many-electron wavefunction $\Psi_{electronic} \equiv \Psi(\mathbf{x}_1,...,\mathbf{x}_N;\mathbf{R}_1,...,\mathbf{R}_M)$ and the total energy of the system are functions of the electronic

spatial and spin coordinates, yet depend *parametrically* on the nuclear coordinates (indicated by the semicolon instead of the comma separating the electronic from the nuclear coordinates and which appear in Eq. 1.16). Henceforth, the subscript "electronic" will be dropped unless the distinction is clearly needed from the context. If we calculate the total energy for a set of fixed nuclear positions, we obtain a *potential energy hypersurface*. It is worth noting at this point that the Hamiltonian is a spin-independent operator. The spin is introduced "by hand" into the symmetry property of the many electron wavefunction, since only in a relativistic treatment does spin arise naturally. (*Dirac, 1958*)

1.4.3 The Variational Principle

Powerful methods for obtaining the best approximate wavefunction of a system include: (i) variational methods, of which the Hartree-Fock theory is one, and (ii) perturbation theory methods which will not be discussed since they were not used in this thesis.

The *variation theorem* states: *the expectation value of the Hamiltonian computed using an approximate (normalizable) wavefunction will always be greater than the lowest eigenvalue (E_0) computed using the exact wavefunction.* The theorem is easily proved (*Smith and Absar, 1977*) and is expressed:

$$E[\Psi_{approx.}] = \frac{\int \Psi^*_{approx.} \hat{H} \Psi_{approx.} d\tau}{\int \Psi^*_{approx.} \Psi_{approx.} d\tau} \geq E_0[\Psi_0] \tag{1.21}$$

where Ψ_0 is the exact wavefunction.

The theorem sets the exact energy as a lower limit to be approached upon improvement of the approximate wavefunction during an electronic structure calculation. Thus, in variational methods, such as the Hartree-Fock method, one starts with a trial function which is a good guess to the exact wavefunction and proceeds iteratively to minimize the energy untill no further reduction in energy is possible within a preset numerical accuracy. It is then said that the calculation has converged.

1.4.4 The Electron Density from the Hartree-Fock Self-Consistent-Field Approximation

Attempts for the direct variational solution for the electronic Hamiltonian are known as *ab initio* methods because they involve no arbitrary parameters. Perhaps the most widely employed *ab initio* method is the Hartree-Fock method developed by Hartree, (*Hartree, 1928a; Hartree, 1928b*), Fock, (*Fock, 1930*) and Slater. (*Slater, 1930*) (See also: (*Blinder, 1965*) and (*Szabo and Ostlund, 1989*)). An approximation central to all molecular orbital based methods is the *orbital approximation* (or the *orbital model*) which is introduced to handle the many electron problem. In the orbital approximation, the complicated many-electron wavefunction ($\Psi(\mathbf{x}_1,\mathbf{x}_2,...\mathbf{x}_N)$) is expressed as a product of one-electron functions (orbitals, ($\psi(\mathbf{x}_i)$, where i=1,2...N)). In doing so, the N-electron (many-body) problem is simplified to N more tractable one-electron problems. A simple product of one-electron functions ($\psi(\mathbf{x}_1)\psi(\mathbf{x}_2)...\psi(\mathbf{x}_N)$) is known as the Hartree product. The Hartree product does not take electron indistinguishability into account, since it labels the electrons by assigning a specific electron to each orbital, nor does it take the Pauli antisymmetry principle (*vide infra*) into account. Indistinguishability and anti-symmetry are incorporatred in the Hartree-Fock self-consistent field (HF-SCF) molecular orbital method, a commonly used *ab initio* method, and the method used to obtain most of the densities analyzed in this thesis.

In the Hartree-Fock theory, the N-electron wavefunction is approximated by an antisymmetrized product of N one-electron wave functions ψ_i; each ψ_i is called a molecular spin orbital (MO). Each spin orbital is the product of a spatial orbital ϕ_i and a spin function which can be either α or β. Mathematically, this is expressed as:

$$\psi_i(\mathbf{x}) = \phi_i(\mathbf{r})\sigma_i(s) = \phi_i(x,y,z)\sigma_i(s), \tag{1.22}$$

where $\psi_i(\mathbf{x}_j)$ is the i^{th} molecular spin orbital, ϕ_i is the spatial one-electron function and σ_i is the function specifying its spin coordinate.

Since electrons are Fermions, an ensemble of electrons - such as one found in a many-electron atom or in a molecule - must obey the *Pauli antisymmetry principle*, which states that:

A *many-electron wavefunction must be antisymmetric to the interchange of the spatial and spin coordinates,* $\mathbf{x}\equiv(x,y,z,\sigma)$*, of any pair of electrons i and j, i.e.:*

$$\Psi(\mathbf{x}_1,\ldots,\underbrace{\mathbf{x}_i,\ldots,\mathbf{x}_j}_{\substack{\text{coordinates}\\ \text{interchange}}},\ldots\mathbf{x}_N) = -\Psi(\mathbf{x}_1,\ldots,\mathbf{x}_j,\ldots,\mathbf{x}_i,\ldots\mathbf{x}_N)\,. \tag{1.23}$$

An elegant way to incorporate the Pauli principle into the orbital model is to approximate the many-electron wavefunction by an *antisymmetrized sum of products* referred to as a *Slater determinant*, Ψ_{SD}:

$$\Psi_0 \approx \Psi_{SD} = \frac{1}{\sqrt{N!}}\begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_N(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_N(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \psi_1(\mathbf{x}_N) & \psi_2(\mathbf{x}_N) & \cdots & \psi_N(\mathbf{x}_N) \end{vmatrix}, \tag{1.24}$$

which is often written in a shorthand form which refers only to the diagonal elements:

$$\Psi_{SD} = \left|\psi_1(\mathbf{x}_1)\quad \psi_2(\mathbf{x}_2)\quad \cdots\quad \psi_N(\mathbf{x}_N)\right|. \tag{1.25}$$

The Slater determinant lets every electron occupy each one of the N spin orbitals thereby making the electrons indistinguishable in an antisymmetrized manner as required for Fermions. In Eq. 1.24, the pre-determinantal factor is a normalization factor and the spin orbitals ψ_i are chosen to form an orthonormal set (the spatial orbitals also form an orthonormal set amongst themselves). Orthonormality of spin- and spatial-orbitals is expressed:

$$\left.\begin{aligned} \langle\psi_i \mid \psi_j\rangle &= \int\psi_i^*\psi_j d\mathbf{x} = \delta_{ij} \\ \langle\phi_i \mid \phi_j\rangle &= \int\phi_i^*\phi_j d\mathbf{r} = \delta_{ij} \end{aligned}\right\} \tag{1.26}$$

where δ_{ij} is the Kronecker delta which equals 1 for $i=j$ and zero otherwise.

The physics embodied in the Slater determinant follows from known algebraic properties of determinants: (i) The interchange of any two rows of the Slater determinant changes the sign of the many-electron wave function in accordance with the requirement of the Pauli antisymmetry principle expressed in Eq. 1.23; (ii) If two rows or columns are equal, i.e. two or more electrons occupy the

same spin orbital, the determinant vanishes, in other words no more than one electron can occupy any spin orbital. The last point is an expression of the Pauli exclusion principle. The Slater determinant also embodies the Fermi correlation, i.e. the correlation of same spin electrons, as will become apparent in a later chapter dealing with the pair density (chapter 8).

An efficient numerical method to obtain a variational approximation to the solution of the Schrödinger equation for a molecule is to approximate the molecular orbitals themselves by a truncated linear combination of atomic basis functions χ_μ, an approach frequently called Linear Combination of Atomic Orbitals (LCAO). The larger the set of atomic basis functions the better the approximation of the molecular spin orbitals. The linear combination of basis functions is expressed:

$$\phi_i = \sum_{\mu=1}^{b} c_{i\mu}\chi_\mu, \tag{1.27}$$

where $c_{i\mu}$ are the expansion coefficients, and b is the number of basis functions (basis set size). A basis set with $b=\infty$ is necessary for an exact expansion of ϕ_i. Thus, the larger the basis set the more accurate is the expansion.

Now it is our task to obtain the best Slater determinant in the variational sense (Eq.1.21). The Hartree-Fock energy is given by:

$$E_{HF} = \left\langle \Psi_{SD} \mid \hat{H} \mid \Psi_{SD} \right\rangle + V_{nn}, \tag{1.28}$$

which, for a closed shell system with an even number of electrons (N), can be shown to yield: (*Szabo and Ostlund, 1989*)

$$\begin{aligned} E_{HF} = {} & 2\sum_{i=1}^{N/2} \int \phi_i^*(1)\left[-\frac{1}{2}\nabla^2 - \sum_A^M \frac{Z_A}{r_{1A}}\right]\phi_i(1)\, d\mathbf{r}_1 \\ & + \sum_{i=1}^{N/2}\sum_{j=1}^{N/2}\left\{2\left[\int\int |\phi_i(1)|^2 \left(\frac{1}{r_{12}}\right)|\phi_i(2)|^2\, d\mathbf{r}_1 d\mathbf{r}_2\right]\right. \\ & \left. - \left[\int\int \phi_i(1)\phi_j^*(1)\left(\frac{1}{r_{12}}\right)\phi_j(2)\phi_i^*(2)\, d\mathbf{r}_1 d\mathbf{r}_2\right]\right\} + V_{nn} \end{aligned} \tag{1.29a}$$

where the integrals are over space coordinates only, and V_{nn} is the nuclear-nuclear repulsion energy. Eq.1.29a can be re-written in a more compact form:

$$E_{HF} = 2\sum_{i=1}^{N/2} H_i + \sum_{i=1}^{N/2}\sum_{j=1}^{N/2}(2J_{ij} - K_{ij}) + V_{nn} \tag{1.29b}$$

where the two-electron integrals, J_{ij} and K_{ij}, are called the *Coulomb* and *exchange integrals* respectively. The one-electron integrals H_i represent the energy of a single electron in the molecular orbital ϕ_i in the field of the nuclei in the absence of the rest of the electrons. The integral H_i is multiplied by 2 to account for the two electrons populating the spatial molecular orbital ϕ_i. The Coulomb integrals represent the electrostatic repulsion energy between the "smoothed-out" electron density due to $\phi_i^*\phi_i$ and $\phi_j^*\phi_j$. (*Pople and Beveridge, 1970*) Thus, one replaces the instantaneous electron-electron repulsion and instead each electron moves in the average field generated by the remaining electrons. The exchange integrals (K_{ij}), which have no classical analogues, appear in the Hartree-Fock expressions as a result of the *exchange correlation* which accounts for the correlation of the motions of electrons of parallel spins. The exchange term corrects for the self-repulsion of the electrons since $J_{kk} = J_{(self)} = K_{kk} = K_{(self)}$ while J_{ij} and K_{ij} appear in Eq. 1.29 with opposite signs.

The target of the Hartree-Fock method is to find the set of orbitals ϕ_i which minimize Eq. 1.29 subject to the orthonormalization constraint (Eq. 1.26). For a closed shell system these Hartree-Fock orbitals are the eigenfunctions of the Fock operator $\hat{F}$, while the eigenvalues are the orbital energies ϵ_i, such that: (*Szabo and Ostlund, 1989*)

$$\hat{F}(1)\phi_i(1) = \varepsilon_i \phi_i(1), \tag{1.30}$$

where

$$\hat{F}(1) = \hat{H}(1) + \sum_{j=1}^{N/2}\left[2\hat{J}_j(1) - \hat{K}_j(1)\right], \tag{1.31}$$

$$\hat{H}(1) \equiv -\frac{1}{2}\nabla^2 - \sum_{A}^{M}\frac{Z_A}{r_{1A}}, \tag{1.32}$$

$$\hat{J}_j(1)\phi_i(1) = \left[\int \phi_j^*(2)\left(\frac{1}{r_{12}}\right)\phi_j(2)\, d\mathbf{r}_2\right]\phi_i(1), \tag{1.33}$$

$$\hat{K}_j(1)\phi_i(1) = \left[\int \phi_j^*(2)\left(\frac{1}{r_{12}}\right)\phi_i(2)\,d\mathbf{r}_2\right]\phi_j(1), \tag{1.34}$$

where the integrals are over all space. $\hat{J}_j$ and $\hat{K}_j$ are termed the Coulomb and exchange operators respectively.

Substitution of the expansion Eq. 1.27 into the Hartree-Fock equations 1.29 gives:

$$\sum_{\mu=1}^{b} c_{i\mu}\hat{F}\chi_\mu = \varepsilon_i \sum_{\mu=1}^{b} c_{i\mu}\chi_\mu, \tag{1.35}$$

which upon multiplication by χ_ν and integration yields the Roothaan equations: (*Roothaan, 1951*)

$$\sum_{\mu=1}^{b} c_{i\mu}(F_{\nu\mu} - \varepsilon_i S_{\nu\mu}) = 0, \qquad \nu = 1,2,\ldots,b \tag{1.36}$$

where

$$F_{\nu\mu} \equiv \langle \chi_\nu \,|\, \hat{F} \,|\, \chi_\mu \rangle, \qquad S_{\nu\mu} \equiv \langle \chi_\nu \,|\, \chi_\mu \rangle. \tag{1.37}$$

The condition to obtain a nontrivial solution of the set of b simultaneous linear homogeneous equations in b unknowns ($c_{i\mu}$, $\mu = 1,2,\ldots,b$) is:

$$\det(F_{\nu\mu} - \varepsilon_i S_{\nu\mu}) = 0. \tag{1.38}$$

Eq.1.38 is the secular equation the roots of which are the orbital energies ϵ_i.

Because of the presence of the Coulomb and exchange operators which appear in the Fock operator, the Hartree-Fock equations are coupled (since the potential in the Fock operator is defined by the orbitals we seek to calculate). Such coupled integro-differential equations must be solved iteratively. This procedure is known as the *self-consistent field* (SCF) method, which means that the calculation converges when the calculated orbitals generate a field that yields the same orbitals to within an allowed numerical error. The steps in a typical self-consistent field Hartree-Fock calculation are outlined in the following paragraph.

The calculation starts with a guess for the occupied molecular orbitals (MO) expressed as a linear combination of the basis functions specified in the basis set (Eq.1.27). The initial guess for the

i^{th} MO consists of a guess of the expansion coefficients ($c_{i\mu}$'s) of this MO in terms of the basis functions. The Fock operator is constructed using this initial guess of the occupied MOs (Eqs.1.31-1.34). The Fock operator is then used to compute the matrix elements (Eq.1.37) which are substituted into the secular determinant to obtain an initial set of orbital energies (ϵ_i's). The ϵ_i's are used in Eq. 1.36 to obtain an improved set of expansion coefficients. The new sets of coefficients define an improved set of MOs which is used to construct an improved Fock operator, and so on. The process is repeated until no further improvement is possible, i.e. the set of MOs used to construct the Fock operator is the same as the one obtained from the solution of the secular equations. When the calculation reaches self-consistency, it is said to have converged. With the resulting orbitals, the calculation of the electron density is trivial:

$$\rho(\mathbf{r}) = \int d\tau' \, \Psi_{SD}^{*} \Psi_{SD} = 2\sum_{j=1}^{N/2} \phi_j^{*}\phi_j = 2\sum_{\nu=1}^{b}\sum_{\mu=1}^{b}\sum_{j=1}^{N/2} c_{\nu j} c_{\mu j}\, \chi_\nu^{*}\chi_\mu = \sum_{\nu=1}^{b}\sum_{\mu=1}^{b} P_{\nu\mu}\chi_\nu^{*}\chi_\mu, \qquad (1.39)$$

where $P_{\nu\mu} = 2\sum_{j=1}^{N/2} c_{\nu j} c_{\mu j}$ are the elements of the *density matrix* (**P**) for a closed shell molecule and Ψ_{SD} is defined in Eq. 1.23. From Eq. 1.39 it is clear that **P** specifies completely the electron density and, indeed, the entire physical situation in the Hartree-Fock scheme. The matrix ½**P** is said to be idempotent, i.e. $(½\mathbf{P})^2 = ½\mathbf{P}$. Detailed discussions about the properties and uses of the density matrix can be found elsewhere.[(*Coleman and Yukalov, 2000; Davidson, 1976*), and references therein].

A single determinant, while completely accounts for the Fermi correlation between same-spin electrons, neglects the Coulombic correlation between opposite-spin electrons. (*McWeeny, 1967*) As a result, the Hartree-Fock energy (E_{HF}) is always greater than the exact non relativistic ground state energy (E_0). The difference between E_0 and E_{HF} is termed the correlation energy (E_C^{HF}):

$$E_C^{HF} = E_0 - E_{HF}. \qquad (1.40)$$

Several schemes designed to incorporate Coulombic correlation present significant improvements beyond the Hartree-Fock limit. Amongst the most rigorous, for example, are the *configuration interaction* (CI) methods. In this type of calculation, an initial SCF calculation yields a set of occupied

and virtual MOs. The MOs are used to form different states (configurations) of the system, e.g. ground state ($^0\Psi_{SD}$), singly excited ($^1\Psi_{SD}$), doubly excited ($^2\Psi_{SD}$), etc., each state being represented by a single Slater determinant (1.23). The CI wavefunction is then written as a linear combination of the configuration functions:

$$\Psi_{CI} = C_0\ {}^0\Psi_{SD} + C_1\ {}^1\Psi_{SD} + C_2\ {}^2\Psi_{SD} + \ldots \tag{1.41}$$

where the C_i's are unknown expansion coefficients that are to be determined variationally by energy minimization (Eq.1.21). It is to be noted that there generally are several configurations of each type, and thus each coefficient/configuration – except C_0 and $^0\Psi$ – is in fact a set of coefficients and configurations that are included in the expansion. A common choice, CI with singles and doubles (CISD), introduces in the sum (Eq.1.41) all the configurations that differ from the ground-state by a single- and double-excitation. The CISD method is an example of a truncated CI. In contrast, a full CI calculation includes all possible configurations of proper symmetry. The main problem with CI calculations is their prohibitive costs.

The Effect of Coulombic Correlation on the Electron Density

A study of the effect of the inclusion of Coulombic correlation on the atomic and bond properties derived from the electron density using the quantum theory of atoms in molecules shows that these effects are rather small. (*Gatti et al., 1988*) In this study, Gatti *et al.* (*Gatti et al., 1988*) studied several carbenes and small hydrocarbon molecules at RHF as well as at the configuration-interaction (CI) level. They found that the correlated densities have the same number and kind of critical points in both the $\rho(\mathbf{r})$ and the $\nabla^2\rho(\mathbf{r})$ fields as are in the corresponding non-correlated densities. In other words, the topology of the density is unaffected by the inclusion of Coulombic correlation. Moreover, the differences in the integrated atomic properties and critical points' properties between the correlated and the corresponding non-correlated densities are found to be small. Thus, fortunately, for most chemical applications, qualitative trends are unaffected by Coulombic correlation.

The Hartree-Fock method yields densities that are sufficiently accurate for most chemical applications as a result of an important property of the SCF Hartree-Fock deteminant, that its matrix elements with any of the determinants produced by a single excitation vanish, i.e. $\left\langle {}^{0}\Psi_{SD} \middle| \hat{H} \middle| {}^{1}\Psi_{SD} \right\rangle = 0$, a property known as *Brillouin's theorem.* (*Brillouin, 1934*) It is because of Brillouin's theorem that the Hartree-Fock method yields all one-electron properties (such as the density or the dipole moment) correct to the second-order since the first-order corrections vanish.

1.4.5 The Electron Density from Density Functional Theory (DFT)

The incorporation of Coulombic correlation effects in a calculated density is computationally costly in terms of resources and time, rendering such a calculation on large molecules of chemical or biochemical interest impossible at the time of writing. In an attempt to circumvent this bottleneck, Hohenberg and Kohn (*Hohenberg and Kohn, 1964*) developed Density Functional Theory (DFT) which was translated into a practical computational scheme by Kohn and Sham (*Kohn and Sham, 1965*). Today, the theory is a powerful tool in the arsenal of the computational chemist who wants to compute correlated yet relatively inexpensive densities of ground states of molecules and crystals. DFT has the merit of rendering such a calculation possible for much larger systems than any other correlated method to date, its main advantage being computational economy and speed. In this section an outline of the theory is presented. For in-depth reviews see for example (*Koch and Holthausen, 2001; Parr and Yang, 1989*).

The idea of expressing one-electron ground-state properties in terms of the density $\rho(\mathbf{r})$ was put forward by Thomas (*Thomas, 1927*) and Fermi (*Fermi, 1928*) quite a long time ago. These workers introduced a variational procedure to determine the ground-state $\rho(\mathbf{r})$ without calculating the many-electron wavefunctions $\Psi(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_N)$. Indeed the motivation of DFT is to use ρ, which is a much simpler 3-dimensional function of space than the cumbersome multidimensional Ψ, as the basic variable defining the ground-state. The price one has to pay for this simplification is that only one-

electron properties which are described by multiplicative operators are accessible. Any property that depends on the pair density cannot be derived from ρ; such properties require the information contained only in the many-body wavefunction but which is lost in the density. The Thomas-Fermi theory did not account for exchange and correlation effects and suffered from poor treatment of the kinetic energy and was abandoned as it failed to reproduce experimental results.

It was not until the sixties that modern DFT was revived with the injection of new ideas. Hohenberg and Kohn presented an elegant *reductio ad absurdum* proof that the ground state electron density $\rho_0(\mathbf{r})$ determines the external potential (the second term in Eq. 1.19, and defined in Eq. 1.42) to within an additive constant and also determines the number of electrons in the system, i.e. $\rho_0(\mathbf{r})$ determines the Hamiltonian which in turn determines the ground-state wavefunction, energy, and other properties. (*Hohenberg and Kohn, 1964*) This mapping between the density and the rest of the properties of the system is known as the *first Hohenberg-Kohn theorem.*

The external potential operator acting on electron i is defined:

$$\hat{V}(\mathbf{r}_i) = -\sum_{A=1}^{M} \frac{Z_A}{r_{iA}}. \tag{1.42}$$

The external potential is so called because it is external to the system of electrons. From the electronic Hamiltonian in Eq. 1.19, the total electronic energy is given as the sum of three terms which are in sequence: the total electronic kinetic energy, the total nuclear-electronic potential energy, and the total electron-electron repulsion energy. This can be expressed as:

$$E_0 = E_V[\rho_0] = \langle T_e[\rho_0] \rangle + \langle V_{ne}[\rho_0] \rangle + \langle V_{ee}[\rho_0] \rangle \tag{1.43}$$

where the angular brackets indicate averages and the square brackets indicate a functional dependence. A functional $F[f(x)]$ is a rule assigning a number to each function $f(x)$. An example of a functional is the action $S[L]$, which is the time-integral of the Lagrangian, since for every path in time there corresponds a single value for the action, the actual path adopted by a classical particle being the one minimizing the action. From Eqs.1.19 and 1.42, one can write:

$$\hat{V}_{ne} = \sum_{i=1}^{N} \hat{V}(\mathbf{r}_i) = -\sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{r_{iA}} \tag{1.44}$$

To obtain the expectation value of the nuclear-electron attraction potential energy operator, one can sandwich it between the ground-state wavefunction and its complex conjugate: (*Levine, 2000*)

$$\langle V_{ne} \rangle = \left\langle \psi_0 \left| \sum_{i=1}^{N} \hat{V}(\mathbf{r}_i) \right| \psi_0 \right\rangle = \int \psi_0^* \sum_{i=1}^{N} \hat{V}(\mathbf{r}_i) \psi_0 \, d\mathbf{r} = \int \rho_0(\mathbf{r}) \, V(\mathbf{r}) \, d\mathbf{r} = \langle V_{ne}[\rho_0] \rangle, \tag{1.45}$$

where $V(\mathbf{r})$ is the nuclear attraction potential energy function for an electron at $\mathbf{r}$. Eq. 1.45 is an explicit expression of the functional $\langle V_{ne}[\rho_0] \rangle$ which appears on the right hand side of Eq. 1.43, however, the two other functionals remain unknowns. Rewriting Eq. 1.43 by substituting the known functional of the nuclear-electronic potential energy gives:

$$E_0 = E_V[\rho_0] = \int \rho_0(\mathbf{r}) \, V(\mathbf{r}) \, d\mathbf{r} + \langle T_e[\rho_0] \rangle + \langle V_{ee}[\rho_0] \rangle = \int \rho_0(\mathbf{r}) \, V(\mathbf{r}) \, d\mathbf{r} + F[\rho_0], \tag{1.46}$$

where in the last step we "dumped" the two unknown functionals into a functional $F[\rho_0]$ which is independent of the external potential. Eq. 1.46 is not useful as such since it contains this unknown functional.

Hohenberg and Kohn proved a second theorem which states that $E_V[\rho_0]$ is minimized by the ground-state density; i.e. $E_V[\rho_{\text{trial}}] \geq E_V[\rho_0]$. (For a proof, see (*Levy, 1979*)). In order to implement in a computational scheme this variational theorem, Kohn and Sham (KS) started by considering a fictitious system consisting of N non-interacting electrons having the same density as the real fully interacting system, that is, $\rho_s = \rho_0$ where the subscript s stands for non-interacting system. The following development summarizes the one provided in (*Levine, 2000*). Since the electrons are non-interacting, the Hamiltonian of this fictitious system $\hat{H}_s$ is equal to the sum of one-electron KS Hamiltonians $\hat{h}_i^{KS}$ such that:

$$\hat{H}_s = \sum_{i=1}^{N} \left[-\frac{1}{2} \nabla_i^2 + V_s(\mathbf{r}_i) \right] \equiv \sum_{i=1}^{N} \hat{h}_i^{KS}. \tag{1.47}$$

Since the system consists of non-interacting Fermions, the corresponding ground state wavefunction $\Psi_{s,0}$ is a Slater determinant (an antisymmetrized product) of the occupied KS spin-orbitals u_i^{KS}, where the spatial parts θ_i^{KS} are eigenfunctions of the one-electron KS Hamiltonians $\hat{h}_i^{KS}$. In other words, one can write:

$$\hat{h}_i^{KS}\theta_i^{KS} = \varepsilon_i^{KS}\theta_i^{KS} \tag{1.48}$$

where the ε_i^{KS} are KS orbital energies.

In order to express $E_V[\rho_0]$ in Eq. 1.46 in terms of quantities that are easy to calculate, Kohn and Sham made the following substitutions/definitions:

$$\langle \Delta T[\rho] \rangle \equiv \langle T[\rho] \rangle - \langle T_s[\rho] \rangle , \tag{1.49}$$

$$\langle \Delta V_{ee}[\rho] \rangle \equiv \langle V_{ee}[\rho] \rangle - \frac{1}{2}\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 , \tag{1.50}$$

where the subscript 0 has been dropped from the density, $\langle \Delta T \rangle$ is the difference between the average ground-state electronic kinetic energy between the real system and the non-interacting system, r_{12} is the distance between the points $\mathbf{r}_1$ and $\mathbf{r}_2$. The second term in Eq. 1.50 represents the *average* electrostatic repulsion energy between the electrons in the system (excluding Coulombic correlation as well as exchange), and $\langle \Delta V_{ee} \rangle$ is the difference between the correlated and uncorrelated electrostatic repulsion energies. The second term in Eq. 1.50 is divided by two to avoid counting the same interaction twice. Since now the unknown quantities are the two functionals $\langle \Delta T \rangle$ and $\langle \Delta V_{ee} \rangle$, they are collected together into the *exchange-correlation functional* $E_{xc}[\rho]$:

$$E_{xc}[\rho] \equiv \langle \Delta T[\rho] \rangle + \langle \Delta V_{ee}[\rho] \rangle . \tag{1.51}$$

Substituting the exchange-correlation functional into Eq. 1.46 yields:

$$E_0 = E_V[\rho] = \int \rho(\mathbf{r})\, V(\mathbf{r})\, d\mathbf{r} + \langle T_s[\rho] \rangle + \frac{1}{2}\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + E_{xc}[\rho], \tag{1.52}$$

Which can be written in a more explicit form:

$$E_0 = -\sum_{A=1}^{M} Z_A \int \frac{\rho(\mathbf{r}_1)}{r_{1A}}\, d\mathbf{r}_1 - \frac{1}{2}\sum_{i=1}^{N}\left\langle \theta_i^{KS}(1) \left| \nabla_1^2 \right| \theta_i^{KS}(1) \right\rangle + \frac{1}{2}\int\int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + E_{xc}[\rho]. \qquad (1.53)$$

The only quantities to be determined in Eq. 1.53 are the KS orbitals θ_i^{KS} and the exchange-correlation functional. Now we are ready to use the second Hohenberg-Kohn theorem to obtain the orbitals variationally. Since the density is expressed as a sum of the individual orbital densities:

$$\rho = \sum_{i=1}^{N} \left| \theta_i^{KS} \right|^2, \qquad (1.54)$$

the variation of the density to minimize the energy can be conveniently effected through the variation of the KS orbitals subject to the constraints: (i) that the integral of the density over all space always yields the total number of electrons in the system (N), and (ii) that the KS orbitals remain orthonormals. It can be shown (*Parr and Yang, 1989*) that the KS orbitals that minimize Eq. 1.53 satisfy the set of Kohn-Sham equations:

$$\left[-\frac{1}{2}\nabla_1^2 - \sum_{A=1}^{M} \frac{Z_A}{r_{1A}} + \int \frac{\rho(\mathbf{r}_2)}{r_{12}}\, d\mathbf{r}_2 + v_{xc}(1) \right] \theta_i^{KS}(1) = \varepsilon_i^{KS}\, \theta_i^{KS}(1), \qquad (1.55a)$$

which are sometimes written:

$$\hat{h}^{KS}\, \theta_i^{KS}(1) = \varepsilon_i^{KS}\, \theta_i^{KS}(1), \qquad (1.55b)$$

where the quantity in the square brackets is known as the effective KS Hamiltonian ($\hat{h}^{KS}$) which contains the *exchange-correlation potential* ($v_{xc}(\mathbf{r})$) defined as the functional derivative of the exchange-correlation energy E_{xc}:

$$v_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}. \qquad (1.56)$$

Thus, if the functional $E_{xc}[\rho]$ is known, its functional derivative is readily found, and so is the exchange correlation potential and the solution of the KS equations can proceed iteratively as seen in the section on Hartree-Fock theory. Three differences between Kohn-Sham theory and Hartree-Fock theory are worth noting at this point. (*Doucet and Weber, 1996; Koch and Holthausen, 2001*) Firstly, the KS Hamiltonian in Eq. 1.55 is a local operator since it is fully determined, in principle, from a

knowledge of the electron density. The "locality" of the KS Hamiltonian is contrasted with the non-local nature of the Fock operator due to the exchange operator (Eq.1.34). Secondly, the KS Hamiltonian incorporates the Coulombic correlation through v_{xc}, a correlation which is ignored in the Hartree-Fock theory. Thirdly, the KS scheme bundles all non-classical contributions (exchange and correlation) into the exchange-correlation potential v_{xc} which is approximated since the functional E_{xc} is unknown. In contrast, the Hartree-Fock scheme treats the exchange contribution exactly (Eq.1.34) but ignores correlation.

The challenge of DFT is to obtain an expression for the exchange-correlation functional, and it is the quality of this functional that determines the quality of the DFT calculation. The solution of Eq. 1.55 consists of a set of KS molecular orbitals which yield the density of the ground state associated with chosen v_{xc}. If one knew the exact E_{xc} and consequently the exact v_{xc} one would obtain the exact ground-state density. It is one of the main drawbacks of DFT that there is no systematic method of improving the E_{xc} functional except by empirical comparison of calculated properties with experiment. The simplest functional introduced by Kohn and Sham was proposed for systems where the density changes smoothly, the *local density approximation* (LDA): (*Kohn and Sham, 1965*)

$$E_{xc}^{LDA}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\, \varepsilon_{xc}(\rho)\, d\mathbf{r} \tag{1.57}$$

where $\varepsilon_{xc}(\rho)$ is the exchange plus correlation energy per electron in a homogeneous electron gas having the density ρ. The integrand in Eq. 1.57 is a functional of only ρ but does not reflect how ρ changes with respect to its surroundings. The LDA will not be discussed further since it is a crude approximation in molecular systems where the densities frequently have substantial curvatures. The next significant improvement over LDA is the *gradient corrected functionals*. In these functionals, sometimes referred to as *generalized gradient approximation* (GGA), the gradient of the density is explicitly included in the integrand:

$$E_{xc}^{GGA}[\rho^{\alpha}(\mathbf{r}), \rho^{\beta}(\mathbf{r})] = \int f(\rho^{\alpha}(\mathbf{r}), \rho^{\beta}(\mathbf{r}), \nabla\rho^{\alpha}(\mathbf{r}), \nabla\rho^{\beta}(\mathbf{r}))\, d\mathbf{r} \tag{1.58}$$

and the E_{xc}^{GGA} is split into an exchange contribution and a correlation contribution, each contribution being modeled separately:

$$E_{xc}^{GGA} = E_{x}^{GGA} + E_{c}^{GGA}. \tag{1.59}$$

Thus, any exchange functional can be added to any correlation functional. For example, a DFT calculation done with the Becke 1988 exchange functional and the Lee-Yang-Parr correlation functional using KS orbitals expanded in a 6-31G* basis set is denoted as BLYP/6-31G*. (Example from (*Levine, 2000*)). The last type of functional that will be mentioned here is the *hybrid functional.* In hybrid functionals, such as the B3LYP, the exchange functional and the correlation functional are both expressed as a weighted sum of different functionals. The weights are adjusted empirically to yield good fit to experimental molecular atomization energies. (See (*Levine, 2000*)).

In the Hartree-Fock scheme, one needs to start with a guess for the orbitals since they enter in the definition of the Fock operator, a situation that can only be solved iteratively. Similarly, a glance at Eq. 1.55 reveals that the operator contains the very density we seek to calculate. Thus, having selected the functional, a DFT calculation starts with an initial guess for the density. The initial guess for ρ is frequently obtained by simple superposition of spherical atomic densities at the desired molecular geometry, a promolecule approach not dissimilar to the one discussed in x-ray crystallography (Section 1.3). The initial guess of ρ is used to obtain an initial guess for $E_{xc}[\rho]$ (using Eq. 1.58 for example) from which the initial guess for $v_{xc}(\mathbf{r})$ is obtained using Eq. 1.56. The initial $v_{xc}(\mathbf{r})$ is then used in the KS equations (Eq.1.55) which are solved to obtain an initial estimate of the KS orbitals. Since the KS orbitals are expanded in a basis set, i.e.

$$\theta_i^{KS} = \sum_{\mu=1}^{b} c_{i\mu}\, \chi_\mu, \tag{1.60}$$

which is very similar to Eq. 1.23, it is not surprising that the solution proceeds from here in a manner similar to the solution of the Hartree-Fock-Roothan equations (Eqs.1.36,1.37):

$$\sum_{\mu=1}^{b} c_{i\mu}(h_{\nu\mu}^{KS} - \varepsilon_i^{KS} S_{\nu\mu}) = 0, \qquad \nu = 1,2,...,b \tag{1.61}$$

where

$$h_{\nu\mu}^{KS} \equiv \left\langle \chi_\nu \left| \hat{h}^{KS} \right| \chi_\mu \right\rangle, \qquad S_{\nu\mu} \equiv \left\langle \chi_\nu \mid \chi_\mu \right\rangle. \tag{1.62}$$

The initial set of KS orbitals obtained from the above procedure is used to obtain an improved density from Eq. 1.54 which is used to obtain an improved v_{xc} which is used in the KS equations to obtain improved KS orbitals, and the process is repeated until convergence.

Since the KS orbitals are constructed for a fictitious non-interacting system, a KS "many-electron wavefunction" is not a Slater determinant and thus not a wavefunction. However, practice has shown that usually the occupied KS orbitals resemble Hartree-Fock molecular orbitals. (*Levine, 2000*)

1.5 Closing Remarks

The electron density is a scalar field filling the real three-dimensional space which is readily accessible from experiment as well as from theory. This field is what a theory of chemistry should be built upon. Having seen in this chapter how the density can be obtained from experiment and from theory, we are left with questions about how to use it. For example:

(1) How can one decode and summarize the wealth of information contained in the density?

(2) How can one account for the empirical observation that atoms and functional groupings of atoms exhibit characteristic properties in different molecules (with very different external potentials) provided their immediate surroundings are similar?

(3) How can one assign a bonded chemical structure from the density? Is there a unique answer?

(4) How can one predict reactivity from the density?

These and other questions will be answered in the following chapter in the context of a wider question posed by Bader: (*Bader, 1990*)

Are there atoms in molecules?

Chapter 2

The Quantum Theory of Atoms in Molecules

> "Or, to take Aristotle's examples, if a man makes a bronze sphere, bronze is the matter, and sphericity is the form; while in the case of a calm sea, water is the matter and smoothness is the form. So far, all is simple.
>
> He goes on to say that it is in virtue of the form that the matter is some one definite thing, and this is the substance of the thing. *What Aristotle means seems to be plain common sense: a "thing" must be bounded, and the boundary constitutes its form.* Take, say, a volume of water: any part of it *can* be marked off from the rest by being enclosed in a vessel, and then this part becomes a "thing", but so long as the part is in no way marked out from the rest of the homogenous mass it is not a "thing" . . . *We* should not naturally say that it is the form that confers substantiality, but that is because the atomic hypothesis is ingrained in our imagination. *Each atom, however, if it is a "thing", is so in virtue of its being delimited from other atoms, and so having, in some sense, a "form".*"
> (Second and last sentences of second paragraph italicized by the present author).
>
> Bertrand Russell [(*Russell, 1945*), p.165]

2.1 The Topology of the Electron Density and the Partitioning of a Molecule into Bounded Atoms

Pure thinking alone led one of the most influential of contemporary philosophers, Bertrand Russell, to agree with his ancient Greek counterpart, Aristotle, that if an atom is to be a "thing" it has to be *bounded*. It is with amazement that the author realizes how these two great minds – separated by over two millennia – have come so close to a "common sense" concept which was rigorously formulated only as late as the seventies with the advent of the quantum theory of atoms in molecules. It is the topology of a physical observable, the electron density, that determines the boundaries of an atom which in turn determine its shape which in turn determines its properties inside a molecule. Without

boundaries, the very idea of the atom as the basic building block of matter, its very "substantiality", dissipates into thin air. In other words, according to common sense, *an atom is a real space-filling object with definite shape defined by its bounding surface.* Such atoms, being bounded by a surface, do not overlap one another.

But what precisely is the boundary of an atom? Indeed what exactly is an atom in a molecule or in an extended system? To begin answering these questions, let's start at the beginning: the topology of the electron density. This chapter is a selective review of only those aspects of the quantum theory of atoms in molecules (QT-AIM) which are most relevant to this thesis. A complete and authoritative account on QT-AIM by its principal investigator can be found elsewhere. (*Bader, 1990*). Introductions to QT-AIM have also appeared in the recent literature (*Gillespie and Popelier, 2001; Matta and Gillespie, 2002; Popelier, 2000*)

The topology of the electron density is dominated by the attractive force exerted on it by the nuclei, a force that endows the density with it principal topological feature – that it exhibits a maximum value at the position of each nucleus, a feature that is true for any plane containing a nucleus. Figure 2.1(a and b) displays two representations (a relief map and a contour plot) of the electron density in the molecular plane of a guanine-cytosine DNA base pair. A consequence of this topological feature of the density is the association of an atom with a region of space, each region being dominated by a given nucleus, with boundaries evident in the minima that exist between the nuclear maxima. The boundaries are determined by the balance in the forces that the neighbouring nuclei exert on the density.

Maxima in the electron density are one type of *critical point* (cp), nuclear cp, that one encounters in the density. A critical point is a point in the three-dimensional density field at which the first derivatives of the density vanish, i.e.:

$$\nabla\rho = \mathbf{i}\frac{d\rho}{dx} + \mathbf{j}\frac{d\rho}{dy} + \mathbf{k}\frac{d\rho}{dz} = \vec{\mathbf{0}} \tag{2.1}$$

where the zero vector signifies that each individual derivative in the gradient operator is zero and not just their sum. From elementary calculus, one can discriminate between a local minimum, a local maximum or a saddle point only by considering the second derivative. Generally, there are nine second derivatives of $\rho(\mathbf{r})$ that can be arranged in the so called *Hessian matrix*, which when evaluated at the critical point $\mathbf{r}_c$ and denoted by $\mathbf{A}(\mathbf{r}_c)$ is written:

$$\mathbf{A}(\mathbf{r}_c) = \left(\begin{array}{ccc} \frac{\partial^2 \rho}{\partial x^2} & \frac{\partial^2 \rho}{\partial x \partial y} & \frac{\partial^2 \rho}{\partial x \partial z} \\ \frac{\partial^2 \rho}{\partial y \partial x} & \frac{\partial^2 \rho}{\partial y^2} & \frac{\partial^2 \rho}{\partial y \partial z} \\ \frac{\partial^2 \rho}{\partial z \partial x} & \frac{\partial^2 \rho}{\partial z \partial y} & \frac{\partial^2 \rho}{\partial z^2} \end{array} \right)_{\mathbf{r}=\mathbf{r}_c} . \tag{2.2}$$

The Hessian matrix can be diagonalized since it is real and symmetric. The diagonalization of $\mathbf{A}(\mathbf{r}_c)$ is equivalent to a rotation of the coordinate system $\mathbf{r}(x,y,z) \rightarrow \mathbf{r}(x',y',z')$ which aligns the new axes x', y', z' with the principal curvature axes of the critical point. The diagonalized Hessian is denoted by $\mathbf{\Lambda}$, and is given by:

$$\mathbf{\Lambda} = \left(\begin{array}{ccc} \frac{\partial^2 \rho}{\partial x'^2} & 0 & 0 \\ 0 & \frac{\partial^2 \rho}{\partial y'^2} & 0 \\ 0 & 0 & \frac{\partial^2 \rho}{\partial z'^2} \end{array} \right)_{\mathbf{r}'=\mathbf{r}_c} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}, \tag{2.3}$$

in which λ_1, λ_2, and λ_3, are the curvatures of the density with respect to the three principal axes x', y', z'. An important property of the Hessian is that it's trace is invariant to rotations of the coordinate system. The trace of the Hessian of the density is known as the Laplacian of the density ($\nabla^2\rho(\mathbf{r})$) and is given by:

$$\nabla^2 \rho(\mathbf{r}) = \nabla \cdot \nabla \rho(\mathbf{r}) = \frac{\partial^2 \rho(\mathbf{r})}{\partial x^2} + \frac{\partial^2 \rho(\mathbf{r})}{\partial y^2} + \frac{\partial^2 \rho(\mathbf{r})}{\partial z^2}. \tag{2.4}$$

The critical points in the electron density are classified according to their *rank* and *signature*. The rank, denoted by ω, is the number of non-zero curvatures of ρ at the critical point. A critical point that

has $\omega<3$ is mathematically unstable, and any perturbation of the density caused by nuclear motion causes it to bifurcate or vanish. The presence of a cp with a rank less than three is indicative of a change in the topology of the density and hence of a change in the molecular structure. As a result, critical points with $\omega<3$ are, in general, not found in equilibrium charge distribution and one finds $\omega=3$ nearly always. The signature, denoted by σ, is the algebraic sum of the signs of the curvatures, i.e. each one of the three curvatures contribute ±1 depending on whether it is a positive or negative curvature respectively. There are four possible types of stable critical points having three non-zero eigenvalues:

(3, -3) Three negative curvatures: ρ is a local maximum.

(3, -1) Two negative curvatures: ρ is a maximum in the plane defined by the corresponding eigenvectors but is a minimum along the third axis which is perpendicular to this plane.

(3, +1) Two positive curvatures: ρ is a minimum in the plane defined by the corresponding eigenvectors and a maximum along the third axis which is perpendicular to this plane.

(3, +3) Three curvatures are positive: ρ is a local minimum.

We are going to identify each type of critical point described above with an element of chemical structure: (3,-3) *nuclear critical point* (ncp); (3,-1) *Bond critical point* (bcp); (3,+1) *ring critical point* (rcp); and (3,+3) *cage critical point* (ccp). The number and type of critical point that can coexist in a molecule follow a strict topological relationship known as the *Poincaré-Hopf relationship* (PH) which states that: (*Bader, 1990*)

$$number\ of\ ncp\ -\ number\ of\ bcp\ +\ number\ of\ rcp\ -\ number\ of\ ccp\ =\ 1. \qquad (2.5)$$

For an infinite lattice the above equation is replaced by the Morse equation in which the equality is changed so that the right hand side is equal to zero. (*Coppens, 1997*)

In physics, there are two important types of fields in three-dimensional space: (i) *Scalar fields* which specify a scalar – *a magnitude* – at every point in space. Examples of scalar fields include the temperature in a room or the electron density $\rho(\mathbf{r})$ in a molecule or in a crystal; (ii) Vector fields which

specify a vector – *both a magnitude and a direction* – at every point in space, examples of vector fields include the magnetic field around a magnet or the gradient vector field of the electron density ($\nabla\rho(\mathbf{r})$) in a molecule or in a crystal.

The gradient ∇ of a scalar function such as $\rho(\mathbf{r})$ (Eq.2.1) at a point in space is a vector pointing in the direction in which $\rho(\mathbf{r})$ undergoes the greatest rate of increase and having a magnitude equal to the rate of increase in that direction.

The definition of an atom and its bounding surface are made both qualitatively and quantitatively apparent in terms of the patterns of trajectories traced out by the gradient vectors of the density. A trajectory map, complementary to the displays of the density, is given in figure 2.1(c). Because $\rho(\mathbf{r})$ exhibits a maximum at each nucleus in any plane that contains the nucleus, the nucleus acts as a *global attractor* in the gradient vector field of the density. As a consequence, the three-dimensional space of the molecule is divided into atomic basins, each basin being defined by the set of trajectories that terminate at a given nucleus. *An atom is defined as the union of a nucleus and its associated basin.* The critical point that appears at the minimum in the density between the maxima for a pair of neighbouring nuclei is a bond critical point (bcp). The density is a minimum at this cp along a line linking the two nuclei, a line that is uniquely defined by the pair of trajectories that originate at this cp and terminate at the two adjoining nuclear maxima. The line so defined represents a line of maximum density with respect to any neighbouring line and for a bound molecule, it is referred to as a *bond path.* A bond path provides a universal indicator of bonding, linking all pairs of bonded atoms, regardless of the nature of the interaction. (*Bader, 1998b*) In chapter 9 we will see how the presence of a bond path is invaluable in resolving cases where the assignment of a bonded pattern in a molecule is ambiguous. (*Bader and Matta, 2001a*) A molecular graph, defined by a molecule's connected set of bond paths, defines a molecular structure with particular critical points defining the presence of bonded rings and cages. The molecular graph of the guanine-cytosine base pair is shown in figure 2.1(d). Nuclear motions can induce topological changes in the density that correspond to the making and breaking of chemical bonds and to a change in molecular structure. Thus the topology of

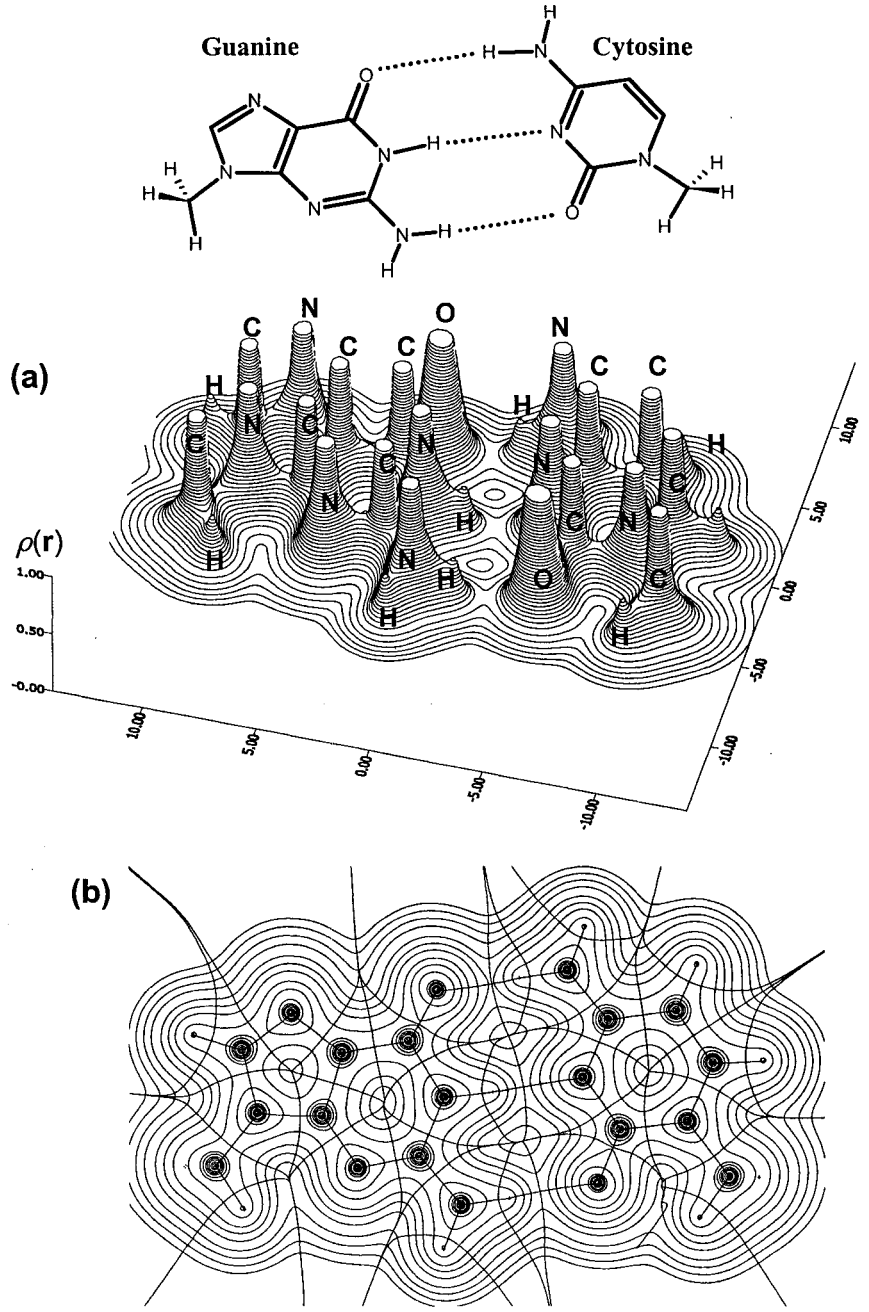


Figure 2.1. Displays of the electron density in the molecular plane of a guanine-cytosine triply hydrogen bonded DNA base pair. **(a)** A relief map of $\rho(\mathbf{r})$ in the plane of the nuclei and the plane coordinates (all values in au). The density is truncated at 1.0 au. **(b)** A corresponding contour map of $\rho(\mathbf{r})$ in the same plane as in (a). The lines connecting the nuclei are the bond paths, and the lines delimiting each atom are the intersection of the interatomic surfaces with the plane. Crosses not linked by bond paths are the projections of the nuclear positions of atoms out of the plane.

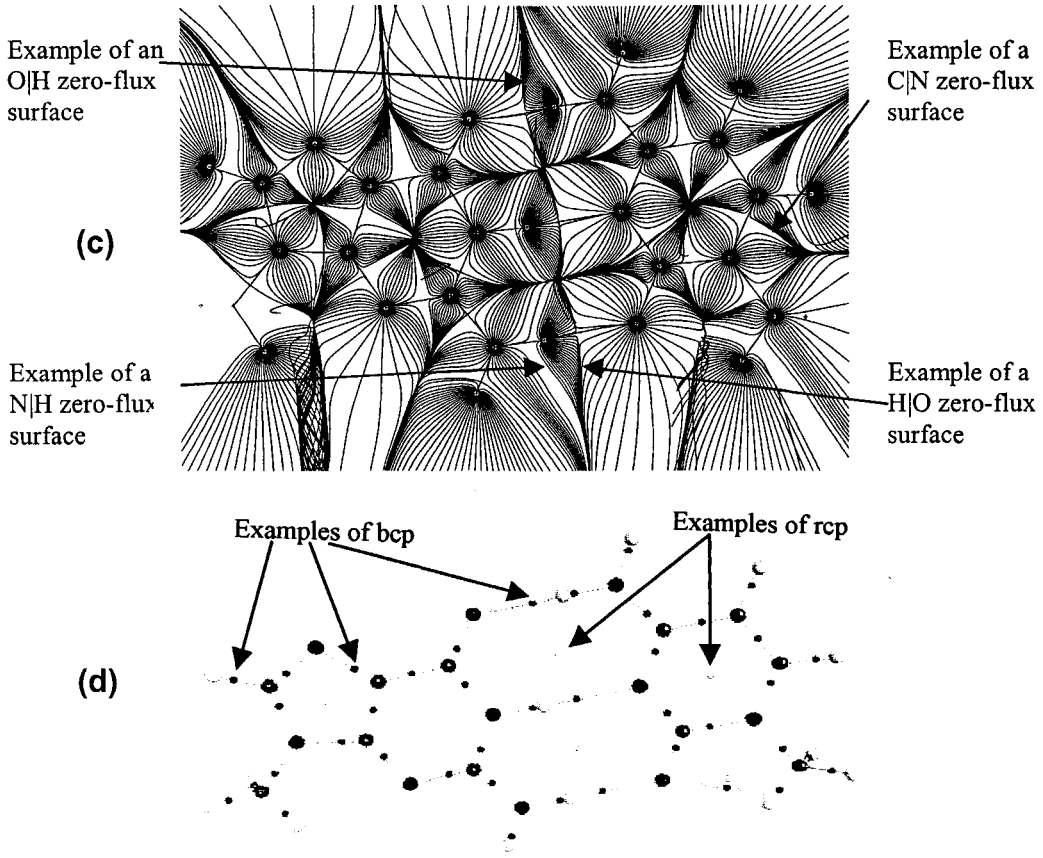


Figure 2.1 (Cont'd). **(c)** Displays of the trajectories of $\nabla\rho(\mathbf{r})$ in the same plane of (a). All the paths in the neighborhood of a given nucleus terminate at that nucleus and define the atomic basin. Some interatomic zero-flux surfaces are indicated and denoted by the vertical bars. **(d)** The molecular graph defining the structure of the guanine-cytosine base pair. The set of bond paths displayed in this figure recover the usual chemical structure, with every bonded atom linked by a unique bond path, the small dots along which being the bond critical point (bcp). The unconnected dots enclosed by rings are the ring critical points (rcp).

the charge distribution and its change, as induced by nuclear motions, embodies the essential elements of structure and structural change. (*Bader, 1990*)

The density is a maximum for all directions perpendicular to the bond path at the position of a bcp, and it thus serves as the terminus for a set of trajectories. The set of trajectories that terminate at a bond critical point define the *interatomic surface* (IAS) that separates the basins of the neighbouring atoms. Since the surface is defined by trajectories of $\nabla\rho(\mathbf{r})$ that terminate at a point, and since

trajectories never cross, an interatomic surface is one of *zero-flux* in the gradient vector field of the density, a property made clear in figure 2.1(c). Each basin is bounded by one or more surfaces of zero flux in $\nabla\rho(\mathbf{r})$, some of which may occur at infinity. Thus, *an atom in a molecule may be alternatively defined as a region of space bounded by one or more zero-flux surfaces.*

2.2 The Zero-Flux Surface and Proper Open Quantum Systems

In the previous section the idea of a zero-flux surface partitioning the molecular space was introduced qualitatively and pictorially. Given a vector normal to the surface $\mathbf{n}(\mathbf{r})$, the mathematical expression characterizing the zero-flux surface is:

$$\nabla\rho(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) = 0 \qquad \forall \ \mathbf{r}\in S(\Omega,\mathbf{r}), \tag{2.6}$$

where Ω refers to the atomic basin and S its bounding surface. The equation means that the normal vector to the surface is orthogonal to the gradient vector field *at every point* on the surface, i.e. the phrase "the zero-flux condition" which is used in this thesis for brevity implies "the condition of *local* zero-flux". This equation plays a central role in the quantum theory of atoms in molecules and will be shown to represent the necessary boundary condition for the definition of proper open quantum systems. The partitioning of molecular space at the zero-flux surfaces is a consequence of the principal topological feature in the density that local maxima in $\rho(\mathbf{r})$ occur at the position of the nuclei, and that each nucleus acts as an attractor in the gradient vector field of the electron density. Since there are no voids or dead-volumes unaccounted for, the resulting partitioning exhausts all space into a disjoint set of mono-nuclear regions (atomic basins), each bounded by an atomic surface satisfying Eq. 2.6. The atomic surface in general consists of a number of interatomic surfaces, there being one such surface separating the basins between every pair of neighbouring atoms. (*Bader, 1990*)

We seek now to answer the question: *How does the zero-flux condition lead to the only physically sound partitioning of a molecule into its composite atoms?*

Before proceeding we must distinguish clearly an *open system*, proper or improper, as opposed to a *closed system*. In thermodynamics, a closed system is an insulated system where the number of particles and energy are constant and cannot be exchanged with the surroundings, to put it differently, it is a bounded system with no flow of current of any kind across its boundaries. An example of a closed system is a perfectly insulated calorimeter. In contrast, an open system is a bounded system that can exchange matter, energy and entropy with its surroundings. An open system is *open* for the flow of currents across its bounding surface. A typical open system is a living organism which utilizes metabolism to free itself from the entropy it generates. In doing so, the organism attempts to minimize its entropy to avoid the state of maximum entropy (death) through a continual flow of heat to its surrounding. (*Schrödinger, 1967*)

An open quantum system is any bounded volume smaller than the molecule's and totally contained within it - i.e. a portion of a molecule in three-dimensional space - and thus in this context will be called a *subsystem*. The closed system will also be referred to as the *total system* when emphasis is to be made on the fact that the total system is composed of the union of several open systems – this semantic distinction should not distract from their synonymy.

For an arbitrary subsystem, that is to say one that is not bounded by a zero-flux surface, the kinetic energy cannot be defined uniquely. Consider the identity:

$$\begin{aligned}\nabla^2(\psi^*\psi) &= \nabla\cdot\left[(\nabla\psi^*)\psi + \psi^*(\nabla\psi)\right] \\ &= (\nabla^2\psi^*)\psi + \psi^*(\nabla^2\psi) + 2\nabla\psi^*\nabla\psi.\end{aligned} \tag{2.7}$$

Rearranging Eq. 2.7, and realizing that $\psi^*\psi=\rho$, one obtains:

$$-\left[\psi\nabla^2\psi^* + \psi^*\nabla^2\psi\right] = 2\nabla\psi^*\cdot\nabla\psi - \nabla^2\rho. \tag{2.8}$$

Using the identity Eq. 2.8 in the many-electron case, we get:

$$-\frac{\hbar^2}{4m}\sum_i\left[\psi\nabla_i^2\psi^* + \psi^*\nabla_i^2\psi\right] = \frac{\hbar^2}{2m}\sum_i\nabla_i\psi^*\cdot\nabla_i\psi - \frac{\hbar^2}{4m}\sum_i\nabla_i^2\rho. \tag{2.9}$$

Since ψ is antisymmetrized and the electrons are indistinguishable, the average of a sum of N one-electron operators can be replaced by N times the average of one of the operators. The average of each

term in Eq. 2.9 is readily found by integrating over the space coordinates of all the electrons but one, and summing over all the spins, a mode of integration that will be denoted in this thesis by $\int d\tau'$. The result of the integration is then multiplied by the number of electrons to yield:

$$\underbrace{-\frac{\hbar^2}{4m}N\int d\tau'\left[\psi\nabla_i^2\psi^* + \psi^*\nabla_i^2\psi\right]}_{K} = \underbrace{\frac{\hbar^2}{2m}N\int d\tau'\,\nabla_i\psi^*\cdot\nabla_i\psi}_{G} - \underbrace{\frac{\hbar^2}{4m}N\int d\tau'\,\nabla_i^2\rho}_{L}, \qquad (2.10)$$

where K is the Schrödinger kinetic energy, G is the gradient kinetic energy, and L is the integrated function of the Laplacian. The integrands in these three terms are the respective densities $K(\mathbf{r})$, $G(\mathbf{r})$, and $L(\mathbf{r})$. Clearly, integrating this equation over all space will result in the equality $K=G=T$, where T is the (well defined) kinetic energy, since the gradient of the density vanishes at infinity ($\nabla^2\rho(\mathbf{r})=\nabla\cdot\nabla\rho(\mathbf{r})=0$ at ∞). On the other hand, integrating Eq. 2.10 over an arbitrary subsystem Ω (which does not satisfy the zero-flux condition) results in the inequality $K(\Omega)\neq G(\Omega)$ which is clearly a non-physical results since calculation of the total kinetic energy ($T(\Omega)$) of the subsystem from the two different expressions of the kinetic energy density provides different answers. In other words, the kinetic energy of the subsystem is not well-defined. Rewriting Eq. 2.10 in a more compact form, and integrating over the volume of an open system Ω we get:

$$K(\Omega) = G(\Omega) - \frac{\hbar^2}{4m}N\int_{\Omega} d\tau'\,\nabla\cdot\nabla\rho. \qquad (2.11)$$

Using the divergence theorem, the volume integral in Eq. 2.11 can be transformed to a surface integral, the integrand being dotted in the normal vector to the surface, giving:

$$K(\Omega) = G(\Omega) - \frac{\hbar^2}{4m}N\oint dS(\Omega,\mathbf{r})\nabla\rho\cdot\mathbf{n}(\mathbf{r}). \qquad (2.12)$$

From Eq. 2.12 it is clear that

$$K(\Omega) = G(\Omega) = T(\Omega) \qquad (2.13)$$

only when the surface integral vanishes. Thus, the average kinetic energy is well defined for an atom in a molecule.

The study of the behaviour of the kinetic energy density resulted in an important observation: when the electron density of a proper open system is transferable, the kinetic energy density is equally transferable. (*Bader and Beddall, 1972*) If one could demonstrate that a virial theorem applies to an atom in a molecule, then one could use it to define the energy of this atom ($E(\Omega)$) in terms of the averaged kinetic energy ($T(\Omega)$), i.e. one would be able to write:

$$T(\Omega) = -E(\Omega) . \tag{2.14}$$

In words, *the average total energy of a proper open quantum system Ω is equal to the negative of its average kinetic energy.* Should Eq. 2.14 be the case, the total energy of the molecule can be partitioned into additive atomic contributions in the same manner as the kinetic energy. Thus, for an equilibrium geometry, the atomic sum of the $E(\Omega)$ equals E, the total energy of the molecule, which includes the nuclear-nuclear repulsion energy:

$$E = \sum_{\Omega} E(\Omega) . \tag{2.15}$$

This result implies that when the density is transferable the kinetic energy density is equally transferable, and so is the total energy (Eq. 2.14). Thus *the form of an atom in real space – as determined by the electron density - determines its contribution to the total molecular energy, and atoms that "look" the same contribute equal amounts to the total energy.* Thus, as a consequence of Eq. 2.14 - which will be proven in the next section – the topological atoms defined in Eq. 2.6 exhibit two fundamental properties that the empirical atoms of chemistry are known to possess: additivity and transferability.

2.3 The Coincidence of the Topological Atom and the Quantum Atom

An important distinction between a closed and an open system is that an operator which is Hermitian when integrated over all space may loose this property when integrated over a subspace. The physical

result of the non-Hermiticity of an operator's average over the open system is the formation of fluxes in property currents across the surface bounding it. This is made apparent by deriving the Heisenberg equation of motion for the observable $\hat{A}$ over an open system. The Heisenberg equation of motion, which determines the evolution of $\hat{A}$ over time, when expressed in the Schrödinger representation is given by:

$$\frac{d\langle \hat{A} \rangle}{dt} = \frac{i}{\hbar} \langle \Psi | [\hat{H}, \hat{A}] | \Psi \rangle. \tag{2.16}$$

The choice of the operator $\hat{A}$ determines the corresponding theorem. The virial theorem, for example is obtained by setting $\hat{A} = \mathbf{r} \cdot \mathbf{p}$, where $\mathbf{r}$ and $\mathbf{p}$ refer to the electron's position and momentum vectors respectively. From the definition of the commutator, one has:

$$\langle \Psi | [\hat{H}, \hat{A}] | \Psi \rangle = \langle \Psi | \hat{H}\hat{A} | \Psi \rangle - \langle \Psi | \hat{A}\hat{H} | \Psi \rangle, \tag{2.17}$$

utilising the Hermiticity property[1] and Schrödinger equation $\hat{H}|\Psi\rangle = E|\Psi\rangle$, one obtains:

$$\begin{aligned} \langle \Psi | [\hat{H}, \hat{A}] | \Psi \rangle &= \langle \hat{H}\Psi | \hat{A}\Psi \rangle - \langle \Psi | \hat{A}\, E | \Psi \rangle \\ &= E\langle \Psi | \hat{A} | \Psi \rangle - E\langle \Psi | \hat{A} | \Psi \rangle = 0 \,. \end{aligned} \tag{2.18}$$

(Note that Ψ does not have to be an eigenfunction of the operator $\hat{A}$). From the result in Eq. 2.18 we have:

$$\langle \Psi | [\hat{H}, \hat{A}] | \Psi \rangle = 0, \tag{2.19}$$

where the commutator vanishes for any Hermitian operator $\hat{A}$. The result in Eq. 2.19 is known as the *hypervirial theorem*. One can no longer assume the Hermiticity of an operator when averaged over an

---

[1] An operator $\hat{O}$ is Hermitian when it is self-adjoint, i.e. $\hat{O} = \hat{O}^{\dagger}$. The matrix representation of a Hermitian operator satisfies $\langle a|\hat{O}|b\rangle = \langle a|\hat{O}^{\dagger}|b\rangle = \langle b|\hat{O}|a\rangle^{*}$. Hermiticity ensures that the eigenvalues of $\hat{O}$ are real.

open system (Ω). In other words, the equality $\left\langle \Psi \middle| \hat{H}\hat{A}\Psi \right\rangle_\Omega = \left\langle \hat{H}\Psi \middle| \hat{A}\Psi \right\rangle_\Omega$ can no longer be assumed. As a result, the two terms in Eq. 2.17 do not cancel in general for an open system, and Eq. 2.19 is replaced by:

$$\left\langle \Psi \middle| \left[\hat{H}, \hat{A}\right] \middle| \Psi \right\rangle_\Omega = \left\langle \Psi \middle| \hat{H}\hat{A}\Psi \right\rangle_\Omega - \left\langle \Psi \middle| \hat{A}\hat{H}\Psi \right\rangle_\Omega \neq 0 \,. \tag{2.20}$$

Adding and subtracting $\left\langle \hat{H}\Psi \middle| \hat{A}\Psi \right\rangle_\Omega$ to this result, and using Schrödinger equation $\hat{H}\Psi = E\Psi$ we get:

$$\left.\begin{aligned} \left\langle \Psi \middle| \left[\hat{H}, \hat{A}\right] \middle| \Psi \right\rangle_\Omega &= \left\langle \Psi \middle| \hat{H}\hat{A}\Psi \right\rangle_\Omega - \left\langle \hat{H}\Psi \middle| \hat{A}\Psi \right\rangle_\Omega + \left\langle \hat{H}\Psi \middle| \hat{A}\Psi \right\rangle_\Omega - \left\langle \Psi \middle| \hat{A}\hat{H}\Psi \right\rangle_\Omega \\ &= \left\langle \Psi \middle| \hat{H}\hat{A}\Psi \right\rangle_\Omega - \left\langle \hat{H}\Psi \middle| \hat{A}\Psi \right\rangle_\Omega + E\left\langle \Psi \middle| \hat{A}\Psi \right\rangle_\Omega - E\left\langle \Psi \middle| \hat{A}\Psi \right\rangle_\Omega \\ &= \left\langle \Psi \middle| \hat{H}\hat{A}\Psi \right\rangle_\Omega - \left\langle \hat{H}\Psi \middle| \hat{A}\Psi \right\rangle_\Omega . \end{aligned}\right\} \tag{2.21}$$

To keep the physics clear, in the following discussion we will treat the case of a single electron, which is done without loss of generality since the many electron case can be obtained by the insertion of $N\int d\tau'$. Writing the Hamiltonian of a single-electron (Eq.1.14), and realizing that the potential energy operator results in an additive contribution which will cancel in the final form of Eq. 2.21, we obtain:

$$\left.\begin{aligned} \left\langle \Psi \middle| \left[\hat{H}, \hat{A}\right] \middle| \Psi \right\rangle_\Omega &= \frac{\hbar^2}{2m} \int_\Omega d\mathbf{r} \left[\left(\nabla^2\Psi^*\right) \hat{A}\,\Psi - \Psi^* \nabla^2\left(\hat{A}\Psi\right) \right] \\ &= \frac{\hbar^2}{2m} \int_\Omega d\mathbf{r}\, \nabla \cdot \left[\left(\nabla\Psi^*\right) \hat{A}\,\Psi - \Psi^* \nabla\left(\hat{A}\Psi\right) \right] . \end{aligned}\right\} \tag{2.22}$$

Using Gauss' theorem to transform the volume integral into a surface integral, one gets:

$$\left\langle \Psi \middle| \left[\hat{H}, \hat{A}\right] \middle| \Psi \right\rangle_\Omega = \oint dS(\Omega, \mathbf{r}) \underbrace{\left\{ \frac{\hbar^2}{2m} \left[\left(\nabla\Psi^*\right) \hat{A}\,\Psi - \Psi^* \nabla\left(\hat{A}\Psi\right) \right] \right\}}_{i\times(\text{negative of the current density of property } A)} \cdot \mathbf{n}(\mathbf{r}) \,. \tag{2.23}$$

The integrand of Eq. 2.23 is identified with i times the negative of the current density vector of the property A ($-\mathbf{j}_A(\mathbf{r})$) through the surface, the quantum mechanical current being defined as:

$$\mathbf{j}_A(\mathbf{r}) = \frac{\hbar}{2mi}\left[\Psi^*\nabla(\hat{A}\Psi) - (\nabla\Psi^*)(\hat{A}\Psi)\right]. \tag{2.24}$$

This result (Eq. 2.23) highlights the clear distinction in the behaviour of an open system as opposed to a closed system: in the open system the average of the commutator $\left\langle \Psi \middle| \left[\hat{H},\hat{A}\right] \middle| \Psi \right\rangle_\Omega$ does not vanish, and instead is given by the flux in the property density of A through the surface. Taking each term together with its complex conjugate (cc) to ensure that the result is a real (observable) quantity, then dividing by two, we get:

$$\frac{1}{2}\left\{\frac{i}{\hbar}\left\langle \Psi \middle| \left[\hat{H},\hat{A}\right] \middle| \Psi \right\rangle_\Omega + cc\right\} = \frac{1}{2}\left[\oint dS(\Omega,\mathbf{r})\ \mathbf{j}_A \cdot \mathbf{n}(\mathbf{r}) + cc\right]. \tag{2.25}$$

Thus whenever $\hat{H}$ and $\hat{A}$ do not commute, there is a flux in the property density of A across the surface bounding the proper open system defined by the surface term in Eq. 2.25.

Schrödinger obtained $\hat{H}\Psi = E\Psi$ as the Euler equation by making the energy functional $J[\Psi]$ stationary with respect to first-order variations in the wavefunction Ψ, i.e. $\delta J[\Psi]=0$, and subjecting it to the normalization constraint. Schrödinger's energy functional is given by: (*Schrödinger, 1926a*)

$$J[\Psi] = \int d\tau \left[\left(\frac{\hbar^2}{2m}\right)\nabla\Psi^* \cdot \nabla\Psi + (\hat{V}+\lambda)\Psi^*\Psi\right], \tag{2.26}$$

where the constant λ is the Lagrange multiplier for the constraint on Ψ which is identified with the negative of the total energy of the system ($E=-\lambda$).

Serbrenik and Bader extended Schrödinger's derivation to an open system by defining the open system analogue of Schrödinger's functional as follows: (*Srebrenik and Bader, 1975*)

$$G[\Psi,\Omega] = \int_\Omega d\mathbf{r}\left[\left(\frac{\hbar^2}{2m}\right)\nabla\Psi^* \cdot \nabla\Psi + (\hat{V}-E)\Psi^*\Psi\right] \equiv \int_\Omega d\mathbf{r}\ f(\Psi,\nabla\Psi), \tag{2.27}$$

where f denotes the integrand in the functional in $G[\Psi,\Omega]$ which is a function of both Ψ and $\nabla\Psi$. Noticing that the variation of the subspace functional must include the variation of the bounding surface, and treating Ψ and Ψ^* as independent variables, one obtains:

$$\delta G[\Psi,\Omega] = \underbrace{\int_{\Omega} d\mathbf{r}\left[\left(\frac{\partial f}{\partial \Psi}\right)\delta\Psi + \left(\frac{\partial f}{\partial \nabla\Psi}\right)\delta\nabla\Psi\right]}_{\text{variation of } f \text{ while the surface is constant}} + \underbrace{\oint dS(\Omega,\mathbf{r})\; f\times\delta S(\Omega,\mathbf{r})}_{\text{variation of the surface while } f \text{ is constant}} + cc, \qquad (2.28)$$

where the arguments of f have been dropped, and δ denotes infinitesimal changes (variations), and where the *cc* results from the variation of Ψ^*. Eq. 2.28 can be shown to yield (see review (*Bader, 1998a*)):

$$\delta G[\Psi,\Omega] = \underbrace{\int_{\Omega} d\mathbf{r}\left[H\Psi - E\Psi\right]\delta\Psi}_{\substack{\text{variation of } f \\ \text{while the surface} \\ \text{is constant}}} + \oint dS(\Omega,\mathbf{r})\left[\left(\frac{\hbar}{2m}\right)\nabla\Psi^*\cdot\mathbf{n}(\mathbf{r})\underbrace{\delta\Psi}_{\substack{\text{variation} \\ \text{of } \Psi \textit{ on} \\ \text{the surface}}} + f\times\underbrace{\delta S(\Omega,\mathbf{r})}_{\substack{\text{variation of the} \\ \text{surface through} \\ \text{the variation of } \Psi}}\right] + cc. \qquad (2.29)$$

If Ω refers to the total system, i.e. when $\Omega=\mathbb{R}^3$, then $S(\Omega,\mathbf{r})$ occurs at ∞ and the surface term vanishes either by requiring that $\delta\Psi=0$ at $\mathbf{r}=\infty$, or by imposing the so called *natural boundary conditions*:

$$\nabla\Psi\cdot\mathbf{n}(\mathbf{r}) = 0 \qquad \text{and} \qquad \nabla\Psi^*\cdot\mathbf{n}(\mathbf{r}) = 0 \qquad \forall\ \mathbf{r} = \infty. \qquad (2.30)$$

It follows that for the total system, Eq. 2.27 reduces to:

$$\delta G[\Psi,\Omega] = \int_{\Omega} d\mathbf{r}\left[H\Psi^* - E\Psi^*\right]\delta\Psi + cc = 0. \qquad (2.31)$$

In order for Eq. 2.31 to be true for any arbitrary variations in Ψ, the bracketed term in this equation must vanish. This requirement yields the two conjugate Schrödinger equations:

$$\hat{H}\Psi = E\Psi \qquad \text{and} \qquad \hat{H}\Psi^* = E\Psi^* \qquad (2.32)$$

From the last result, one concludes that Schrödinger's equation applies at the point of variation and therefore, the first term in Eq. 2.29 will not contribute to δG for a subsystem as well as for the total system. Thus, at the point of variation Eq. 2.29 reduces to:

$$\delta G[\Psi,\Omega] = \oint dS(\Omega,\mathbf{r})\left[\left(\frac{\hbar}{2m}\right)\nabla\Psi^*\cdot\mathbf{n}(\mathbf{r})\delta\Psi + f(\Psi,\nabla\Psi)\times\delta S(\Omega,\mathbf{r})\right] + cc. \qquad (2.33)$$

From the above discussion, one concludes that unlike the total system for which $\delta G[\Psi]=0$, the variation of this functional does not vanish for an open system but rather is equal to two terms: one

proportional to the variation of Ψ at the boundaries of the open system, and the other proportional to the variation of the boundary itself (Eq. 2.33). This retention of the variation in Ψ at the boundary of the system and the variation of the boundary itself is identical to the step taken by Schwinger to generalise the variation of the action integral to obtain the principle of stationary action. (*Schwinger, 1951*)

Equation 2.33 is not a useful result as such since it contains a surface term which cannot be calculated. To rid the equation of this term is equivalent to restricting the final result to a particular subclass of open systems. Consider an alernative expression for the integrand $f(\Psi,\nabla\Psi)$ which involves the Hamiltonian operator. Including the complex conjugate term in Eq. 2.33, the function $f(\Psi,\nabla\Psi)$ appears twice, and using Eq. 2.9 which relates the two forms of kinetic energy density, one has:

$$2f(\Psi,\nabla\Psi)=\left[(\hat{H}\Psi)^*\Psi+\Psi^*\hat{H}\Psi\right]-2E\Psi^*\Psi+2\left(\frac{\hbar^2}{4m}\right)\nabla^2(\Psi^*\Psi), \tag{2.34}$$

and since Schrödinger equations are assumed to apply, then:

$$f(\Psi,\nabla\Psi)=\left(\frac{\hbar^2}{4m}\right)\nabla^2\rho(\mathbf{r})\,. \tag{2.35}$$

This is an important result. *Equation 2.35 shows that when $\hat{H}\Psi = E\Psi$, the Schrödinger energy functional reduces to a term proportional to the Laplacian of the charge density.* Substituting this result into Eq. 2.33:

$$\delta G[\Psi,\Omega]=\oint dS(\Omega,\mathbf{r})\left[\left(\frac{\hbar^2}{4m}\right)\left(2\nabla\Psi^*\cdot\mathbf{n}(\mathbf{r})\delta\Psi+\delta S(\Omega,\mathbf{r})\,\nabla^2\rho\right)\right]+cc. \tag{2.36}$$

One cannot progress from this expression to obtain a general variational principle for an arbitrary region of space. It will next be demonstrated that Eq. 2.36 is transformed into the atomic statement of the hypervirial theorem when the subsystem Ω is restricted to one satisying a particular variational constraint on the form of the open system. The variational constraint will be introduced in terms of a trial function Φ.

A region $\Omega(\Phi)$ is defined in terms of the trial function Φ that is bound by a zero-flux surface (Eq. 2.6) in the trial density ($\rho_\Phi = \int d\mathbf{r}\ \Phi^* \Phi$). Recalling the topological definition of an atom in Eq. 2.6, a region of space bounded by a zero-flux surface is henceforth called an atom. It is required that as $\Phi \rightarrow \Psi$, $\Omega(\Phi)$ is continuously deformable into the region $\Omega(\Psi)$ associated with the atom. The region $\Omega(\Phi)$ thus represents the atom in the varied total system, which is described by the trial function Φ, just as $\Omega(\Psi)$ represents the atom when the total system is in the state described by Ψ. Requiring the fulfilment of these conditions amounts to imposing the variational constraint that $\mathrm{div}(\nabla\rho_\Phi) = \nabla\cdot\nabla\rho_\Phi$ integrates to zero at all stages of the variation, i.e.: (*Bader, 1990*)

$$\int_{\Omega(\Phi)} \nabla^2 \rho_\Phi(\mathbf{r})\, d\mathbf{r} = 0 \tag{2.37}$$

for all admissible Φ, which implies:

$$\delta\left[\int_{\Omega} \nabla^2 \rho_\Phi(\mathbf{r})\, d\mathbf{r}\right] = 0. \tag{2.38}$$

The variation in Eq. 2.38 requires a variation in both the surface of Ω as well as the variation of the integrand, similar to Eq. 2.28. Since the sum of these two variations vanishes as required by the constraint, one is able to equate one to the negative of the other:

$$\frac{\hbar^2}{4m} \oint dS(\Omega,\mathbf{r})\ \delta S(\Omega,\mathbf{r})\ \nabla^2\rho(\mathbf{r}) = -\frac{\hbar^2}{4m}\int_{\Omega} d\mathbf{r}\ \delta\left[\nabla^2\rho(\mathbf{r})\right] \tag{2.39}$$

$$= -\frac{\hbar^2}{4m}\int_{\Omega} d\mathbf{r}\ \nabla\cdot\left[\nabla\Psi^*\delta\Psi + \Psi^*\nabla\delta\Psi\right] \tag{2.40}$$

$$= -\frac{\hbar^2}{4m}\oint dS(\Omega,\mathbf{r})\left[\nabla\Psi^*\delta\Psi + \Psi^*\nabla\delta\Psi\right]\cdot\mathbf{n}(\mathbf{r})\,, \tag{2.41}$$

where in going from Eq. 2.40 to Eq. 2.41 Gauss' theorem was utilised. Now, substituting the right hand side of Eq. 2.41 into Eq. 2.36, we get:

$$\delta G[\Psi,\Omega] = \frac{\hbar^2}{4m}\oint dS(\Omega,\mathbf{r})\left[(\nabla\Psi^*)\,\delta\Psi - \Psi^*\delta(\nabla\Psi)\right]\cdot\mathbf{n}(\mathbf{r}) + cc. \tag{2.42}$$

The last expression shows how Schrödinger's energy functional can be expressed in terms of the surface integral of the current density, which is defined in Eq. 2.24 when $\hat{A} = \hat{1}$. The variation of Eq. 2.24 yields:

$$\delta \mathbf{j}(\mathbf{r}) = \frac{\hbar}{2mi}\left[\Psi^* \delta(\nabla \Psi) - (\nabla \Psi^*)\delta \Psi\right]. \tag{2.43}$$

Substituting Eq. 2.43 into Eq. 2.42, we obtain:

$$\delta G[\Psi, \Omega] = -\frac{i\hbar}{2} \oint dS(\Omega, \mathbf{r})\ \delta \mathbf{j}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) + cc. \tag{2.44}$$

Schwinger uses *infinitesimal unitary transformations*, acting separately on the state function or observables, to generate all possible physical changes in the dynamical variables of a quantum system. The operator for an infinitesimal unitary transformation and its inverse (its Hermitian conjugate) are given by:

$$\hat{U} = \hat{1} - \frac{i\varepsilon}{\hbar}\hat{G} \qquad \text{and} \qquad \hat{U}^{-1} = \hat{1} + \frac{i\varepsilon}{\hbar}\hat{G}, \tag{2.45}$$

where ε is an infinitesimal and $\hat{G}$ is any linear Hermitian operator (observable) which is referred to as the *generator of the transformation*. The first-order changes in Ψ and Ψ^* are equated to the action of the generator $\hat{G}$ on them according to:

$$\delta \Psi = -\frac{i\varepsilon}{\hbar}\hat{G}\Psi \qquad \text{and} \qquad \delta \Psi^* = \frac{i\varepsilon}{\hbar}\hat{G}\Psi^* \quad . \tag{2.46}$$

Replacing the variation of $\delta\Psi$ and $\delta\Psi^*$ in Eqs.2.43 (and 2.44) by the action of infinitesimal generators (Eq.2.47) yields: (*Bader, 1994*)

$$\delta G[\Psi, \Omega] = -\frac{\varepsilon}{2} \oint dS(\Omega, \mathbf{r})\ \delta \mathbf{j}_G(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) + cc. \tag{2.47}$$

By comparing Eq. 2.25 and Eq. 2.47 we can write:

$$\delta G[\Psi, \Omega] = -\frac{\varepsilon}{2}\left\{\frac{i}{\hbar}\left\langle \Psi \middle| \left[\hat{H}, \hat{G}\right] \middle| \Psi \right\rangle_{\Omega} + cc\right\}. \tag{2.48}$$

This result is a statement of the principle of stationary action for a system in a stationary state. Eq. 2.48 applies equally to the total closed system and to its composing proper open systems, treating the total systems and its proper subsystems on an equal quantum mechanical footing.

Thus, *the topological and quantum definition of an atom in a molecule are brought into coincidence through the zero-flux condition expressed in Eq. 2.6.*

The topological atoms have been identified with the atoms of chemistry because they recover the essential ideas associated with the atomic concept: (*Bader et al., 1994*) (i) The atomic properties, defined by the principle of stationary action, are characteristic and additive, summing to yield the corresponding values for the total system; this is expressed:

$$M = \sum_{\Omega} M(\Omega) , \tag{2.49}$$

where M is any molecular property and $M(\Omega)$ is the corresponding additive atomic counterpart; (ii) The properties of an atom are as transferable from one system to another as is the form of the atom in real space - if the density distribution of the atom is transferable, so are its properties. The latter property in particular is what identifies the topological atom with the atoms of chemistry. There are many instances in experimental chemistry where one observes what are termed additivity schemes, wherein the value of a property, such as the heat of formation, the electric polarizability or the magnetic susceptibility of a molecule can be expressed in terms of additive group contributions. In a homologous series, such as the linear hydrocarbons, the additivity principle is satisfied to within experimental accuracy. One finds that in those cases *where the density distribution of an atom or group is essentially transferable between systems, all property densities, including the energy density and those induced by external fields, exhibit the same degree of transferability.* Thus the truism that two identical macroscopic samples of matter must exhibit the same values for all properties, applies at the atomic level as well and indeed, the observation of identical properties for two macroscopic samples can be viewed as a consequence of the behaviour of matter at the atomic level. Without the possibility of relating the value of every observable property of any system, finite or extended, to the

sum of the contributions of its individual composite atoms, the concept of the atom as the building block of matter loses its usefulness.

2.3 The Atomic Virial Theorem

In this section, the virial theorem of an open system is derived as an example showing how the statement of stationary action (Eq. 2.48) can be used to obtain the quantum mechanics of a proper open quantum system.

Setting the generator in Eq. 2.48 equal to the virial operator $\hat{\mathbf{r}}\cdot\hat{\mathbf{p}}$, where $\hat{\mathbf{r}}$ and $\hat{\mathbf{p}}$ are the position and momentum operators for a single electron, and equating Eqs. 2.47 and 2.48 one obtains:

$$\frac{N}{2}\left\{\frac{i}{\hbar}\left\langle\Psi\middle|\left[\hat{H},\hat{\mathbf{r}}\cdot\hat{\mathbf{p}}\right]\middle|\Psi\right\rangle_{\Omega}+cc\right\}=-\frac{N}{2}\left[\oint dS(\Omega,\mathbf{r})\ \mathbf{j}_{\hat{\mathbf{r}}\cdot\hat{\mathbf{p}}}(\mathbf{r})\cdot\mathbf{n}(\mathbf{r})+cc\right]. \tag{2.50}$$

When the commutator is evaluated, the L.H.S. of Eq. 2.50 yields:

$$-2N\left(\frac{\hbar^2}{4m}\right)\int_{\Omega}d\tau\int d\tau'\left[\Psi^*\nabla^2\Psi+(\nabla^2\Psi^*)\Psi\right]+N\int_{\Omega}d\tau\int d\tau'\,\Psi^*(-\mathbf{r}\cdot\nabla\hat{V})\Psi \tag{2.51}$$

$$=2T(\Omega)+\mathscr{V}_b(\Omega), \tag{2.52}$$

where $\mathscr{V}_b(\Omega)$ is the integrated value of the virial of the Ehrenfest force acting on an electron in the basin of atom Ω, and $T(\Omega)$ is the integrated kinetic energy density over Ω.

Now we evaluate the surface integral on the R.H.S of Eq. 2.50. From Eq. 2.47, we have:

$$\frac{N}{2}\left[\oint dS(\Omega,\mathbf{r})\ \mathbf{j}_{\hat{\mathbf{r}}\cdot\hat{\mathbf{p}}}(\mathbf{r})\cdot\mathbf{n}(\mathbf{r})+cc\right]$$

$$=-\frac{N\hbar^2}{4m}\left\{\oint dS(\Omega,\mathbf{r})\int d\tau'\Psi^*\nabla(\mathbf{r}\cdot\nabla\Psi)-\nabla\Psi^*(\mathbf{r}\cdot\nabla\Psi)+\Psi\nabla(\mathbf{r}\cdot\nabla\Psi^*)-\nabla\Psi(\mathbf{r}\cdot\nabla\Psi^*)\right.$$

$$=\underbrace{-\oint dS(\Omega,\mathbf{r})\ \mathbf{r}\cdot\vec{\boldsymbol{\sigma}}(\mathbf{r})\cdot\mathbf{n}(\mathbf{r})}_{-\mathscr{V}_S(\Omega)}\underbrace{-\frac{\hbar^2}{4m}\oint dS(\Omega,\mathbf{r})\nabla\rho(\mathbf{r})\cdot\mathbf{n}(\mathbf{r})}_{L(\Omega)}, \tag{2.53}$$

where the last result is obtained by the use of the identity $\nabla(\mathbf{r}\cdot\nabla\Psi) = \nabla\Psi + \mathbf{r}\cdot\nabla\nabla\Psi$. The second term of the R.H.S. of Eq. 2.53 is $L(\Omega)$ as defined in Eq. 2.11. The quantum stress tensor $\ddot{\boldsymbol{\sigma}}(\mathbf{r})$, which appears in the surface integral, is defined for the one-electron case as:

$$\ddot{\boldsymbol{\sigma}}(\mathbf{r}) = -\frac{\hbar^2}{4m}\left[\Psi^*\nabla(\nabla\Psi) + \nabla(\nabla\Psi^*)\Psi - \nabla\Psi^*\nabla\Psi - \nabla\Psi\nabla\Psi^*\right], \tag{2.54}$$

which can be easily generalized to the many electron case by the insertion of $N\int d\tau'$ as usual.

The negative of the first term of Eq. 2.53 is labeled $\mathscr{V}_S(\Omega)$ and identified with the virial of the Ehrenfest forces acting on the surface of the atom. The quantity $\ddot{\boldsymbol{\sigma}}(\mathbf{r})\cdot\mathbf{n}(\mathbf{r})$ is the outwardly directed force per unit area of the surface and $\mathbf{r}\cdot\ddot{\boldsymbol{\sigma}}(\mathbf{r})\cdot\mathbf{n}(\mathbf{r})$ is the virial of this force, which is defined:

$$\mathscr{V}_S(\Omega) = \oint dS(\Omega,\mathbf{r})\ \mathbf{r}\cdot\ddot{\boldsymbol{\sigma}}(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}), \tag{2.55}$$

Equating the commutator (Eq. 2.52), and surface result (Eq. 2.53), and realizing that for a proper open system $L(\Omega) = 0$ - i.e. the second term in (Eq. 2.53) vanishes, one obtains:

$$-2T(\Omega) = \mathscr{V}_b(\Omega) + \mathscr{V}_S(\Omega) = \mathscr{V}(\Omega), \tag{2.56}$$

where the $\mathscr{V}(\Omega)$ is the total atomic virial (basin + surface virials). Eq. 2.56 is the virial theorem for an atom in a molecule.

In terms of the virial theorem of an open system, we can define the electronic energy $E_e(\Omega)$:

$$E_e(\Omega) = T(\Omega) + \mathscr{V}(\Omega). \tag{2.57}$$

For systems in electrostatic equilibrium, when there are no external (Hellmann-Feynman) forces acting on the nuclei, the virial equals the average potential energy of the molecule, i.e. $\mathscr{V} = V$. Under this condition Eq. 2.56 becomes:

$$-2T(\Omega) = V(\Omega), \tag{2.58}$$

where $V(\Omega)$ is the potential energy of atom Ω, and Eq. 2.57 becomes:

$$E(\Omega) = E_e(\Omega) = T(\Omega) + V(\Omega) = -T(\Omega), \tag{2.59}$$

where $E(\Omega)$ is the total energy of atom Ω. Thus, it is because of the atomic statement of the virial theorem that for an equilibrium geometry, where the forces on the nuclei vanish, the electronic energy of an atom is equal to the total energy of this atom, i.e. $E_e(\Omega)=E(\Omega)$.[1] The result expressed in Eq. 2.59 is extraordinary. The equation reprents a quantum mechanical spatial partitioning of all of the interactions in a molecule: electron-nuclear, electron-electron, and nuclear-nuclear, into a sum of atomic contributions, a sum yielding the total energy of the molecule.

2.5 Properties Derived from the Quantum Theory of Atoms in Molecules

2.5.1 Definition of Bond Path and Bond Properties

The zero-flux surface is defined by a particular set of $\nabla\rho(\mathbf{r})$ trajectories which satisfies Eq. 2.6. All of the members of this set terminate at a single point, a critical point in the density where $\nabla\rho(\mathbf{r}) = 0$. There is one such critical point (bond critical point, bcp) between each pair of atoms that are bonded, i.e. one linked by a bond path and share a common interatomic surface. In addition to the set of trajectories which terminate at the bcp and define an interatomic surface, a pair of trajectories originate at the bcp with each member of the pair terminating at one of the neighbouring nuclei. This latter pair of trajectories defines the bond path - a line through space along which the density is a maximum with respect to any neighbouring line and the atoms so linked are bonded to one another. (*Bader, 1998b*) Only atoms that share a common interatomic surface are linked by a bond path and are therefore, bonded to one another. The ability to uniquely delineate the molecular graph for any density distribution is of particular importance in the case of weak interactions such as hydrogen bonding,

---

[1] It is for this reason that henceforth the symbols $E_e(\Omega)$ and $E(\Omega)$ will be used interchangeably in this thesis unless otherwise stated, since we will always be dealing with molecule at their equilibrium geometries.

since the presence of a bond path unambiguously establishes the presence or absence of bonding (see Fig.2.1 for example).

The interaction between a pair of bonded atoms is characterized by the properties exhibited by the density at the bcp:

- *The electron density at the bcp* (ρ_b) provides a measure of bond order for the bonding between a given pair of atoms. In general, $\rho_b > 0.20$ atomic units (au) for shared or polar interactions and $<$ 0.10 au for interactions between closed shell atoms, such as ionic and hydrogen bonded interactions. The distance of a bcp from nucleus A determines the *bonded radius* of atom A with respect to the interaction defined by the bcp and is denoted by r_b(A). If the bond path is coincident with the internuclear axis, then the sum of the two associated bond radii, termed the bond path length, equals the bond length. If however, the bond path is curved, strained in the chemical sense, then the bond path length will exceed the bond length. Examples of this latter behaviour are found for hydrogen bonded interactions or in the case of bonding within strained cyclic molecules.
- *The Laplacian of the electron density at the bcp* ($\nabla^2\rho_b$) is the sum of the three curvatures of the density at the critical point, the two perpendicular to the bond path, λ_1 and λ_2 being negative while the third, λ_3, lying along the bond path, is positive. The negative curvatures measure the degree to which density is concentrated along the bond path, while the positive curvature measures the extent to which it is depleted in the region of the interatomic surface and concentrated in the individual atomic basins. In a shared interaction, density is accumulated between the nuclei and concentrated along the bond path so that ρ_b is large and $\nabla^2\rho_b < 0$, as exemplified by C-H for which $\rho_b = 0.29$ au and $\nabla^2\rho_b = -1.1$ au. For a closed-shell interaction density is removed from the region of contact of the two atoms and hence ρ_b is small and $\nabla^2\rho_b > 0$, as exemplified by the hydrogen bond formed between N-H and C=O for which $\rho_b = 0.01$ au and $\nabla^2\rho_b = +0.03$ au. Polar interactions such as C=O, while characterized by significant charge accumulations between the

nuclei typical of shared interactions, are dominated by charge transfer and the bcp falls in the region where the valence density borders the core of the electropositive atom. The Laplacian of the density rises steeply in this region, undergoes a change in sign, and $\nabla^2\rho_b$ can be of either sign for polar interactions.

- *The bond ellipticity* (ϵ) measures the extent to which density is preferentially accumulated in a given plane containing the bond path. It is defined as $\epsilon = (\lambda_1 / \lambda_2 - 1)$ where λ_1 is the perpendicular curvature of greatest magnitude. The ellipticity provides a measure of double bond character with ϵ = 0.0, 0.23 and 0.45 respectively at a C-C bcp in ethane, benzene and ethene, respectively.
- *Energy densities at the bcp* determined by the one-electron density matrix (as opposed to just the density, its diagonal element) are used to summarize the mechanics of the interaction. Theory defines a potential energy density $\mathcal{V}(\mathbf{r})$, the virial field, (Eq.2.50) that is negative everywhere and which integrates to the total potential energy of the molecule, the sum $G(\mathbf{r})+\mathcal{V}(\mathbf{r})$ yielding the energy density $E(\mathbf{r})$ that integrates to $E_e(\Omega)$. The magnitude of the virial field exhibits the same topology as does the electron density. This observation has the important implication that every bond path linking a pair of nuclei is mirrored by a virial path, a line connecting the same nuclei along which the potential energy density is maximally attractive. (*Keith et al., 1996*) The local statement of the virial theorem:

$$\left(\frac{\hbar^2}{4m}\right)\nabla^2\rho(\mathbf{r}) = 2G(\mathbf{r}) + \mathcal{V}(\mathbf{r}) \tag{2.60}$$

ties the kinetic and potential energy densities to the topology of the density where $G(\mathbf{r})>0$ and $\mathcal{V}(\mathbf{r})<0$. Thus interactions for which $\nabla^2\rho_b < 0$ are dominated by a local lowering of the potential energy, while those for which $\nabla^2\rho_b > 0$ are dominated by a local excess in the kinetic energy as measured by the two to one ratio required for the local satisfaction of the virial theorem, which for the total system states $2T = -\mathcal{V}$. The energy density $H(\mathbf{r}) = G(\mathbf{r}) + \mathcal{V}(\mathbf{r})$ evaluated at a bcp compares

the kinetic and potential energies on an equal footing and is found to yield negative values for all interactions with significant sharing of the electron density, its magnitude reflecting the "covalent character" of the interaction. (*Cremer and Kraka, 1984*)

2.5.2 Atomic Properties

The average of a property A over an atomic basin $A(\Omega)$ is calculated from:

$$A(\Omega) = \left\langle \hat{A} \right\rangle_{\Omega} = \frac{N}{2} \int_{\Omega} d\tau \int d\tau' \left[\psi^* \hat{A} \psi + (\hat{A}\psi)^* \psi \right]. \tag{2.61}$$

Some examples of commonly computed atomic properties include:

- *Electron population* ($N(\Omega)$): obtained by setting $\hat{A} = \hat{1}$ in Eq. 2.53. This yields:

$$N(\Omega) = \int_{\Omega} \rho(\mathbf{r}) d\mathbf{r} \,. \tag{2.62}$$

- *Net atomic charge* ($q(\Omega)$): calculated by subtracting $N(\Omega)$ from the nuclear charge Z_{Ω}

$$q(\Omega) = Z_{\Omega} - N(\Omega) \,. \tag{2.63}$$

- *Atomic volume* ($v(\Omega)$): the space bounded by the intersection of the zero-flux surface and a chosen isodensity envelope. In this thesis, an isodensity of $\rho = 0.001$ au is used unless otherwise stated.

- *Schrödinger kinetic energy* ($K(\Omega)$):

$$K(\Omega) = -\frac{\hbar^2}{4m} N \int_{\Omega} d\tau \int d\tau' \left[\psi \nabla^2 \psi^* + \psi^* \nabla^2 \psi \right]. \tag{2.64}$$

- *Gradient kinetic energy* ($G(\Omega)$):

$$G(\Omega) = \frac{\hbar^2}{2m} N \int_{\Omega} d\tau \int d\tau' \, \nabla_i \psi^* \cdot \nabla_i \psi. \tag{2.65}$$

- *Laplacian* ($L(\Omega)$): used as an indicator of the accuracy of atomic integration. Deviations from zero are a measure of the numerical integration error. It is computed from:

$$L(\Omega) = K(\Omega) - G(\Omega)$$

$$= -\frac{\hbar^2}{4m}\int_{\Omega} d\mathbf{r}\left[\nabla^2\rho(\mathbf{r})\right]$$

$$= -\frac{\hbar}{4m}\oint dS(\Omega,\mathbf{r})\ \nabla\rho(\mathbf{r})\cdot\mathbf{n}(\mathbf{r}) = 0 \tag{2.66}$$

- *Total atomic energy* ($E_e(\Omega)$): calculated from the virial theorem

$$E_e(\Omega) = -T(\Omega) = \frac{1}{2}\mathscr{V}(\Omega) \tag{2.67}$$

- *Atomic dipolar polarization* ($\boldsymbol{\mu}(\Omega)$): which is also known as the *first atomic electrostatic moment* is the atomic space average of the electronic position vector. It is a three-dimentional vector with components and magnitude defined in Eq. 2.68 and 2.69, respectively:

$$\boldsymbol{\mu}(\Omega) = \begin{pmatrix}\mu_x\\ \mu_y\\ \mu_z\end{pmatrix} = \begin{pmatrix}-e\int_{\Omega} x\rho(\mathbf{r})\,d\mathbf{r}\\ -e\int_{\Omega} y\rho(\mathbf{r})\,d\mathbf{r}\\ -e\int_{\Omega} z\rho(\mathbf{r})\,d\mathbf{r}\end{pmatrix} = -e\int_{\Omega}\mathbf{r}_{\Omega}\,\rho(\mathbf{r})\,d\mathbf{r}\,, \tag{2.68}$$

$$\left|\boldsymbol{\mu}(\Omega)\right| = \sqrt{\mu_x^{\,2}+\mu_y^{\,2}+\mu_z^{\,2}}\,, \tag{2.69}$$

with the origin for the vector $\mathbf{r}_{\Omega}$ at the nucleus of atom Ω, i.e. $\mathbf{r}_{\Omega} = \mathbf{r} - \mathbf{R}_{\Omega}$, $\mathbf{r}$ being the electronic coordinates and $\mathbf{R}_{\Omega}$ the nuclear coordinates of atom Ω. The first moment measures the polarization of the charge density, that is to say the departure from sphericity of the electron density. Dipolar polarization of molecules and of extended systems such as crystals or polymers will be discussed in detail in chapter 7.

- *Atomic quadrupolar polarization* ($\mathbf{Q}(\Omega)$): which is also known as the *second atomic electrostatic moment* is a symmetric traceless tensor when defined:

$$\mathbf{Q}(\Omega)=\begin{pmatrix} Q_{xx} & Q_{xy} & Q_{xz} \\ Q_{yx} & Q_{yy} & Q_{yz} \\ Q_{zx} & Q_{zy} & Q_{zz} \end{pmatrix}=-\frac{e}{2}\begin{pmatrix} \int_{\Omega}(3x_{\Omega}^2-r_{\Omega})\rho(\mathbf{r})\,d\mathbf{r} & 3\int_{\Omega}x_{\Omega}y_{\Omega}\rho(\mathbf{r})\,d\mathbf{r} & 3\int_{\Omega}x_{\Omega}z_{\Omega}\rho(\mathbf{r})\,d\mathbf{r} \\ 3\int_{\Omega}y_{\Omega}x_{\Omega}\rho(\mathbf{r})\,d\mathbf{r} & \int_{\Omega}(3y_{\Omega}^2-r_{\Omega})\rho(\mathbf{r})\,d\mathbf{r} & \int_{\Omega}y_{\Omega}z_{\Omega}\rho(\mathbf{r})\,d\mathbf{r} \\ 3\int_{\Omega}z_{\Omega}x_{\Omega}\rho(\mathbf{r})\,d\mathbf{r} & \int_{\Omega}z_{\Omega}y_{\Omega}\rho(\mathbf{r})\,d\mathbf{r} & \int_{\Omega}(3z_{\Omega}^2-r_{\Omega})\rho(\mathbf{r})\,d\mathbf{r} \end{pmatrix}, \quad (2.70)$$

where again the origin is at the nucleus. If the atomic electron density possesses spherical symmetry, then $\int_{\Omega}x_{\Omega}^2\rho(\mathbf{r})\,d\mathbf{r}=\int_{\Omega}y_{\Omega}^2\rho(\mathbf{r})\,d\mathbf{r}=\int_{\Omega}z_{\Omega}^2\rho(\mathbf{r})\,d\mathbf{r}=\frac{1}{3}\int_{\Omega}r_{\Omega}^2\rho(\mathbf{r})\,d\mathbf{r}$, and $Q_{xx}=Q_{yy}=Q_{zz}=0$. Thus, the quadrupole moment is another measure of the deviation of the atomic electron density from spherical symmetry. For example, if a diagonal component of **Q** is <0, this indicates that the electron density is concentrated along that axis and *vice versa*. It is always possible to find a coordinate system such that the original tensor in Eq. 2.70 (**Q**(Ω)) is diagonalized ($\mathcal{Q}(\Omega)$) that is to say that the off-diagonal components vanish. The diagonalization of **Q**(Ω) corresponds to a rotation of the original coordinate system. The diagonalized quadrupole tensor corresponding to Eq. 2.70 is written:

$$\mathcal{Q}(\Omega)=\begin{pmatrix} \mathcal{Q}_{x'x'} & 0 & 0 \\ 0 & \mathcal{Q}_{y'y'} & 0 \\ 0 & 0 & \mathcal{Q}_{z'z'} \end{pmatrix}, \quad (2.71)$$

where $\mathcal{Q}_{x'x'}$, $\mathcal{Q}_{y'y'}$, and $\mathcal{Q}_{z'z'}$ are the principal values of the quadrupole moment with respect to the principal (rotated) axes x', y', and z' – axes which would correspond to axes of symmetry if they exist in the electron density distribution (henceforth, the primes will be dropped for simplicity).

The traceless property of the tensor defined in Eq. 2.70 (or in its diagonalized form, Eq. 2.71) is a consequence of the equality:

$$r_{\Omega}^2 = x_{\Omega}^2 + y_{\Omega}^2 + z_{\Omega}^2, \quad (2.72)$$

which is always true in any coordinate system. Therefore:

$$(Q_{xx}+Q_{yy}+Q_{zz}) = (\mathcal{Q}_{xx}+\mathcal{Q}_{yy}+\mathcal{Q}_{zz}) = 0 \quad (2.73)$$

and only five independent components completely specify $\mathbf{Q}(\Omega)$ in the original coordinate system and only two are sufficient to specify its diagonalized form $\mathbf{\Theta}(\Omega)$.

The magnitude of the quadrupole moment is defined as: (*Gray and Gubbins, 1984*)

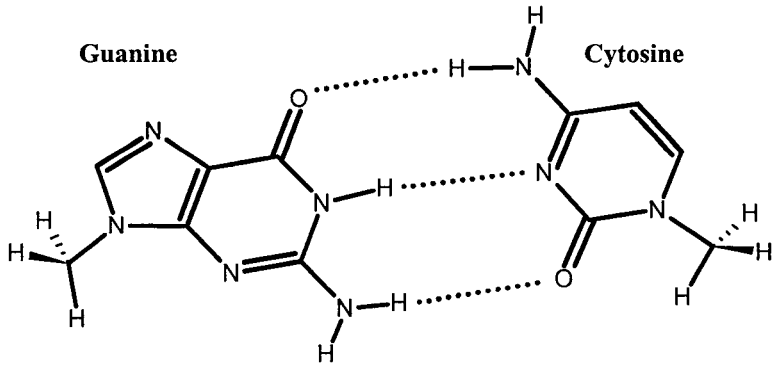
$$|\mathbf{Q}| = \sqrt{\frac{2}{3}\left(Q_{xx}^2 + Q_{yy}^2 + Q_{zz}^2\right)} = \sqrt{\frac{2}{3}\left(\Theta_{xx}^2 + \Theta_{yy}^2 + \Theta_{zz}^2\right)}. \qquad (2.74)$$

2.5.3 The Laplacian of the Electron Density

The Laplacian of the density $[\nabla^2\rho(\mathbf{r})]$ measures the local curvature of the electron density in all of its three dimensions (Eq.2.4). Where the Laplacian is positive, the density is locally depleted and the electron density is expanded relative to its average distribution; where it is negative the density is locally concentrated and the density is tightly bound and compressed above its average distribution. A local charge concentration is a Lewis base (nucleophile), while a local charge depletion is a Lewis acid (electrophile). The Laplacian reproduces the spherical shell structure of isolated atoms in terms of alternating shells. The spherical nodes in the Laplacian are envelopes bounding regions of density depletion or concentration. The outer shell of charge concentration, which is followed by a shell of charge depletion extending to infinity, is called the valence shell charge concentration (VSCC). When an atom is involved in bonding the spherical symmetry of the VSCC is broken and it may become punctured. A chemical reaction corresponds to the combination of a "lump" in the VSCC of the base with the "hole" in the VSCC of the acid. The reactive surface of a biological molecule serves as a "complementary template" for its receptor: where there is a region of charge depletion (holes) in the reactive surface of the molecule there should be a region of charge concentration (lumps) in the receptor, and vice versa.

Figure 2.2 displays the reactive surface envelope of a guanine-cytosine DNA base pair which is overlaid on the van der Waal's outer density envelope. Holes in the Laplacian represent the sites susceptible to nucleophilic attack while the lumps indicate sites of electrophilic attack. The sites

indicated by the arrow correspond exactly to the sites of attack and adduct formation by chemical carcinogens. (See for example: (*Archer, Tannock and Hill, 1987*) and references therein).



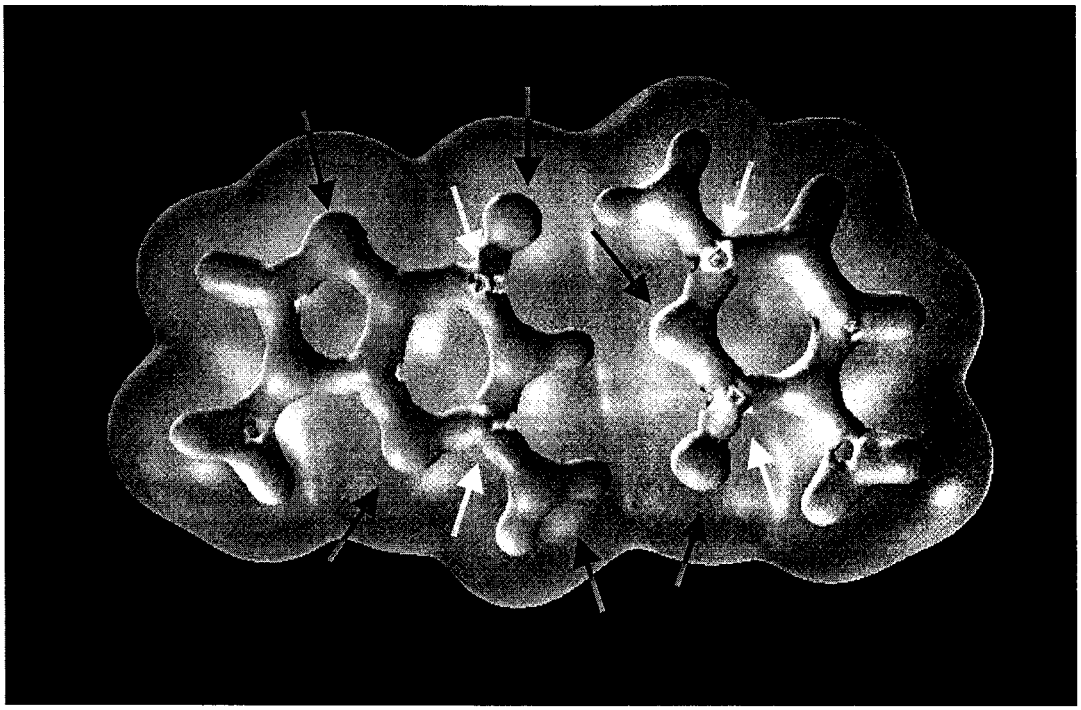
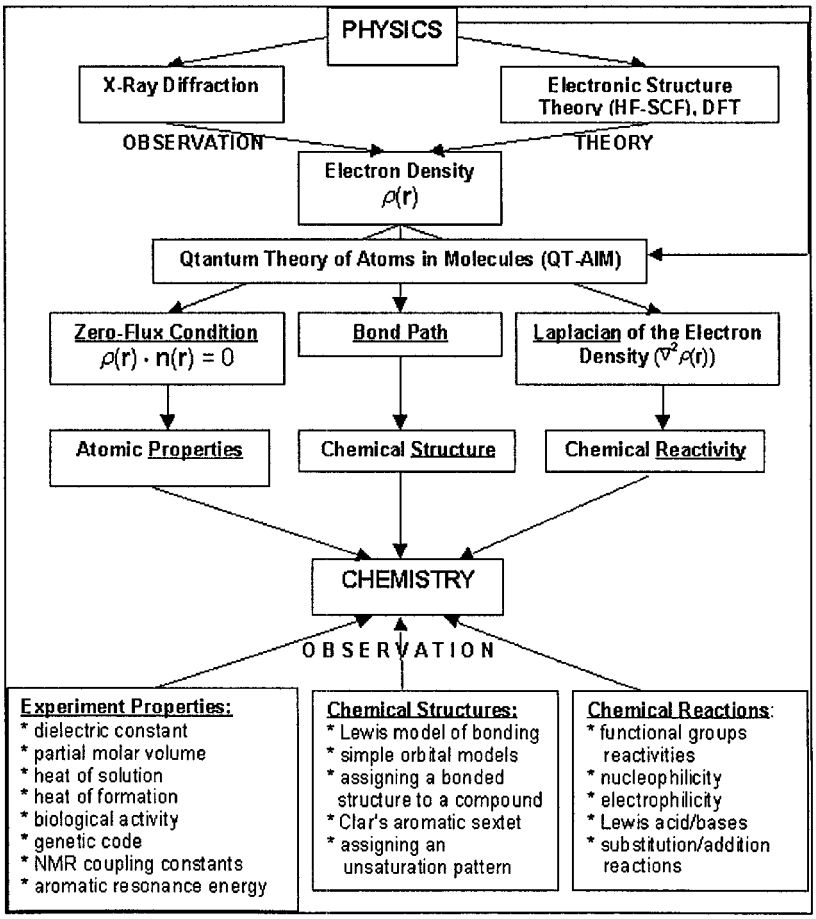


Figure 2.2 A display of the zero-Laplacian surface also known as the reactive surface (inner solid/opaque envelope) of a guanine-cytosine DNA base pair. The van der Waal's surface corresponding to the isodensity envelope of $\rho(\mathbf{r})$=0.001 au (outer transparent envelope) is overlaid to provide a sense of the spatial extent of the reactive surface. Holes in the Laplacian (regions of charge depletion, are indicated by yellow arrows, and lumps (regions of charge concentration) are indicated by red arrows.

2.6 Closing Remarks

The atoms of theory do coincide with the atoms of topology and the atoms of topology do coincide with the atoms of chemistry. Theory provides a physical basis for the philosopher's "common sense": that if an atom is a "thing" – i.e. possesses a set of well defined properties determined by its form in space – it ought to be bounded. The bounded atoms of theory are the atoms of chemistry because they recover the chemist's empirical observation: that atoms and functional groups contribute characteristic additive and essentially transferable properties to the molecules they constitute. Scheme 2.1 expresses the relationships between some experimental quantities which will be dealt with in the following chapters (bottom of the diagram) to aspects of theory discussed in this part of the thesis.

Scheme 2.1



PART II:

COMPREHENSIVE STUDY OF THE FREE AMINO ACIDS

Chapter 3

Effects of Proton Transfer and of Conformation on Geometric, Atomic and Bond Properties of the Genetically-Encoded Amino Acids

3.1 Statement of the Problem

One of the most important concepts in chemistry is that of a functional group, the idea that a linked grouping of atoms can exhibit a set of characteristic geometric and chemical properties. These properties enable one to detect the presence of a group in a molecule and to predict the effect its presence has on the molecule's chemistry. Nowhere is this concept more useful than in understanding and cataloguing the chemistry of biological macromolecules which consist of polymeric combinations of a small number of building blocks: amino acids, nucleotides, simple sugars and phosphates. We know from experiment that the geometric parameters - bond lengths and bond angles - of the individual members of these sets and combinations of them can be assigned with some certainty when present in a macromolecule, but our ability to predict the chemical consequences of their presence is less precise.

The quantum theory of atoms in molecules (QT-AIM) recovers the basic empirical cornerstone of chemistry: that atoms and functional groups possess characteristic and additive

properties that in many cases exhibit a remarkable transferability between different molecules. As a result, the theory enables the theoretical synthesis of a large molecule and the prediction of its properties by joining fragments which are predetermined as proper open systems (chapter 6). The present part of the thesis explores this possibility for polypeptides by determining the transferability of the building blocks: the amino acid residues.

The transferability of group properties requires transferability of the electron density $\rho(\mathbf{r})$ which in turn requires the transferability of the geometric parameters. This chapter demonstrates that the geometric parameters are conformation-insensitive for a representative amino acid, leucine, and that the atomic and bond properties exhibit a corresponding transferability. The effects of hydrogen bonding are determined and a set of geometric conditions for the occurrence of such bonding is identified. The effects of transforming neutral leucine into its zwitter-ionic form on its atomic and bond properties are shown to be localized primarily to the sites of ionization. This chapter is based on the paper: (*Matta and Bader, 2000*).

3.2 Functional Groups and the Properties of Macromolecules

Atoms defined as proper open systems and described in chapter 2 are identified with the atoms of chemistry because: (i) their properties are characteristic and additive, summing to yield the corresponding values for the total system, (ii) they are as transferable from one system to another as are their forms in real space and, (iii) the surface that defines an atom in a molecule has the important property of maximising the transferability of its form and hence its properties between molecules. Point (ii) is a consequence of the truism that two identical pieces of matter exhibit identical properties, coupled with the realisation that the form of matter is determined by its spatial distribution of charge. Thus to both understand and utilise the concept of a functional group through theory the group must be defined in real space by its charge distribution, which is what physics does.

The contributions of a group to the total properties of a molecule are accessible to experimental measurement in those instances where the properties of the group, as well as being additive, are transferable between molecules. It is these experimental group contributions to the volume, energy, polarizability and magnetic susceptibility that are recovered by theory, providing the ultimate test of QT-AIM. (*Bader, 1990*) It is not generally appreciated just how high a degree of transferability the form of a group and hence its properties can exhibit. The methyl and methylene groups of saturated hydrocarbons played a historically important role in the development of the notion group contributions to molecular properties, with early experiments providing evidence of the additivity of their volumes, polarizabilities, heats of formation and magnetic susceptibilities. A theoretical calculation of the properties of such groups defined as proper open systems at various computational levels with and without electron correlation, demonstrate that their charge distributions and hence their geometries are essentially superimposable through a homologous series of hydrocarbons, with their electron populations differing by less than 0.001e and their energies by less than one kcal/mole, the latter result being particularly remarkable considering that the total energy of a methylene group is of the order of 25,000 kcal/mole.

This part of the thesis reports the characteristic properties of the atoms and functional groupings of atoms found in the genetically-encoded amino acids. The tabulated atomic and group properties, including a set of atomic multipole moments, should facilitate the modelling of the behaviour of a polypeptide and its intermolecular interactions. The atomic properties obtained in this study represent the set with the maximal degree of transferability. In the present chapter, leucine is used to study the dependence of the geometric, bond, and atomic properties on molecular conformation and ionization-state. This is a necessary prelude to the remainder of this part which tabulates and compares the geometric, atomic and bond properties for each of the genetically-encoded amino acids in a stable conformation.

3.3 Modelling the Molecular Electrostatic Potential (MEP)

In the past the modelling of the molecular force field was frequently restricted to a set of atom-based point charges (monopoles). These consist of charges derived by fitting them to an empirical force field that reproduces the crystal geometry; (*Hagler et al., 1977*) or of Mulliken atomic charges obtained from SCF calculations; (*Clementi et al., 1977; Sordo et al., 1987*) or of overlap normalized partial charges obtained from CNDO/2 calculations; (*Momany et al., 1975*) or of charges derived by a fitting of an independently derived potential. (*Weiner et al., 1986*) An interatomic transfer of electronic charge results in a polarization of the atomic densities in a direction counter to the charge transfer and the resulting multipolar field cannot be adequately represented by a set of monopoles. More recently, there has been a resurgence in methods used to represent the electrostatic field in terms of a set of multipole moments. In one approach, cumulative atomic multipole moments (CAMMs) are derived from the corresponding set of molecular multipole moments, (*Sokalski and Poirier, 1983*) which were used to obtain libraries of amino acid atomic multipole moments. (*Sokalski et al., 1989*) Price *et al.* (*Price et al., 1991*) use the distributed multipole analysis (DMA) of Stone (*Stone, 1981; Stone and Alderton, 1985*) in the determination of the multipole representations of the charge densities of blocked residues of the naturally occurring amino acids. These authors use standard force field bond lengths and bond angles to fix the molecular geometries. The results reported in the present part of this thesis represent the first set of moments and other atomic properties determined at fully optimized geometries, and in an entirely non-empirical manner, using QT-AIM.

With the exception of a few cases, (*Cooper and Stutchbury, 1985; Kosov and Popelier, 2000a; Messer, 1977; Popelier, 1996*) all of the previously defined charges and multipole moments have been expressed in terms of spatially overlapping contributions such as atomic centred basis functions, as opposed to the disjoint partitioning of space used to define the atomic multipole moments in Eqs.2.63, 2.68, 2.69, and 2.70. Stone's DMA can be implemented using moments defined for spatially disjoint regions or in terms of a set of many-centred basis functions, but Stone and Alderton

argue that the latter procedure should in general produce faster convergence to the true potential since atomic boundaries are generally non-spherical. This conclusion is, however, not borne out in practice as shown in the work of Cooper and Stutchbury in the modelling of van der Waals complexes and more recently in Popelier's direct comparison of the electrostatic potential for a number of molecules using the multipole moments obtained from the disjoint partitioning in QT-AIM and a basis set representation of DMA. Popelier concludes that the atomic multipoles obtained from QT-AIM reproduce the exact *ab initio* electrostatic potential to the required accuracy. (*Kosov and Popelier, 2000a; Kosov and Popelier, 2000b; Popelier et al., 2001*)

As important as the ability of the QT-AIM multipole moments to reproduce an electrostatic potential field, is their insensitivity to a change in the basis set, a property not shared by moments expanded in terms of a set of basis functions. Stone and Alderton state that the "penalty" one pays in using a set of basis functions in the implementation of DMA is the sensitivity of the resulting moments to the choice of basis set. An even more stringent test of the stability of QT-AIM moments to the choice of basis set is provided by their use in the calculation of distributed polarizabilities both static and frequency dependent, at the coupled perturbed Hartree-Fock level of theory. (*Hättig et al., 1997; Stone et al., 1996; Ángyán et al., 1994*) These moments are necessary in the determination of intermolecular interaction potentials derived from perturbation theory. The authors found that basis set partitioning schemes yielded polarizability parameters so unstable with respect to basis set extension as to be unusable. They demonstrated two important advantages that result from the use of the disjoint atoms defined in QT-AIM for the determination of the induced dipole and charge flow contributions to the atomic polarizabilities: (i) the atomic contributions to the moments maintain a remarkable stability with respect to basis set extension, exhibiting a stability equal to that exhibited by the total polarizability components and (ii), in the case of the n-alkanes, the atomic contributions yield methyl and methylene group polarizability parameters that are transferable between any members of the series.

It is readily shown that a quantum mechanical definition of any property for an open system requires that it be bounded in real space and that the failure of the previous methods to meet this theoretical requirement is their principal shortcoming. (*Bader, 1998a*) Any partitioning of a molecule into spatially overlapping atoms eschews the very property of the density that forms the core of the concept of an atom in a molecule - namely, that an atom exhibits characteristic properties in spite of changes in its neighbours, a property that is faithfully recovered by the atoms of theory. However, in the case of contributions defined by overlapping atoms, the density assigned to atom A when bonded to B contains contributions from the basin of B while when bonded to a different atom C, it contains contributions from the basin of atom C. In addition to these differences, the density within the basin of A also differs in the two cases, as it loses density to B in one case and to C in the other. Thus the density distribution assigned to atom A is markedly different in each of the two cases and A is assigned different properties. Models of overlapping atoms or overlapping atomic contributions, as used in the definition of Mulliken charges, do not account for the essential observation that atoms and functional groups can exhibit characteristic properties in spite of changes in their environments. Mulliken charges can appear transferable (but still basis set dependent) between molecules in those cases where the environments of the atoms remain unchanged, a situation encountered in the repeating units of biological molecules. However, even in these instances the charges are strongly dependent on conformation.

The properties of the Cα atom in an amino acid exhibit a remarkable transferability as the side-chain is changed through the twenty amino acids (chapter 5). The average charge, the magnitude of the dipole, and the atomic energy of this atom are 0.581 (0.015), 0.499 (0.023), and -37.497 (0.014) au respectively, values in parentheses being standard deviations (σ). Other atomic properties exhibit similar transferability, as evidenced by the small σ, and will be discussed in fuller detail in chapter 5. The atomic charges assigned on the basis of Mulliken population or DMM do not exhibit the same degree of transferability. For example, Price *et al.* (*Price et al., 1991*) report Mulliken net charges (q) for the CαH group for 14 residues blocked by CH_3CO- and -$NHCH_3$ and calculated using a standard

set of geometries. A comparison of these charges with those obtained using QT-AIM at the optimized geometries (values in brackets) show that the latter are more transferable: av. q = 0.079e (0.577e), σ = 0.044e (0.023e), max(q) - min(q) = 0.169e (0.087e).

In this chapter, the effects of conformation of the side chain on the total energy, geometric parameters, bond, and atomic properties of non-zwitter-ionic[1] leucine are studied. The changes in these properties incurred on the formation of the zwitter-ion are also investigated. In chapters 4 and 5, the genetically-encoded amino acids are studied in their neutral forms. The neutral forms are chosen to eliminate the effects of charge separation on Cα since the environment of this atom in a polypeptide does not involve formal charge separation.

3.4 The Conformational Problem

Amino acids are floppy molecules which can fall into numerous stable conformations as they explore their respective potential energy surfaces (PES). Each well on a PES corresponds to a local minimum. The lowest energy local minimum is known as the "global minimum". There is no systematic method at the present that guarantees locating the global minimum except for the smallest molecules. The global minimum is located usually by geometry optimization after generating different starting rotamers either by: (i) random statistical sampling of the geometric parameter space or (ii) systematically rotating groups around single bonds to get different staggered conformers. The number of rotamers scale very rapidly with molecular size, prohibiting an exhaustive *ab initio* PES scan for most molecules of biological interest.

The rotamers of the side chains of the amino acids have been repeatedly shown to be separated by low energy barriers on the PES. In a conformational study at the MP2/6-31+G*//HF/6-31G* level, Gronert and O'Hair obtained 10 energy minima for alanine, 51 for serine, and 42 for

---

[1] Henceforth non-zwitter-ionic amino acids will be referred to as "neutral" amino acids (**n**) which is to be distinguished from amino acids with non-ionized side chains.

cysteine, all of which are relatively small amino acids. (*Gronert and O'Hair, 1995*) More recently, the same group found 26 conformations of valine at the same level of theory. (*Shirazian and Gronert, 1997*) In both studies, the minima of each molecule are found to span an energy range of about 10 kcal/mol and the results indicate that many of these have energies near the global minimum. In the case of glycine, seven local minima on the potential energy surface were identified using a 4-31G basis set spanning a range 10.7 kcal/mol wide. (*Ramek and Cheng, 1992*) This difference between the global minimum and the least stable conformer in glycine is 0.006% of the total energy of the global minimum reported in their work. The effect of conformation on the total SCF energy has been found to be small in the present work as well: nine neutral leucine rotamers span a range of 5.1 kcal/mol while their zwitter-ionic counterparts span a range of 3.9 kcal/mol.

The small differences between the energies of rotamers of the side chains account for their floppiness and for their ability to populate several rotameric conformations even in the solid state. For example, the side chains of crystalline Leu-enkephalin are shown to co-exist in four different conformations while the peptide backbones adopt essentially the same conformation. (See for example: (*Karle and Karle, 1983*)). In protein crystallography, the electron density of some side chains cannot be determined presumably as a result of the numerous conformations that they adopt in the crystal. (*Creighton, 1983*) Moreover, the gas-phase global minimum or relative energies between conformers are generally different than in the water-solvated phase. (See for example: (*Watanabe et al., 1997*).

It is concluded that the location of the global minimum is not essential for a study of the geometric parameters, atomic, and bond properties of the amino acids using QT-AIM. A reasonable local minimum is anticipated to yield properties which are accurate enough when compared to experiment. It is one of the main goals of this chapter to demonstrate this conformational insensitivity and to show that one can rely upon geometric, bond and atomic properties of a reasonable local minimum. This is accomplished through a comparison of these properties in a number of leucine rotamers.

3.5 Computational Methods

3.5.1 Choice of Basis Set

Unconstrained geometry optimizations were performed at the single-determinant Hartree-Fock level using a 6-31+G* basis set whose suitability for such optimizations was demonstrated in a benchmark paper by Head-Gordon *et al.* (*Head-Gordon et al., 1991*) The wavefunctions were then recalculated at these geometries using a 6-311++G** basis set for the determination of the density and the bond and atomic properties, a similar strategy having been used in a study of hydrogen bonding in related molecules. (*Popelier and Bader, 1992*) The diffuse functions present in the large basis set are necessary to describe the density of anionic forms of aspartate and glutamate, and were applied to all amino acids in this study for consistency. (See pages xx-xxi).

3.5.2 Calculations

For a fixed conformation of the -CαH(NH_2)COOH group, the leucine side chain can exist in nine (3×3) staggered rotamers around the Cα-Cβ and Cβ-Cγ bonds. A starting geometry for leucine was fully optimized at the Hartree-Fock level implemented in Gaussian94 (*Frisch et al., 1995*) using the 3-21G basis set, followed by a second optimization using a 6-31+G* basis set. This optimized rotamer was used to generate the starting geometries of the remaining eight rotamers by varying the necessary dihedral angles. The new rotamers were optimized directly at the 6-31+G* level. A single point calculation was performed at each of these nine geometries to obtain 6-311++G**//6-31+G* densities. The optimized structures of the nine neutral rotamers were then modified to the corresponding zwitter-ionic forms and optimized using a 6-31+G* basis set followed by a single-point calculation using a 6-311++G** basis set. The AIMPAC suite of programs (*Biegler-König et al., 1982*) was used to calculate the bond and atomic properties from the calculated electron densities. Bond paths were followed to their nuclear termini to ensure correct assignment of the two contributing atoms in the case of hydrogen bonded interactions.

As discussed in chapter 2, deviations in $L(\Omega)$ from zero provide a stringent test of the numerical accuracy of the atomic integrations (Eq. 2.66). All integrations reported in this chapter have $|L(\Omega)| < 5.0 \times 10^{-3}$. The average $|L(\Omega)|$ for all atoms (264 atoms) is $\approx 7.0 \times 10^{-4}$. At this accuracy, the sums of the atomic populations $N(\Omega)$ for the leucine rotamers differ on average by 0.005e from the total number of electrons N. The sum of the atomic energies $E(\Omega)$ for each conformer differs from the SCF total energy on average by 0.7 kcal/mol.

3.6 Results and Discussion

3.6.1 Effects of Conformation and of Tautomerization (Proton Transfer) on Energy and on geometric Parameters

Figure 3.1(a) depicts the geometries of the lowest energy rotamers of both neutral and zwitter-ionic leucine along with the numbering scheme, and Fig 3.1(b) displays the remaining eight optimized minimum energy structures of neutral leucine (their optimized zwitter-ionic counterparts are not shown as they adopt the same side-chain conformation).

The geometry optimizations described in the present work provide nine local minima (rotamers) for the neutral set, designated as (**n**), and nine corresponding minima for the zwitter-ionic set, designated as (**z**). The rotamers are labeled **A** to **I** in order of increasing energy, an ordering which is identical for both the **z**- and **n**-set. The relative energies and molecular dipole magnitudes of the rotamers are given in table 3.1 and their respective torsion angles in table 3.2.

The optimized **z**-tautomers exhibit side-chain geometries that are essentially the same as their **n** counterparts, justifying the use of the same label. The nine minima of **n**-leucine span an energy range of 5 kcal/mol, while the **z**-rotamers span an energy range of 4 kcal/mol. The change in the total energy upon the migration of the proton from the neutral carboxylic group to the neutral amino group to form the zwitter-ion will be referred to as energy of tautomerization and is given by:

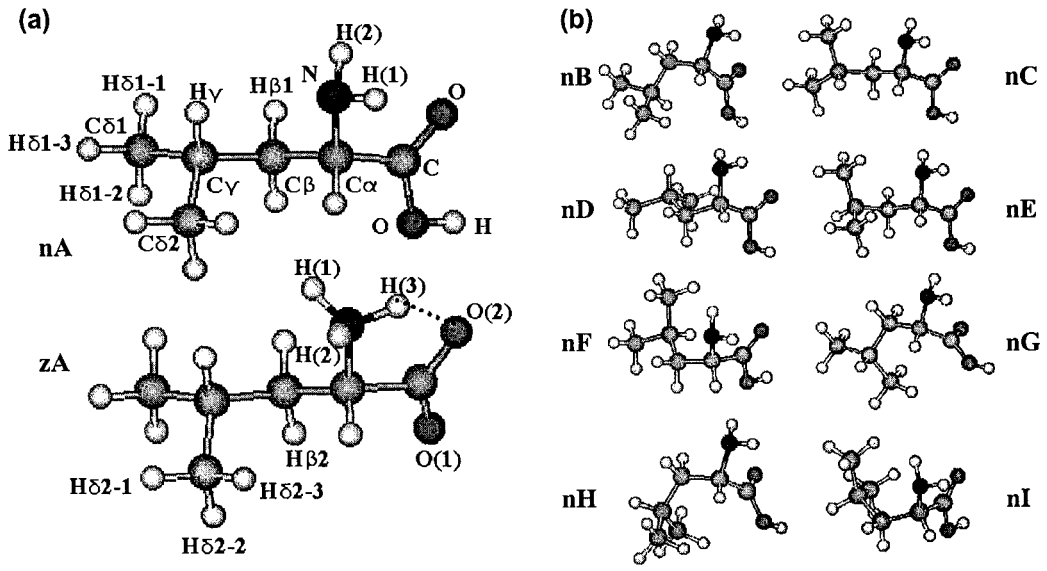


Figure 3.1 (a) Optimized geometries and numbering system for rotamers **nA** and **zA** (the global minima of the neutral and zwitter-ionic set). The numbering of the other rotamers follow the same conventions (summarized in the Appendix). Note the close geometrical similarity of the side chains of the two tautomers. **(b)** The remaining rotamers of the neutral (**n**)-set: **nB-nI**. Rotamers of the zwitter-ionic (**z**)-set are not shown, since the geometry of their side-chains are similar to the corresponding rotamers of the **n**-set.

$$E_\tau = E_z - E_n \ , \tag{3.1}$$

where E_n and E_z are the total energies of the **n** and the **z** forms of the amino acid with the same conformation. The average E_τ over the nine rotamers is +25.5(±0.6) kcal/mol showing that the **z**-form is less stable than the **n** counterpart in the gas-phase (value in parenthesis is σ). This destabilization is more than compensated for by the energy of aqueous solvation estimated to be about 150 kcal/mol (*Kebarle, 1976*) for the solvated carboxyl and quaternized amino groups. The small value of σ indicates that E_τ is rather independent of the conformation of the side chain. For comparison, E_τ was calculated for glycine at the same level of theory and found to be +28.3 kcal/mol. These results show that the replacement of one of the hydrogen atoms bonded to the Cα of glycine by the side-chain of leucine reduces E_τ by less than 3 kcal/mol, that is by less than 10%.

The side-chain conformations of rotamers **A** and **B** are in good agreement with the x-ray crystallographic structure of leucine (see table 3.2). The crystallographic side-chain dihedral angles differ on the average by 6° from their optimized counterparts, which is a relatively small difference

which could be accounted for, at least in part, by crystal packing effects. The crystallographic bond lengths are generally slightly longer than the calculated values (since the latter refer to the vibration-less molecule) while the calculated bond angles and lengths are in good agreement with experiment (tables 3.3-3.6).

The average value for each of the geometric parameters over the set of rotamers is given in tables 3.3-3.6 along with the σ The optimized bond lengths and angles show a marked constancy across rotamers of both **n** and **z** sets, σ ranging from 0.000 to 0.005Å for bond lengths and from 0.0 to 2.5° for bond angles for both the **n** and the **z** sets. Comparison of the optimized geometric parameters of the side chains of each **n** rotamer with its **z** counterpart shows that ionization has little effect and that the geometric parameters are almost constant within each set and from one set to the other. This transferability of bond lengths and angles leads one to anticipate an underlying transferability of the electron density and hence of the atomic and bond properties.

Table 3.1 Molecular dipole moment magnitudes |μ| (debyes) and relative energies (kcal/mol) of the neutral (**n**) and zwitter-ionic (**z**) leucine rotamers

Rotamer	|μ| **n**	Rel.*E* **n***	|μ| **z**	Rel. *E* **z***
A	1.279	0	10.521	0
B	1.581	0.7	9.923	0.7
C	1.410	1.9	10.442	0.8
D	1.390	2.0	10.114	1.3
E	1.412	2.5	10.470	1.5
F	1.483	2.7	10.007	1.6
G	1.369	3.1	10.014	3.1
H	1.416	3.9	10.057	3.8
I	1.520	5.1	10.018	3.9
E (au) *		-439.1015		-439.0599

* Total SCF energy in au of the global minimum.

Table 3.2 Torsion angles (°) of the neutral (**n**) and zwitter-ionic (**z**) leucine rotamers and the corresponding crystallographic values*

Rotamer	ψ(**n**)	ψ(**z**)	χ_1(**n**)	χ_1(**z**)	χ_{21}(**n**)	χ_{21}(**z**)	χ_{22}(**n**)	χ_{22}(**z**)
A	-17.1	3.9	-58.8	-65.3	175.2	172.9	-61.6	-63.8
B	-40.4	-14.9	-177.3	-178.8	62.4	60.3	-174.5	-176.5
C	-18.7	1.8	-67.8	-81.4	81.5	-173.7	-155.2	62.4
D	-1.3	6.5	54.6	50.7	66.0	58.4	-170.8	-177.9
E	-17.4	2.6	-63.9	-78.0	-50.2	-63.5	77.8	63.9
F	-12.5	-3.8	80.6	74.8	175.4	172.2	-61.2	-64.3
G	-77.3	-20.9	-157.3	179.3	174.4	141.9	-61.7	-94.9
H	-57.7	-23.9	-168.1	-177.7	-70.9	-81.6	57.1	46.6
I	-11.2	-8.2	70.3	65.9	-58.5	-61.9	70.8	67.0
Cryst.**B1**		-26.8(1)		-176.8(1)		64.6(1)		-174.3(1)
Cryst.**B2**		-32.3(1)		-170.0(1)		71.0(2)		-166.9(1)
Enk.a **A1**			-64(2)		173(2)		-67(2)	
Enk.b **A2**			-62(2)		165(2)		-69(2)	
Enk.d **A3**			-80(2)		179(2)		-63(2)	

* Crystal **B1** and **B2** refer to two closely related but crystallographically independent low-temperature x-ray structures (*Görbitz and Dalhus, 1996b*) that correspond to the **B** rotamer in the present study. Enk. **A1-A3** refer to the geometries of the leucine residue in three crystallographically independent x-ray crystallographic structures of Leu-enkephalin, small letters refer to labeling in the experimental paper. (*Karle and Karle, 1983*) The conventions for labelling the geometrical parameters are described in appendix 3.1.
This footnote also serve as one for tables 3.3-3.6.

Table 3.3 Bond lengths (Å) of the -CαH(NH_2)COOH group in the neutral (**n**) and zwitter-ionic (**z**) leucine rotamers and corresponding crystallographic values

		Cα-C	Cα-N	Cα-H	C=O	C-OH	O-H	C=O(1)	C=O(2)	NH(1,2)*	N-H(3)
n	av	1.524	1.448	1.084	1.190	1.331	0.953			1.002	
	σ	0.001	0.005	0.002	0.000	0.000	0.000			0.001	
z	av	1.569	1.515	1.080				1.211	1.244	1.005	1.043
	σ	0.003	0.001	0.001				0.001	0.001	0.001	0.005
	Cryst.B1	1.530(2)	1.494(1)					1.255(1)	1.258(1)	0.84(2)	0.88(2)
	Cryst.B2	1.534(2)	1.491(2)					1.252(1)	1.263(1)	0.83(3)	0.94(2)

* Bond length N-H(1,2) is an average of the averages of N-H(1) and N-H(2).

Table 3.4 Bond lengths (Å) of the side-chain of the neutral (**n**) and zwitter-ionic (**z**) leucine rotamers and corresponding crystallographic values*

		Cα-Cβ	Cβ-H	Cβ-Cγ	Cγ-H	Cγ-Cδ1	Cδ1-H	Cγ-Cδ2	Cδ2-H
n	av.	1.540	1.086	1.540	1.088	1.533	1.086	1.533	1.086
	σ	0.004	0.001	0.003	0.001	0.001	0.000	0.001	0.000
z	av.	1.529	1.087	1.542	1.088	1.533	1.086	1.533	1.086
	σ	0.004	0.002	0.003	0.003	0.001	0.001	0.001	0.001
	Cryst. **B1**	1.535(1)		1.529(2)		1.525(2)		1.530(2)	
	Cryst. **B2**	1.533(2)		1.531(2)		1.524(3)		1.520(2)	
	Enk.a **A1**	1.540(30)		1.528(30)		1.560(30)		1.601(30)	
	Enk.b **A2**	1.521(30)		1.478(30)		1.423(30)		1.561(30)	
	Enk.d **A3**	1.526(30)		1.478(30)		1.607(30)		1.591(30)	

* For Cβ-H and Cδ-H bonds, the values given are averages of averages.

Table 3.5 Bond angles (°) involving the terminal Cα(NH_2)COOH groups of the neutral (**n**) and zwitter-ionic (**z**) leucine rotamers and corresponding crystallographic values

		O-C-O	O(1)-C-Cα	O(2)-C-Cα	O-C-Cα*	N-Cα-C
n	av.	122.2	112.7	125.2	118.9	112.1
	σ	0.1	0.4	0.4	0.0	1.4
z	av.	132.2	112.7	115.1	113.9	104.1
	σ	0.1	0.2	0.3	0.1	0.5
	Cryst. B1	124.7(1)	118.0(1)	117.3(1)	117.7	109.2(1)
	Cryst. B2	125.5(1)	117.8(1)	116.8(1)	117.3	107.8(1)

* Averages of O(1)-C-Cα and O(2)-C-Cα.

Table 3.6 Bond angles (°) of the side chains of the neutral (**n**) and zwitter-ionic (**z**) leucine rotamers and corresponding crystallographic values

		N-Cα-Cβ	C-Cα-Cβ	Cα-Cβ-Cγ	Cβ-Cγ-Cδ1	Cβ-Cγ-Cδ2	Cδ1-Cγ-Cδ2
n	av.	111.1	111.0	117.7	111.6	112.7	110.4
	σ	1.7	2.0	1.9	2.3	1.9	0.4
z	av.	111.2	114.8	117.7	112.4	111.7	110.5
	σ	2.5	1.4	1.7	1.8	2.3	0.4
	Cryst.B1	107.5(1)	112.0(1)	116.1(1)	111.3(1)	109.2(1)	109.6(1)
	Cryst.B2	107.8(1)	111.1(1)	115.1(1)	110.9(1)	109.8(3)	110.3(1)

3.6.2 The Effect of Conformation on Bond Properties

The presence of a bond path joining the nuclei of a pair of atoms provides a rigorous criterion that the two atoms are bonded to one another, a criterion that is applicable to all types of bonding. The properties of the electron and energy densities at the bcp characterize the bonding (See chapter 2). With the exception of hydrogen bonds, the amino acids exhibit shared interactions with $0.21 < \rho_b < 0.44$ au, $\nabla^2\rho_b < 0$ and $H(\mathbf{r}_b) < 0$. The value of $\nabla^2\rho_b$ for the polar C=O bond in the ionized leucine rotamers is slightly positive, $H(\mathbf{r}_b)$ however, remains negative. Bond properties are summarized in Table 3.7. Considerable accumulation of electronic charge between the nuclei distinguishes a shared interaction from one between two closed-shells, as found for hydrogen-bonded interactions. In a hydrogen bond the hydrogen atom is linked by bond paths to two other atoms and shares two interatomic surfaces. The primary interaction, the acidic one, exhibits the characteristics of a shared or polar interaction while the weaker one between the hydrogen and the base atoms, exhibits the characteristics of a closed-shell interaction: $0.00<\rho_b<0.06$ au, positive $\nabla^2\rho_b$, and positive or near zero $H(\mathbf{r}_b)$ - table 3.8.

The energy density at the bcp, $H(\mathbf{r}_b)$, becomes increasingly more negative as the polarity of a shared interaction increases, the polarity increasing in the order C-C < C-H < C-N < C-O < C=O. (The net charges on the atoms, $q(\Omega)$, are given in table 3.9). The bonded hydrogens illustrate the interdependence of q(H) and $H(\mathbf{r}_b)$; the values in au are respectively: -0.04 and -0.31 for C-H; +0.36 and -0.54 for N-H, +0.65 and -0.77 for O-H. The decrease in $H(\mathbf{r}_b)$ is accompanied by a decrease in bond length. Thus the increasing energetic stability of the density at a bcp parallels the increase in the bond strength.

The bond properties are remarkably transferable from one rotamer to the other. In general, the fluctuations in these properties are too small across each rotamer set to exhibit physically meaningful trends. For this reason, only the average value of each property and its standard deviation are given in table 3.7. The transferability of the bond length across rotamers is echoed by a parallel

transferability in the bond properties. For example, the high transferability of O-H bond lengths in the **n** set (σ=0.000) is associated with a high transferability in ρ_b and the remaining bond properties (see small σ values in table 3.7). In general, the transferability of one property entails the transferability of the others as well, since the various properties are inter-dependent.

With the exception of the hydrogen bonds, the bond paths lie along the internuclear axes and consequently the bond path lengths equal the corresponding bond lengths and there is no bond strain. Hydrogen bonding in these molecules results in the formation of ring structures in which the bond paths exhibit deviations from linearity.

3.6.3 Hydrogen Bonding

The ubiquity of hydrogen bonds in biology can hardly be overemphasized. Some observations are presented here about the hydrogen bonding in amino acids as illustrated by the leucine rotamers. The identification of conditions for such bonding is important if fragments are to be used to build a total density, since hydrogen bonding perturbs the properties of the atoms involved and their interactions with their neighbors.

All the ionic rotamers show at least one hydrogen bond between H(3) of the quaternized amino group and oxygen O(2). This hydrogen bond closes a 5-membered ring (5-MR) involving Cα, N, H(3), O(2) and C. It is the strongest such interaction found in this study in terms of its ρ_b value. Other weaker hydrogen bonds are also present, the properties of all are summarized in table 3.8. From this table one can identify distinct geometric conditions for each type of hydrogen bond with length given as H to base (*B*) separation and the angle *A*-H-*B* where *A* denotes the acid to yield the following classification: (1) Nine N-H$\cdots$O=C in 5-MR with lengths 1.6 - 1.8 Å, angles 123 - 130° and $\rho_b \sim$ 0.046 - 0.059 au, this type is the strongest hydrogen bond found in this study: it exists in all zwitter-ionic rotamers, and is expected to occur in all zwitter-ionic amino acids; (2) Six C-H$\cdots$O=C in 6-and 7-MR with lengths 2.4 - 2.8 Å, angles 117 - 137° and $\rho_b \sim$ 0.006 - 0.011 au; (3) Two C-H$\cdots$O(H) in 7-MR with lengths 2.7 - 3.0 Å, angles 134 -119° and $\rho_b \sim$ 0.007 and 0.004 au respectively;

Table 3.7 Selected bond properties averaged over the nine neutral (**n**) and nine zwitter-ionic (**z**) leucine rotamers, respectively (distances in Å, all other values in au)

A-B		r_b(A)	r_b(B)	R(A-B)	ρ_b	$\nabla^2\rho_b$	G(**r**)	H(**r**)
Cα-C								
n	av	0.685	0.839	1.524	0.268	-0.790	0.059	-0.256
	σ	0.001	0.001	0.001	0.001	0.005	0.000	0.001
z	av	0.848	0.721	1.569	0.248	-0.670	0.049	-0.217
	σ	0.001	0.002	0.003	0.001	0.008	0.000	0.002
Cα-Cβ								
n	av	0.803	0.738	1.540	0.250	-0.660	0.051	-0.216
	σ	0.002	0.003	0.004	0.002	0.012	0.001	0.004
z	av	0.809	0.720	1.529	0.254	-0.683	0.055	-0.225
	σ	0.002	0.003	0.004	0.002	0.009	0.001	0.003
Cα-H								
n	av	0.686	0.398	1.084	0.294	-1.095	0.040	-0.313
	σ	0.001	0.001	0.002	0.002	0.014	0.001	0.003
z	av	0.689	0.392	1.080	0.296	-1.106	0.038	-0.315
	σ	0.000	0.001	0.001	0.001	0.008	0.001	0.001
Cα-N								
n	av	0.540	0.908	1.448	0.281	-0.900	0.139	-0.364
	σ	0.004	0.003	0.005	0.004	0.024	0.003	0.009
z	av	0.487	1.028	1.516	0.216	-0.208	0.233	-0.285
	σ	0.001	0.002	0.001	0.001	0.015	0.003	0.002
Cβ -Cγ								
n	av	0.784	0.757	1.540	0.250	-0.656	0.050	-0.214
	σ	0.004	0.004	0.003	0.002	0.008	0.001	0.003
z	av	0.793	0.749	1.541	0.249	-0.652	0.051	-0.214
	σ	0.014	0.012	0.004	0.002	0.010	0.001	0.003
Cγ-Cδ1								
n	av	0.764	0.769	1.533	0.252	-0.670	0.051	-0.219
	σ	0.003	0.003	0.001	0.000	0.002	0.000	0.001
z	av	0.766	0.767	1.533	0.252	-0.670	0.051	-0.219
	σ	0.010	0.011	0.001	0.001	0.003	0.000	0.001
Cγ-Cδ2								
n	av	0.763	0.769	1.533	0.252	-0.670	0.051	-0.219
	σ	0.003	0.003	0.001	0.000	0.002	0.000	0.001
z	av	0.767	0.766	1.533	0.252	-0.669	0.051	-0.219
	σ	0.010	0.010	0.001	0.001	0.003	0.000	0.001
C-O								
n	av	0.430	0.900	1.331	0.307	-0.095	0.438	-0.462
	σ	0.000	0.000	0.000	0.000	0.012	0.002	0.001
C=O								
n	av	0.397	0.794	1.190	0.434	0.131	0.792	-0.759
	σ	0.000	0.000	0.000	0.000	0.012	0.002	0.002
C=O(1)								
z	av	0.404	0.807	1.211	0.417	-0.080	0.701	-0.721
	σ	0.000	0.001	0.001	0.001	0.004	0.003	0.002
C=O(2)								
z	av	0.413	0.831	1.244	0.386	-0.194	0.598	-0.647
	σ	0.000	0.001	0.001	0.001	0.003	0.002	0.002

Table 3.7 (Continued)

A-B		r_b(A)	r_b(B)	R(A-B)	ρ_b	$\nabla^2\rho_b$	G(**r**)	H(**r**)
Cβ-H(1)								
n	av	0.680	0.406	1.086	0.286	-1.037	0.044	-0.303
	σ	0.002	0.004	0.002	0.002	0.013	0.001	0.003
z	av	0.675	0.416	1.090	0.282	-1.010	0.046	-0.299
	σ	0.002	0.002	0.001	0.001	0.007	0.001	0.002
Cβ-H(2)								
n	av	0.679	0.407	1.086	0.287	-1.041	0.044	-0.304
	σ	0.001	0.001	0.001	0.001	0.006	0.000	0.002
z	av	0.688	0.397	1.085	0.289	-1.055	0.041	-0.305
	σ	0.006	0.008	0.003	0.003	0.019	0.002	0.003
Cγ-H								
n	av	0.675	0.413	1.088	0.287	-1.044	0.045	-0.306
	σ	0.002	0.003	0.001	0.002	0.009	0.001	0.002
z	av	0.677	0.411	1.088	0.287	-1.047	0.045	-0.306
	σ	0.007	0.011	0.003	0.004	0.027	0.003	0.004
Cδ1-H(2)								
n	av	0.675	0.412	1.087	0.282	-1.012	0.046	-0.298
	σ	0.002	0.002	0.001	0.001	0.006	0.001	0.002
z	av	0.678	0.408	1.087	0.283	-1.018	0.045	-0.299
	σ	0.005	0.008	0.003	0.003	0.021	0.002	0.003
Cδ1-H(3)								
n	av	0.676	0.410	1.086	0.283	-1.018	0.045	-0.300
	σ	0.001	0.001	0.001	0.001	0.004	0.000	0.001
z	av	0.677	0.409	1.086	0.283	-1.020	0.045	-0.300
	σ	0.002	0.003	0.001	0.001	0.006	0.001	0.001
Cδ2-H(1)								
n	av	0.676	0.410	1.086	0.283	-1.017	0.045	-0.299
	σ	0.001	0.001	0.000	0.000	0.002	0.000	0.000
z	av	0.677	0.408	1.086	0.283	-1.021	0.045	-0.300
	σ	0.002	0.002	0.001	0.001	0.004	0.001	0.000
Cδ2-H(2)								
n	av	0.676	0.411	1.086	0.282	-1.012	0.045	-0.299
	σ	0.003	0.003	0.002	0.001	0.008	0.001	0.003
z	av	0.677	0.410	1.087	0.282	-1.014	0.045	-0.299
	σ	0.005	0.007	0.002	0.003	0.019	0.002	0.003
Cδ2-H(3)								
n	av	0.675	0.411	1.086	0.283	-1.019	0.045	-0.300
	σ	0.002	0.002	0.001	0.001	0.005	0.001	0.001
z	av	0.677	0.409	1.086	0.283	-1.023	0.045	-0.300
	σ	0.005	0.007	0.002	0.003	0.017	0.002	0.003
N-H(1)								
n	av	0.735	0.266	1.002	0.354	-1.769	0.059	-0.502
	σ	0.000	0.001	0.000	0.000	0.007	0.000	0.002
z	av	0.756	0.250	1.005	0.354	-1.992	0.045	-0.543
	σ	0.001	0.001	0.001	0.001	0.004	0.000	0.001
N-H(2)								
n	av	0.738	0.264	1.002	0.354	-1.800	0.058	-0.507
	σ	0.001	0.001	0.001	0.001	0.016	0.000	0.003
z	av	0.757	0.249	1.005	0.354	-2.004	0.045	-0.546
	σ	0.001	0.001	0.001	0.001	0.016	0.000	0.004
N-H(3)								
z	av	0.827	0.217	1.044	0.311	-2.010	0.039	-0.541
	σ	0.005	0.000	0.005	0.005	0.039	0.001	0.009

Table 3.7 (Continued)

A-B		r_b(A)	r_b(B)	R(A-B)	ρ_b	$\nabla^2\rho_b$	G(**r**)	H(**r**)
O-H								
n	av	0.782	0.171	0.953	0.375	-2.825	0.062	-0.768
	σ	0.000	0.000	0.000	0.000	0.001	0.000	0.000

Table 3.8 Classification of the hydrogen-bonds based on geometric and bond properties (distances in Å, angles in (°), all other values in au)

Hydrogen bond type		R(H...B)	∠(A-H...B)	ρ_b	$\nabla^2\rho_b$	ϵ	G(**r**)	H(**r**)
N-H···O(=C)								
z A	N-H3···O2	1.65	129	0.057	0.169	0.052	0.051	-0.009
z B	N-H3···O2	1.69	126	0.052	0.166	0.070	0.047	-0.006
z C	N-H3···O2	1.64	130	0.059	0.171	0.048	0.052	-0.010
z D	N-H3···O2	1.67	128	0.054	0.166	0.057	0.048	-0.007
z E	N-H3···O2	1.64	130	0.058	0.170	0.049	0.052	-0.009
z F	N-H3···O2	1.68	128	0.053	0.165	0.058	0.047	-0.006
z G	N-H3···O2	1.73	124	0.048	0.162	0.091	0.044	-0.003
z H	N-H3···O2	1.75	123	0.046	0.161	0.105	0.042	-0.002
z I	N-H3···O2	1.69	127	0.052	0.165	0.064	0.047	-0.005
methyne C-H ··· O(=C)								
z B	Cγ-H···O2	2.54	117	0.010	0.034	0.211	0.008	0.001
methyne C-H···H-N								
z D	Cγ-H···H(2)	2.07	129	0.010	0.042	1.798	0.009	0.002
methyl C-H···H-N								
n F	Cδ2-H(3)···H(2)	2.11	132	0.009	0.033	0.184	0.007	0.001
n I	Cδ1-H(2)···H(2)	2.15	130	0.009	0.033	0.387	0.007	0.001
z F	Cδ2-H(3)···H(2)	2.29	104	0.009	0.036	0.565	0.007	0.002
methyl C-H···N								
n C	Cδ1-H(1)···N	2.55	119	0.011	0.035	0.150	0.008	0.001
n E	Cδ2-H(2)···N	2.61	113	0.010	0.033	0.198	0.007	0.001
z C	Cδ1-H(1)···N	2.72	108	0.007	0.030	0.245	0.006	0.001
z E	Cδ2-H(2)···N	2.78	104	0.007	0.029	0.405	0.006	0.001
methyl C-H···O(H)								
n B	Cδ1-H(1)···O(H)	2.95	119	0.004	0.016	0.250	0.003	0.001
n H	Cδ1-H(2)···O(H)	2.72	134	0.007	0.024	2.201	0.005	0.001
methyl C-H···O(=C)								
n D	Cδ1-H(3)···O(=C)	2.76	137	0.007	0.028	1.214	0.006	0.001
n I	Cδ2-H(2)···O(=C)	2.81	133	0.007	0.026	0.975	0.005	0.001
z G	Cδ2-H(3)···O(2)	2.57	126	0.009	0.030	0.040	0.006	0.001
z H	Cδ1-H(2)···O(2)	2.42	136	0.011	0.037	0.063	0.008	0.001
z H	Cδ2-H(2)···O(2)	2.79	128	0.006	0.020	0.086	0.004	0.001
N-H···C (methyl)								
z I	N-H2···Cδ1	2.56	118	0.009	0.038	1.276	0.008	0.002
C-H···C								
n G	Cδ2-H(3)···C	2.68	108	0.007	0.029	1.257	0.006	0.002
z D	Cδ1-H(1)···C	2.64	111	0.007	0.030	0.934	0.006	0.002
z I	Cδ2-H(2)···C	2.64	111	0.007	0.031	1.185	0.006	0.002

(4) Four C-H···N in 6-MR with lengths 2.6 - 2.8 Å, angles 104 - 119° and ρ_b ~ 0.007 - 0.011 au; (5) Three C-H···C in 6-MR with lengths 2.6 - 2.7 Å, angles 108 - 111° and ρ_b ~ 0.007 au; (6) Four unusual hydrogen-to-hydrogen C-H···H-N bonds (i.e. hydrogen-bonds in which the acceptor atom is a negatively charged hydrogen atom): three in 7-MR with lengths 2.1 - 2.3Å, angles 104 - 132° and ρ_b ~ 0.009 au; and one in a 6-MR with length 2.1Å, angle 129° and ρ_b = 0.010 au. The latter dihydrogen bond has recently been discovered experimentally (*Bodige et al., 1999*) and in calculated densities. (*Alkorta et al., 1998; Garbowski, 2000; Garbowski, 2001; Popelier, 1998*)

Carrol and Bader (*Carroll and Bader, 1988*) reported a wide range of hydrogen bonds between various bases and HF, the strongest being H_3N···HF (ρ_b~ 0.034), an intermediate being HF···HF (ρ_b ~ 0.021) and the weakest being HCl···HF (ρ_b ~ 0.007). Hydrogen bonds occurring between a methyl group and a negatively charged oxygen in creatine and carbamoyl sarcosine have ρ_b values of 0.035 and 0.033 respectively. (*Popelier and Bader, 1992*) In open and cyclic formamide dimers, the values of ρ_b are 0.027 and 0.021 respectively. (*Cheeseman et al., 1988*) Destro *et al.* report x-ray crystallographic ρ_b values for N-H···O bonds between different L-alanine molecules calculated from densities determined from a multipole refinement at 23K. (*Destro et al., 1991*) These workers found two inter-molecular O···H···N hydrogen bonds of length 1.83Å and ρ_b 0.03 au each; these two hydrogen-bonds bifurcate at oxygen O(2) (our numbering). They also found a shorter and stronger inter-molecular O···H···N hydrogen-bond of 1.72Å and ρ_b of 0.04 au. In comparison to these reported theoretical and experimental values of ρ_b, the hydrogen bonds found in this study fall in the range of weak interaction ($0.004 \lesssim \rho_b \lesssim 0.011$) with the exception of N-H···O=C bonds ($0.046 \lesssim \rho_b \lesssim 0.059$).

Popelier and Bader found similar hydrogen bonds between the same atoms (N-H···O=C), but in a non-ionic state, in the blocked triglycyl peptide HC(=O)|Gly|Gly|Gly|NH_2 using the same level of theory as in the current work. (*Popelier and Bader, 1994*) These bonds are found to exhibit values of ρ_b ranging from 0.013 to 0.022 au and all have positive Laplacians ($\nabla^2\rho_b$), as is the case for the Laplacians reported in this work, as expected for such closed shell interactions. In this paper, the

authors also emphasize that the depletion of electron density on the hydrogen bonded hydrogen is an important contributor to the strength of a hydrogen bond. The charge on the non-hydrogen-bonded H atoms on a quaternized ammonium group in **z**-leucine rotamers is around +0.4, but is +0.6 for the hydrogen bonded one. (See table 3.9). Arranging the hydrogen bonded hydrogens in order of their charge reproduces the ordering of the strength of the hydrogen-bonds. The charge on the H for the hydrogen-bonds found by Popelier and Bader is 0.48 (weakest) and 0.50 (strongest), while it is 0.59 on the H involved in the strongest hydrogen-bond in the present work.

In addition to stabilizing and holding certain conformations that might have otherwise been unfavorable, hydrogen-bonding perturbs the properties of the atoms involved and their interactions with their neighbors. For instance, of the two oxygens bonded to the carboxylic carbon in the hydrogen-bonded zwitter-ions, the bond to the one that is hydrogen bonded (O(2)=C) is longer by 0.033Å and has a ρ_b value smaller by 0.031au, i.e. it is longer and weaker than its non-hydrogen-bonded counterpart. In addition, O(2)=C is found to have a more negative Laplacian, is almost perfectly cylindrical (ϵ=0.00) in contrast to 0.06 for O(1)=C, indicating less double bond character, and it also has a larger (less negative) $H(\mathbf{r}_c)$. On the other hand, compared to O(1), O(2) is 0.014e more populated, is destabilized by 22kcal/mol, is smaller in volume by 6.1 au, has a dipolar polarization weaker by 0.11au, and has a quadrupolar polarization which is stronger by 0.25au. Hydrogens involved in hydrogen-bonds have consistently smaller volumes, smaller dipolar polarization and higher quadrupolar polarization.

3.6.4 The Effect of Conformation on Atomic Properties

The atomic properties listed in table 3.9 show a marked transferability for all atoms from rotamer to rotamer within each set, and for atoms of the side chains across the two sets. Only those atoms which are involved in the tautomerism and the atoms to which they are bonded are non-transferable across the **n** and the **z** sets. Conformation has minor effects on the studied atomic properties, as can be seen by the small magnitudes of the σ values. For example, atomic energies show a near-constancy.

Typically, the energies of the same atom in different rotamers differ by a few kcal/mol, a very small number when compared to the magnitude of the atomic energy (e.g. $\sim 5\times10^4$ kcal/mol for an oxygen atom). Relatively, properties with very small magnitudes are the ones that show the larger σ. Otherwise, a slightly elevated σ always originates from the inclusion of the hydrogen-bonded atoms in the averaging, since the properties of these atoms are generally the most outlying ones. In order to show how hydrogen bonding affects the atomic properties, table 3.9 lists the properties of those atoms that are involved in hydrogen bonding explicitly in addition to the averages over all rotamers.

This remarkable transferability of atomic properties, originating from the transferability of the density of the same atom in different rotamers and from the zero-flux condition used to partition the molecule, contrasts with the extreme conformational sensitivity and basis-set dependence of Mulliken-type partitioning schemes.

3.6.5 Effect of Tautomerization on Bond and Atomic Properties

The effect of ionization which leads to the formation of the zwitter-ionic tautomer is significant only for the bond and atomic properties involving the ionized atoms or those directly bonded to them. Generally, bond and atomic properties beyond Cβ change little in response to ionization. As an example, we consider the effect of ionization on rotamer **A**. A comparison of the bond properties and atomic properties of **nA** with the corresponding ones in **zA** show how little the side chain is perturbed by tautomerization. A plot of the values of bond length (Å) and ρ_b (au) for **nA** form minus the corresponding value for zA results in the difference plot shown in figure 3.2(a). The plot immediately reveals that the perturbations in these two properties are localized mainly at the N-Cα and Cα-C bonds. The Cα-Cβ bond is still slightly perturbed by the ionization but no bond beyond Cβ-Cγ shows a significant change in properties. Both the C-Cα and the Cα-N bonds are shorter and stronger (higher ρ_b) in the neutral form.

Table 3.9 Selected atomic properties (au) of nine neutral (**n**) and three (**A**, **E**, and **I**) zwitterionic (**z**) leucine rotamers (listed are the averages over the rotamer set except when the atom is involved in hydrogen bonding)

Atom (Ω)		Rotam.	H-bonded to:	$N(\Omega)$	$E(\Omega)$	$v(\Omega)$	$q(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\lvert\mathbf{Q}(\Omega)\rvert$	$\lvert L(\Omega)\rvert$
C										
	n	G	Hδ2-3	4.231	-36.718	28.91	1.769	1.009	1.485	0.0034
	n	av		4.233	-36.716	30.38	1.767	1.006	1.514	0.0022
		σ		0.003	0.003	0.79	0.003	0.003	0.014	0.0012
	z	av		3.900	-36.485	26.59	2.100	0.580	1.547	0.0013
		σ		0.002	0.002	0.90	0.002	0.007	0.047	0.0010
Cα										
	n	av		5.436	-37.505	40.99	0.564	0.506	1.058	0.0020
		σ		0.008	0.007	0.40	0.008	0.006	0.101	0.0008
	z	av		5.717	-37.663	47.65	0.283	0.586	1.686	0.0010
		σ		0.002	0.003	0.13	0.002	0.002	0.023	0.0003
Cβ										
	n	av		5.869	-37.746	52.34	0.131	0.099	0.408	0.0018
		σ		0.004	0.007	0.36	0.004	0.004	0.040	0.0011
	z	av		5.870	-37.747	53.22	0.130	0.128	0.565	0.0028
		σ		0.002	0.009	0.31	0.002	0.007	0.013	0.0018
Cδ1										
	n	av		5.862	-37.723	62.31	0.138	0.038	0.081	0.0009
		σ		0.005	0.003	0.41	0.005	0.006	0.010	0.0006
	z	av		5.865	-37.725	62.09	0.135	0.037	0.146	0.0009
		σ		0.010	0.004	0.64	0.010	0.026	0.081	0.0006
Cδ2										
	n	av		5.861	-37.722	62.25	0.139	0.037	0.094	0.0009
		σ		0.003	0.002	0.37	0.003	0.004	0.023	0.0006
	z	av		5.863	-37.724	62.01	0.137	0.049	0.183	0.0011
		σ		0.006	0.001	0.27	0.006	0.021	0.025	0.0005
Cγ										
	n	av		5.849	-37.705	46.94	0.151	0.043	0.189	0.0023
		σ		0.006	0.004	0.53	0.006	0.017	0.052	0.0016
	z	av		5.857	-37.708	47.43	0.143	0.071	0.328	0.0038
		σ		0.003	0.002	0.56	0.003	0.018	0.049	0.0015
H (-O)										
	n	av		0.352	-0.327	18.61	0.648	0.142	0.066	0.0000
		σ		0.001	0.000	0.09	0.001	0.000	0.000	0.0000
H(1)										
	n	av		0.641	-0.486	31.54	0.359	0.186	0.111	0.0001
		σ		0.003	0.001	0.22	0.003	0.001	0.006	0.0000
	z	av		0.561	-0.444	27.53	0.439	0.158	0.091	0.0001
		σ		0.001	0.001	0.20	0.001	0.001	0.002	0.0000
H(2)										
	n	F	Hδ2-3	0.627	-0.482	28.37	0.373	0.175	0.136	0.0001
	n	I	Hδ1-2	0.627	-0.482	28.74	0.373	0.176	0.131	0.0000
	n	av		0.629	-0.480	30.38	0.371	0.181	0.103	0.0001
		σ		0.003	0.002	1.02	0.003	0.003	0.017	0.0000
	z	I	Cδ1	0.555	-0.439	25.83	0.445	0.151	0.165	0.0034
	z	av		0.558	-0.441	26.84	0.442	0.156	0.113	0.0012
		σ		0.002	0.002	0.73	0.002	0.004	0.037	0.0015

Table 3.9 (Continued)

Atom (Ω)		Rotam.	H-bonded to:	$N(\Omega)$	$E(\Omega)$	$\nu(\Omega)$	$q(\Omega)$	$\lvert\boldsymbol{\mu}(\Omega)\rvert$	$\lvert\mathbf{Q}(\Omega)\rvert$	$\lvert L(\Omega)\rvert$
H(3)										
	z	av*	O=C	0.406	-0.346	14.85	0.594	0.114	0.168	0.0003
		σ		0.004	0.003	0.37	0.004	0.002	0.004	0.0001
Hα										
	n	av		1.012	-0.646	45.77	-0.012	0.121	0.381	0.0001
		σ		0.001	0.002	0.87	0.001	0.003	0.007	0.0001
	z	av		0.995	-0.637	45.537	0.005	0.12	0.409	0.0001
		σ		0.000	0.001	0.594	0.000	0.00	0.010	0.0000
Hβ1										
	n	av		1.039	-0.646	48.92	-0.039	0.131	0.408	0.0001
		σ		0.013	0.004	1.13	0.013	0.001	0.008	0.0000
	z	av		1.064	-0.655	50.32	-0.064	0.131	0.421	0.0001
		σ		0.009	0.004	0.45	0.009	0.002	0.001	0.0000
Hβ2										
	n	av		1.040	-0.648	47.79	-0.040	0.125	0.420	0.0001
		σ		0.004	0.003	0.75	0.004	0.003	0.010	0.0000
	z	av		0.973	-0.619	44.63	0.027	0.123	0.399	0.0001
		σ		0.012	0.005	1.15	0.012	0.002	0.004	0.0000
Hδ1-1										
	n	B	O (H)	1.045	-0.645	48.98	-0.045	0.121	0.489	0.0000
	n	C	N	1.004	-0.630	45.28	-0.004	0.120	0.508	0.0000
	n	av		1.044	-0.642	49.75	-0.044	0.126	0.479	0.0001
		σ		0.016	0.005	1.95	0.016	0.006	0.021	0.0000
	z	av		1.056	-0.645	51.02	-0.056	0.129	0.467	0.0001
		σ		0.025	0.010	1.00	0.025	0.002	0.003	0.0000
Hδ1-2										
	n	H	O (H)	1.057	-0.651	48.36	-0.057	0.116	0.515	0.0000
	n	I	H(2)	1.064	-0.653	49.70	-0.064	0.123	0.462	0.0001
	n	av		1.057	-0.646	50.92	-0.057	0.130	0.507	0.0001
		σ		0.006	0.004	1.28	0.006	0.006	0.123	0.0000
	z	av		1.056	-0.646	50.28	-0.056	0.133	0.427	0.0016
		σ		0.013	0.006	0.69	0.013	0.005	0.020	0.0014
Hδ1-3										
	n	D	O=C	1.049	-0.643	50.75	-0.049	0.131	0.466	0.0001
	n	av		1.052	-0.644	51.07	-0.052	0.131	0.468	0.0001
		σ		0.004	0.001	0.25	0.004	0.000	0.001	0.0000
	z	av		1.039	-0.639	50.40	-0.039	0.130	0.463	0.0001
		σ		0.009	0.004	0.46	0.009	0.000	0.005	0.0000
Hδ2-1										
	n	av		1.052	-0.644	51.02	-0.052	0.131	0.468	0.0001
		σ		0.004	0.001	0.28	0.004	0.000	0.002	0.0000
	z	av		1.044	-0.641	50.63	-0.044	0.130	0.464	0.0001
		σ		0.013	0.005	0.55	0.013	0.001	0.004	0.0000
Hδ2-2										
	n	E	N	1.012	-0.632	46.08	-0.012	0.125	0.480	0.0001
	n	I	O=C	1.048	-0.647	48.20	-0.048	0.116	0.502	0.0000
	n	av		1.047	-0.642	50.45	-0.047	0.129	0.464	0.0001
		σ		0.014	0.005	1.91	0.014	0.006	0.016	0.0000

Table 3.9 (Continued)

Atom (Ω)		Rotam.	H-bonded to:	$N(\Omega)$	$E(\Omega)$	$\nu(\Omega)$	$q(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\lvert Q(\Omega)\rvert$	$\lvert L(\Omega)\rvert$
Hδ2-2										
	z	**A****		1.034	-0.635	50.44	-0.034	0.130	0.452	0.0001
	z	E	N	1.087	-0.660	51.31	-0.087	0.127	0.474	0.0000
	z	I	C	1.009	-0.631	45.65	-0.009	0.115	0.482	0.0000
	z	av.		1.043	-0.642	49.13	-0.043	0.124	0.470	0.0000
		σ		0.033	0.013	2.49	0.033	0.007	0.013	0.0000
Hδ2-3										
	n	F	H(2)	1.055	-0.651	48.30	-0.055	0.116	0.492	0.0001
	n	G	C	1.066	-0.652	49.47	-0.066	0.112	0.533	0.0031
	n	av.		1.056	-0.647	50.508	-0.056	0.126	0.479	0.0004
		σ		0.006	0.003	1.064	0.006	0.007	0.021	0.0009
	z	av.		1.052	-0.645	50.51	-0.052	0.128	0.468	0.0001
	z	σ		0.022	0.009	1.04	0.022	0.002	0.016	0.0000
Hγ										
	n	av.		1.068	-0.666	48.97	-0.068	0.128	0.404	0.0004
		σ		0.011	0.004	0.98	0.011	0.003	0.008	0.0007
	z	av.		1.078	-0.670	49.71	-0.078	0.129	0.405	0.0001
		σ		0.026	0.009	1.13	0.026	0.001	0.012	0.0000
N										
		C	Hδ1-1	8.147	-54.926	107.76	-1.147	0.194	2.066	0.0002
		E	Hδ2-2	8.149	-54.927	107.96	-1.149	0.192	2.072	0.0001
	n	av.		8.146	-54.914	111.43	-1.146	0.207	2.311	0.0006
		σ		0.006	0.011	2.18	0.006	0.015	0.148	0.0010
	z	E	Hδ2-2	8.221	-55.131	90.61	-1.221	0.119	1.102	0.0008
	z	av.		8.220	-55.130	91.29	-1.220	0.113	1.091	0.0005
		σ		0.001	0.002	0.63	0.001	0.008	0.008	0.0003
O (H)										
	n	B	Hδ1-1	9.302	-75.634	119.18	-1.302	0.310	0.939	0.0007
	n	H	Hδ1-2	9.299	-75.624	120.18	-1.299	0.304	0.930	0.0005
	n	av.		9.301	-75.625	119.78	-1.301	0.306	0.962	0.0005
		σ		0.001	0.004	0.49	0.001	0.003	0.021	0.0007
O (=C)										
	n	D	Hδ1-3	9.358	-75.679	134.12	-1.358	0.583	0.518	0.0000
	n	I	Hδ2-2	9.357	-75.678	134.76	-1.357	0.583	0.518	0.0001
	n	av.		9.356	-75.677	135.37	-1.356	0.582	0.554	0.0001
		σ		0.002	0.004	0.62	0.002	0.003	0.044	0.0000
O(1)										
	z	av.		9.387	-75.620	139.63	-1.387	0.486	0.721	0.0001
		σ		0.001	0.001	0.21	0.001	0.002	0.040	0.0001
O(2)										
	z	av.		9.401	-75.585	133.55	-1.401	0.377	0.972	0.0001
		σ		0.002	0.001	0.25	0.002	0.001	0.054	0.0000

* All the (z) rotamers exhibit this particular hydrogen-bond.

** This is the only one of the three averaged atoms that is *not* hydrogen bonded.

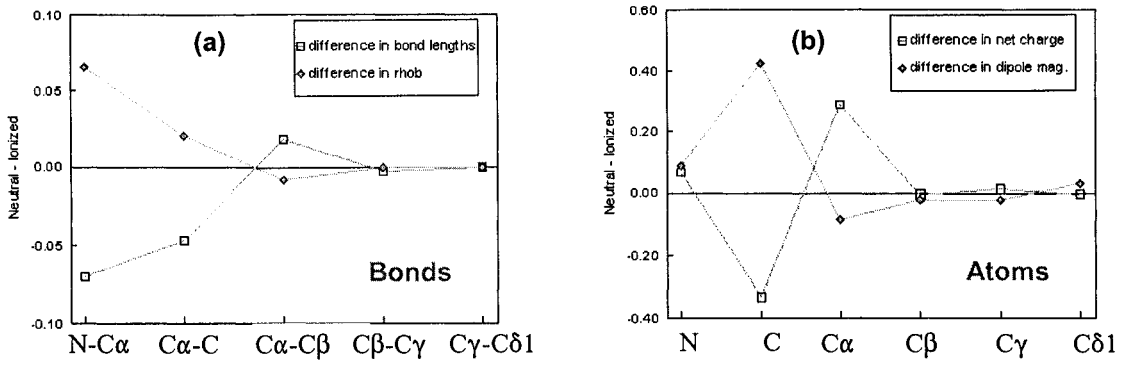


Figure 3.2 The effect of tautomerization on: **(a)** bond properties, and **(b)** atomic properties. These are plots of the difference between the bond or atomic property of the neutral A rotamer (**nA**) and the corresponding value in its zwitter-ionic counterpart (**zA**). Both plots indicate the short-range nature of the perturbations induced by tautomerization on bond and atomic properties, in other words, groups directly involved in the tautometrization are the only ones that are significantly perturbed.

The short-range nature of the perturbation is also reflected in the atomic properties. For example, figure 3.2(b) shows the rapid damping of the perturbing effect of ionization on the atomic charge and the magnitude of the dipole moment beyond Cβ. The atoms primarily responsible for the increase in energy incurred in the formation of the zwitter-ion of rotamer **A** (E_τ = +26 kcal/mol) are readily accounted for by comparing the atomic energies in the two tautomers. The formation of the zwitter-ion results in the destabilization of atoms C, O(2), H(1) and H(2) by 146, 56, 26, and 24 kcal/mol respectively; and in the stabilization of atoms Cα and N by 101 and 130 kcal/mol respectively. The combined effect of ionization on these atoms alone is a destabilization energy of about 20 kcal/mol accounting for 77% of the destabilizing E_τ. In other words, the destabilization upon ionization is mostly localized at the ionized atoms and their immediate neighbourhood. The only atoms whose populations are affected significantly on going from the **n**-tautomer to the **z**-tautomer (i.e. absolute difference is greater than 0.08e) are: C which loses 0.334e, Cα which gains 0.286e, and O(H) which gains 0.086e upon the loss of its hydrogen to the nitrogen. The nitrogen experiences a slight increase (0.069e) in population, and hence is slightly more negative upon protonation. It acheives this increase in population by extracting electrons from the neighbouring Cα forcing the less electronegative carbon to bear the increase in positive charge on the quaternized amino group. The

functional group properties are the sums of the atomic properties of the individual atoms making up the functional group. The functional group charges are $-NH_2$ (-0.420e), -COOH (-0.238e), $-NH_3^+$ (+0.257e), $-COO^-$ (-0.683e). In other words, $-NH_3^+$ is more positive relative to $-NH_2$ by 0.658e, and $-COO^-$ is more negative by 0.445e relative to -COOH. The CαH fragment loses 0.270e upon **nA** → **zA** tautomerization. The corresponding values for glycine are very similar: the amino group gains 0.672 positive charge, the carboxylic group gains 0.442 negative charge and the CαH fragment loses 0.266e. A comparison of the changes in the functional group populations upon tautomerization in leucine and in glycine shows how little the side chain perturbs the functional groups' response to this dramatic change in structure.

3.7 Conclusions

This chapter has demonstrated how theory can be used to define the properties of atoms in molecules and to determine the degree of sensitivity of their properties to changes in conformation and to changes in the atoms' environments. It should be emphasised that the model-independent nature of this theory, a result of its basis in the measurable charge distribution, provides a physical measure of an atom and its properties. The theory is a realisation of the obvious truism that the properties of matter are determined by its form in real space. Therefore, the transferability of an atom is determined by the degree of transferability of its charge distribution, a distribution that is found within a bounded region of real space. Clearly, any degree of transferability that is exhibited by a group in any particular situation will be necessarily and hence faithfully recovered by the theory of atoms in molecules.

The present results show that the numerous possible conformations that are energetically accessible to the side chain do not present a serious problem in the use of amino acid residues in the modelling of proteins. Based on our results and those of earlier workers, the conformers accessible from an energy minimum that is characterised by normal bond lengths and bond angles, differ little in energy and the geometric parameters are not significantly altered as various dihedral angles are

sampled. This conservation of geometry in turn ensures the conservation of the form of the density distribution of corresponding atoms in different rotamers. As a consequence, the bond, atomic and group properties defined by theory are found to be equally insensitive to conformational changes. The results of this chapter set the stage for the two following ones, wherein the geometric, bond, and atomic properties of each of the free genetically-encoded amino acids are determined for a single minimum energy conformer.

The presence of strong hydrogen bonds, those with ρ_b values in excess of 0.02 au, do change the properties of the atoms directly involved in the interaction. The results presented here provide geometric parameters characterising such interactions for each set of ρ_b values, that enable one to predict the presence of such interactions and the resulting changes in the atomic properties.

APPENDIX 3.1: Numbering Conventions

The IUPAC-IUB conventions (*IUPAC-IUB Commission on Biochemical Nomenclature, 1970a; IUPAC-IUB Commission on Biochemical Nomenclature, 1970b*) were adopted unless otherwise stated. The hydrogens of the leucine rotamers were numbered as follows: (i) nitrogen's H(1) has the smallest H-N-Cα-Hα angle; (ii) Cβ's H(1) has the smallest Hβ-Cβ-Cα-Hα angle; and (iii) Cδ's H(1) has the smallest Hδ-Cδ-Cγ-Hγ angle; these angles are those necessary to rotate the H clockwise when viewed from its side. For the nitrogen of the zwitterion, the hydrogen involved in the H-bonding with the carboxyl oxygen (O(2)) is H(3), the two others are labelled as described above.

Chapter 4

Molecular Geometries of the Genetically-Encoded Amino Acids

4.1 Statement of the Problem

This chapter represents a necessary step in the characterization of the properties of the genetically-encoded amino acids in terms of their electron density by providing a full examination of their molecular geometries. This study lays the foundations for a determination of the atomic and bond properties using the quantum theory of atoms in molecules (QT-AIM). What is remarkable, as is demonstrated by this theory, is the exceptional degree of transferability that the charge distribution and hence the properties of a functional grouping of atoms can exhibit. A prerequisite for this transferability of the density is the transferability of the geometric parameters, since one can think of the molecular geometry as the skeleton and the density as the flesh making the molecular body. Since the geometry is determined by the Hellmann-Feynman forces exerted on the nuclei and these forces are completely determined by the electron density distribution, the condition of a transferable geometry is both a necessary and sufficient condition for the electron density to be equally transferable. (*Bader, 2002*) One may also appeal to density functional theory which establishes that the electron density determines all measurable properties of a system. In particular, DFT establishes a one-to-one

correspondence between the density and the external potential, the potential that embodies the molecular geometry. (*Bader, 2002*)

Thus, it is the goal of this chapter to examine the extent of transferability of the geometric parameters of the main chain group and of the functional groups present in the side chains amongst the members of the amino acid series. The following chapter uses the electron densities determined for the optimized geometries described here to define the atoms and their properties and correlate these properties with experimental physicochemical properties. The present chapter is based on the manuscript: (*Matta and Bader, 2002a*)

4.2 Introduction

This part of the thesis focuses on the properties of the amino acid side-chains, the portions of amino acids which stamp them with their individual identity. These properties are determined and studied in their smallest natural "mold" consisting of the *free amino acids* in their non-zwitter-ionic form, H_2N-CαHα(R)-COOH, where R denotes the side-chain. Henceforth, this form is denoted by AA. The AA form yields an environment of the side-chain that is free of any field effects of the charged groups present in the zwitter-ion and it thus serves as a suitable approximation to the electronic environment of the side-chains of the corresponding *amino acid residues* (labelled Aa) in a polypeptide. This approximation is assessed by a detailed comparison of the geometries of the R group in a free amino acid AA and as part of the residue Aa in the tripeptide Gly-Aa-Gly, where Aa stands for any one of the twenty amino acid residues. The comparison shows that the side chain geometries of the side chains are not significantly changed upon incorporation into a peptide. This comparison demonstrates that Cα effectively damps the perturbations of the properties of R including the geometric parameters caused by the formation of the amide linkages in the main-chain atoms and the side-chain properties are only slightly perturbed by the alterations of the main-chain group (-CαHα(NH_2)COOH) induced by peptide bond formation. The results presented in chapter 3 - showing that when geometry is transferable the

atomic and bond properties are also transferable - lead one to anticipate the transferability of the atomic and bond properties on going from the free amino acid to the amino acid residue in a peptide if the geometries are transferable.

This chapter considers the following topics: (i) The transferability of the geometric parameters that characterize the bonding between the main-chain atoms, the atoms common to all amino acids; (ii) The determination of the geometries of the individual side-chains and a comparison of the geometric parameters for functional groups common to two or more amino acids; (iii) A comparison of the theoretically optimized geometries with experimental geometries; (iv) A selective review of the literature and a compilation of literature sources for the experimental geometries of the genetically-encoded amino acids; and (v) The introduction of labeling conventions to complement the IUPAC-IUB conventions by uniquely identifying every atom and every bond in the genetically-encoded amino acid series with respect to their stereochemistry.

4.3 Previous *Ab Initio* Computational Studies of the Genetically-Encoded Amino Acids

Several excellent computational studies of the amino acids or of models of the side-chains have been reported (see for example Refs: (*Cao et al., 1995; Chipot et al., 1992; Chipot et al., 1993a; Chipot et al., 1993b; Gronert and O'Hair, 1995; Price et al., 1991; Schäfer et al., 1990a; Schäfer et al., 1990b; Shirazian and Gronert, 1997; Siam et al., 1984; Sokalski et al., 1989; Van Alsenoy et al., 1988*)) and reviewed ((*Sapse, 2000*) a review which we will not duplicate). (See also section 3.3 of the previous chapter). Many of the earlier studies consider only the side-chains terminated with a $C\alpha H_3$ group and most address only certain subclasses of the genetically-encoded amino acids. The literature is missing a comprehensive theoretical study of the entire family of the genetically-encoded amino acids at one and the same level of theory, a void this work seeks to fill.

4.4. Computational Strategy and Methods

All geometry optimizations were performed using the Gaussian94 program. (*Frisch et al., 1995*) Only those ionization states of the side-chains which predominate at biological pH are considered. In humans, the pH of extracellular fluid lies between 7.35 and 7.45 while estimates of intracellular pH suggest values around 6.9. (*Levinsky, Wilson et al., 1991*) The percent ionization of acidic and basic side-chains in the pH range (6.90-7.45) for amino acids with ionizable side-chains was calculated. The calculation shows that all the side-chains are either almost completely ionized or completely unionized at these pH values except histidine which has a degree of ionization of *ca.* 7% in this range, and hence it is studied in both its unionized (His) and ionized [$His^{(+)}$] states. In addition, neutral histidine occurs in two tautomeric forms depending on which nitrogen atom in the imidazole ring is bonded to a hydrogen atom and which one is unsaturated. See figure 4.1. The tautomer with the hydrogen atom bonded to Nδ1 (His(I)) is more stable by 1.8 kcal/mol than the tautomer with the hydrogen atom bonded to Nϵ2 (His(II)). Both tautomers of neutral histidine are studied, but only the values for His(I) are included in the statistics presented in the tables.

Initial geometries were subjected to full optimization first at the restricted Hartree-Fock (RHF)/3-21G level, followed by another full optimization at the RHF/6-31+G* level to obtain final optimized geometries. The symbol (+) in the basis set description indicates that diffuse functions were incorporated to ensure adequate representation of the diffuse density of anionic amino acids such as $Asp^{(-)}$ and $Glu^{(-)}$ and for the detection of intramolecular hydrogen-bonding. Several previous studies indicated the adequacy of this level of theory in the study of amino acids and peptides. (*Bader and Martín, 1998; Head-Gordon et al., 1991; Martín, 2001; Popelier and Bader, 1994*) Head-Gordon *et al.* have shown that RHF/6-31+G*-optimized geometries provide an excellent level to map the conformational potential energy surface of small dipeptides. (*Head-Gordon et al., 1991*) The optimized geometries presented here are used in chapter 5 to obtain electron density distributions using the larger 6-311++G** basis set. Popelier and Bader have shown that wavefunctions obtained using

this large basis set at the RHF/6-31+G* geometries recover the hydrogen bonding anticipated in the 3.6_{13} α-helical geometry of N-formyltriglycine. (*Popelier and Bader, 1994*) These authors have also shown that the charge neutrality of an amino acid residue defined by QTAIM, a necessary prerequisite for its transferability, is preserved at this level upon twisting of the peptide. (*Martín, 2001; Popelier and Bader, 1994*) Thus, the RHF/6-31+G* level of theory is not only adequate but will also facilitate comparison with a substantial body of results. (*Bader and Martín, 1998*)

In view of its increasing usage in the literature, we have performed density functional theory (DFT) optimization for glycine and tyrosine, to compare with the RHF results. The DFT calculations were performed in the manner described above using the B3LYP hybrid functional and a 6-31+G* basis set.

4.5 Results and Discussion

4.5.1 Calculated Geometries

The optimized geometries determined in this study are displayed in figure 4.1, placed at the end of this chapter after an appendix describing the atom and bond labeling conventions. The same geometries and labeling conventions also serve chapter 5. The IUPAC-IUB conventions (*IUPAC-IUB Commission on Biochemical Nomenclature, 1970a; IUPAC-IUB Commission on Biochemical Nomenclature, 1970b*) are adopted in this work and supplemented with additional rules to label all atoms and bonds uniquely according to their stereochemistry (figure 4.1 and appendix 4.1).

4.5.1.1 The Geometry of the Main-Chain Group [CαHα(NH$_2$)COOH]

The bond lengths found in the main-chain groups, their average values and standard deviations, are listed in table 4.1. The conformations of these groups are defined by the dihedral angle ψ (∠N-Cα-C=O1) (table 4.2), an angle which equals zero in the case of glycine. While most amino acids generally tend to retain the conformation of glycine, they exhibit values for the dihedral angle ψ

different from zero. A negative value indicates that the C=O1 bond is on the same side of the N-Cα-C plane as the side-chain bond **1**. A positive value indicates that the C-O2 bond is on the same side of the N-Cα-C plane as the Cα-Hα bond.

Table 4.1 Bond lengths (Å) of the main-chain group

	C=O	C-OH	O-H	Cα-C	Cα-N	Cα-Hα	N-H[b]
Gly	1.189	1.330	0.953	1.516	1.438	1.084	1.001
Gly(DFT)[a]	*1.213*	*1.357*	*0.976*	*1.527*	*1.450*	*1.097*	*1.018*
Ala	1.190	1.331	0.953	1.523	1.443	1.085	1.002
Val	1.190	1.331	0.953	1.524	1.445	1.085	1.001
Leu	1.191	1.330	0.953	1.524	1.449	1.081	1.002
Ile	1.191	1.330	0.953	1.525	1.449	1.083	1.002
Met	1.190	1.332	0.953	1.525	1.456	1.083	1.002
Cys	1.190	1.329	0.953	1.523	1.455	1.082	1.002
Cystine	1.190	1.324	0.953	1.526	1.449	1.082	1.002
Ser	1.188	1.336	0.953	1.525	1.443	1.085	1.002
Thr	1.188	1.331	0.952	1.527	1.441	1.087	1.002
$Lys^{(+)}$	1.192	1.324	0.953	1.527	1.451	1.085	1.002
$Arg^{(+)}$	1.188	1.328	0.954	1.526	1.445	1.083	1.002
His(I)	1.194	1.327	0.953	1.524	1.454	1.083	1.002
His(II)	1.190	1.326	0.953	1.525	1.450	1.082	1.002
$His^{(+)}$	1.199	1.310	0.955	1.529	1.443	1.087	1.002
Asn(I)	1.189	1.326	0.953	1.524	1.449	1.084	1.003
Asn(II)	1.194	1.319	0.953	1.522	1.454	1.081	1.001
$Asp^{(-)}$	1.194	1.338	0.951	1.510	1.452	1.081	1.003
$Glu^{(-)}$	1.194	1.332	0.952	1.521	1.453	1.084	1.002
Gln	1.191	1.330	0.952	1.523	1.448	1.083	1.001
Phe	1.190	1.331	0.953	1.524	1.443	1.083	1.002
Tyr	1.190	1.330	0.953	1.523	1.444	1.083	1.002
Tyr(DFT)[a]	*1.214*	*1.358*	*0.977*	*1.531*	*1.457*	*1.095*	*1.019*
Trp	1.190	1.330	0.953	1.526	1.452	1.082	1.002
Pro	1.190	1.330	0.952	1.510	1.446	1.086	0.998
av(AA)	1.191	1.329	0.953	1.523	1.448	1.084	1.002
σ(AA)	0.003	0.005	0.001	0.004	0.005	0.002	0.001
max	1.199	1.338	0.955	1.529	1.456	1.087	1.003
min	1.188	1.310	0.951	1.510	1.438	1.081	0.998
|max-min|	0.012	0.029	0.004	0.019	0.018	0.006	0.005
av(Aa)[c]	1.199			1.531	1.449	1.080	0.997
σ(Aa)[c]	0.002			0.002	0.003	0.000	0.000

(a) Entries in italics are not included in the statistics.
(b) Values for N-H are averages over N-H1 and N-H2.
(c) Averages and standard deviations of the corresponding bond length in amino acid residues (Aa) calculated in tripeptide molds of the type Gly|Aa|Gly, taken from (*Martín, 2001*).

The bond lengths linking the main-chain atoms exhibit a high degree of transferability, as seen from the values of the standard deviations given in table 4.1 which never exceed 0.005 Å and the small spread in their maximum-minimum values. Moreover, there is no obvious trend in the bond lengths reflecting the overall charge on AA, whether negative, positive, or neutral. The bond lengths of the main-chain group in histidine, the only amino acid studied in both of its side-chain ionization states, exhibit little change between the two states. Even the main-chain bond lengths for proline, in which the nitrogen atom of the main-chain amino group is bonded to the side-chain to form a ring structure (for structure, see figure 4.1) do not show large deviations from the average values.

Selected bond angles and dihedral angles of the main-chain atoms are given in table 4.2. The angles formed by the bonds between the main-chain atoms exhibit only small standard deviations, of 1.6° or less, with exceptions for the angles involving Cα in Asn and $Asp^{(-)}$ resulting from the presence of hydrogen bonds as discussed below. All AA, with three exceptions, exhibit negative values of ψ, with eight having $|\psi| < |30^{\circ}|$, 13 with $|\psi| < |60^{\circ}|$, and the remainder exhibiting larger magnitudes. Of those with $\psi > 0$, the values for serine and threonine are close to zero, while glutamine exhibits an unusual ψ angle of 136.6°. The dihedral angle O2-C-Cα-N, Tor1, is always close to complementarity with ψ, showing that carbon atom C and the atoms bonded to it lie nearly in a plane. The last dihedral angle shown in table 4.2 is H-O2-C=O1, Tor2, which indicates that the COOH group is close to planar in all the amino acids with maximum deviations from planarity exhibited by $His^{(+)}$ and $Glu^{(-)}$ where $Tor2 = -6^{\circ}$.

4.5.1.2 The Apparent Anomaly of the C-N Bond Length in Amino Acids

It has long been known that the Cα-N bond found in peptides as well as other C-N single bonds are shorter than the $C\text{-}NH_3^{(+)}$ bond. This result appears to some to be counterintuitive since larger charge separation is expected to favour shorter bonds. (*Hahn, 1957*) Participation in intermolecular hydrogen bonding within the crystal was suggested as a possible reason for the anomaly. (*Hahn, 1957*) Since the

effect is recovered by theoretical calculations on gas-phase molecules, the crystal effects are ruled out as the cause. To explore the origin of this apparent anomaly, RHF wavefunctions using the large basis

Table 4.2 Selected bond angles and dihedral angles of the main-chain group (values in (°))

	O-C=O	C-O-H	O=C-Cα	O-C-Cα	C-Cα-N	N-Cα-R	ψ	Tor1	Tor2
Gly	122.7	109.1	125.6	111.7	115.4	109.9	-0.1	180.0	0.1
Gly(DFT)	*122.8*	*107.5*	*125.7*	*111.5*	*115.8*	*109.7*	*0.1*	*-179.8*	*0.0*
Ala	122.3	108.9	125.4	112.3	113.3	110.2	-14.5	166.0	-0.6
Val	122.1	108.8	125.4	112.5	112.6	112.1	-12.7	167.2	-0.5
Leu	122.2	108.9	125.3	112.5	112.9	111.0	-17.1	163.4	-0.9
Ile	122.0	108.7	124.8	113.2	111.7	110.5	-29.0	150.6	-1.2
Met	122.2	108.9	126.0	111.8	110.1	109.7	-96.2	80.6	0.9
Cys	122.4	109.1	125.6	111.9	110.6	109.2	-98.8	77.9	-0.0
Cystine	123.1	109.2	123.7	113.2	112.0	107.9	-46.5	132.8	-3.3
Ser	121.7	108.9	125.4	112.9	112.5	109.6	4.0	-176.0	1.2
Thr	122.3	108.7	125.1	112.6	113.0	111.1	2.8	-180.0	-1.4
$Lys^{(+)}$	122.8	109.8	125.6	111.6	112.0	109.5	-116.2	62.8	1.4
$Arg^{(+)}$	122.9	109.3	123.9	113.2	112.4	108.5	-46.3	134.9	3.0
His(I)	122.0	109.3	126.1	111.8	110.6	109.4	-108.2	69.3	0.2
His(II)	122.7	108.6	124.0	113.3	111.9	108.4	-41.3	137.3	-6.5
$His^{(+)}$	122.7	110.5	124.5	112.8	114.5	108.4	-142.7	38.8	-0.4
Asn(I)	122.9	109.3	124.7	112.4	112.4	111.2	-16.9	163.5	-1.2
Asn(II)	122.8	109.2	124.1	113.0	106.0	114.0	-93.7	85.1	2.5
$Asp^{(-)}$	120.6	107.6	126.3	113.1	108.9	115.3	-34.7	147.8	-1.9
$Glu^{(-)}$	121.3	108.3	125.3	113.4	111.7	109.7	-37.3	141.9	-5.9
Gln	122.1	109.0	124.8	113.1	114.2	109.3	136.6	-44.6	0.4
Phe	122.4	109.0	125.2	112.5	112.7	110.8	-19.7	161.3	-0.9
Tyr	122.3	108.9	125.2	112.5	112.7	110.8	-19.6	161.4	-0.8
Tyr(DFT)	*122.4*	*107.4*	*125.3*	*112.3*	*112.8*	*110.7*	*-23.0*	*158.3*	*-1.0*
Trp	122.2	108.6	124.1	113.6	111.1	108.9	-56.5	121.4	-2.0
Pro	122.2	108.9	125.6	112.2	112.8	104.0	-3.1	175.7	0.1
av(AA)	122.3	109.0	125.1	112.6	112.4	109.6	-37.7	94.9	-0.5
σ(AA)	0.5	0.5	0.7	0.6	1.3	1.6	56.0	99.6	1.8
max	123.1	110.5	126.3	113.6	115.4	112.1	136.6	180.0	3.0
min	120.6	107.6	123.7	111.6	110.1	104.0	-142.7	-180.0	-5.9
|max-min|	2.5	2.9	2.6	2.0	5.3	8.1	279.4	359.9	8.9
av(Aa)*			121.3		113.4				
σ(Aa)*			0.6		0.9				

ψ = dihedral angle ∡ O1=C-Cα-N, Tor1 = dihedral angle ∡ O2-C-Cα-N, and Tor2 = dihedral angle ∡ H-O2-C=O1. Entries in italics are not included in the statistics. The C-Cα-N and N-Cα-R angles for Asn(II) and $Asp^{(-)}$ are not included in the average since the amino group in these two molecules adopts a different conformation than for the rest of the amino acids, where the acute H1-N-H2 angle faces the side chain rather than the main chain i.e. is bisected roughly by the Cα-Cβ bond instead of the Cα-C bond.

* Averages and standard deviations of the corresponding bond length in amino acid residues (Aa) calculated in tripeptide molds of the type Gly|Aa|Gly, taken from (*Martín, 2001*).

Table 4.3 Charge separation and Cα-N bond lengths in four ionization states of gas-phase glycine

Molecule	BL	q(Cα)	q(N)	$\|q(C\alpha)-q(N)\|$	q(amino)
Gly(**n**)	1.438	0.617	-1.153	1.770	-0.414
$Gly^{(-)}$	1.460	0.520	-1.179	1.699	-0.502
$Gly^{(+)}$	1.501	0.471	-1.202	1.672	0.347
Gly(**z**)	1.505	0.331	-1.228	1.559	0.258

BL is the bond length in Å, q stands for the atomic or functional group charge in au, the charge separation is given by the absolute value of the difference in the charge of the nitrogen atom and the α-carbon atom, and q(amino) refers to the total charge on the amino group in its two ionization states (amino refers to NH_2 in both Gly(**n**) and $Gly^{(-)}$, and refers to NH_3 in both $Gly^{(+)}$ and Gly(**z**)).

set 6-311++G**//6-31+G* were obtained for glycine in the following closed-shell ionization states: neutral non- zwitter-ionic Gly(**n**), zwitter-ionic Gly(**z**), glycine anion $Gly^{(-)}$ and protonated glycine cation $Gly^{(+)}$ and the atomic populations determined using QTAIM. The results are given in table 4.3 where they are sorted in order of increasing bond length. The positive charge on Cα undergoes a significant decrease in parallel with the increase in bond length, while the negative charge on N undergoes a parallel but less dramatic increase. The net result is a decrease in charge separation, $|q(C\alpha)-q(N)|$, that parallels the increase in bond length and the anomaly is removed if one relates the increase in bond length to the atomic charges on the Cα and N atoms rather than to the formal charge on the amino group. Thus, $Gly^{(+)}$ and Gly(**z**), even though they possess an overall positively charged NH_3 group (+0.35 and +0.26 au respectively), possess a nitrogen atom that is more negative than the nitrogen atom in Gly(**n**) and $Gly^{(-)}$.

The atomic properties are considered in detail in the next chapter but attention is drawn here to further examples of the bond length decreasing in parallel with an increase in charge separation. The atom Cζ in $Arg^{(+)}$ bears a charge of +2.185 and is bonded to three sp^2-nitrogen atoms. The three associated bond lengths decrease with an increase in charge on the bonded N atom, the charge on Nϵ being the most negative with $q(N\epsilon)$ = -1.462 e. The bond lengths (BL) are found to be linearly related to the charge on N and can be fitted to a simple linear equation with an r^2= 0.95. The same trend is

found in $His^{(+)}$, where Nδ1and Nϵ2, both bonded to Cϵ1, have charges of –1.545 e and –1.504 e respectively, with associated bond lengths of 1.310 and 1.315 Å.

4.5.1.3 The Geometry of the Side-Chain (R)

It is more difficult to provide a systematic comparison of the geometries of the side-chains, since these are generally composed of different combinations of functional groups. Thus, statistics are given in the tabulation of bond lengths, table 4.4, only for bonds between the same atoms. Table 4.5 provides the principle bond angles in the side-chains.

A Cα-Cβ bond (bond 1), is present in all AA but glycine. Despite the widely different substituents on the β carbon atom, the length of this bond does not change significantly from one amino acid to another, as seen from the small standard deviation and the small difference between the maximum and minimum lengths exhibited by this bond. It is worth noting that the Cβ-Sγ bond lengthens upon dimerization of cysteine into cystine and that the shortest S-C bond is the one between a methyl group and the divalent sulfur atom, bond 4 of methionine.

Bond 5 in histidine (Nϵ2-Cϵ1) lengthens significantly upon protonation of Nϵ2, while the neighbouring bond 41 (Cϵ1-Nδ1) undergoes a concomitant shortening. Protonation of the imidazole ring decreases the bond length alternation within the ring and consequently, the protonated form of histidine is expected to exhibit more aromatic character. The carboxylate groups in $Asp^{(-)}$ and $Glu^{(-)}$ exhibit similar bond lengths and angles, including the angles formed with the neighbouring methylene group. The plane defined by the carboxylate group bisects the HCH angle and the resulting pairs of C-C-O angles, are identical for the two molecules, with C-C-O1, trans to the hydrogens of the methylene group, equal to 117° and that for C-C-O2 equal to 114°. The formyl groups terminating the side-chains of Asn(I and II) and Gln exhibit bond lengths with only small differences: the CO, CN, NH and CC bond lengths exhibiting deviations of 0.005, 0.001, 0.0005 and 0.003 Å, respectively.

The bond lengths of hydrogen to second-row atoms (X) are transferable with $\sigma=\pm0.002$ Å, (data not shown). Typical values include: 1.08Å for C-H, 0.95 Å for alcoholic or phenolic O-H, and

1.00Å for N-H. The bond lengths and bond angles of the methyl, methylene, methine and phenyl groups exhibit transferable geometries.

Table 4.4 Side-chain bond lengths (values in Å)

	1	21	22	31	32	41	42
Ala	1.532						
Val	1.549	1.530	1.533				
Leu	1.539	1.537		1.532	1.532		
Ile	1.552	1.539	1.533	1.530			
Met	1.527	1.529		1.817		1.808	
Cys	1.525	1.820					
Cystine	1.535	1.828		2.052			
Ser	1.539	1.395					
Thr	1.533	1.412	1.522				
$Lys^{(+)}$	1.528	1.531		1.533		1.521	
$Arg^{(+)}$	1.536	1.531		1.525		1.472	
His(I)	1.532	1.499		1.374	1.357	1.348	1.371
His(II)	1.541	1.499		1.375	1.355	1.290	1.373
$His^{(+)}$	1.542	1.502		1.385	1.346	1.310	1.383
Asn(I)	1.540	1.525		1.203	1.347		
Asn(II)	1.536	1.520		1.202	1.353		
$Asp^{(-)}$	1.539	1.555		1.232	1.239		
$Glu^{(-)}$	1.541	1.526		1.554		1.236	1.236
Gln	1.536	1.528		1.519		1.355	1.201
Phe	1.546	1.514		1.391	1.392	1.387	1.388
Tyr	1.546	1.513		1.395	1.387	1.382	1.390
Tyr(DFT)	*1.559*	*1.513*		*1.405*	*1.401*	*1.398*	*1.393*
Trp	1.532	1.506		1.353	1.449	1.400	1.400
Pro	1.539	1.531		1.534		1.470	
av	1.537	1.513	1.529				
std	0.007	0.038	0.005				
max	1.552	1.555	1.533				
min	1.525	1.395	1.522				
|max-min|	0.027	0.160	0.011				

	43	51	52	53	61	62
$Lys^{(+)}$		1.519				
$Arg^{(+)}$		1.317			1.325	1.329
His(I)		1.291				
His(II)		1.347				
$His^{(+)}$		1.315				
Phe		1.388	1.386			
Tyr		1.389	1.384		1.354	
Tyr(DFT)		*1.399*	*1.397*		*1.374*	
Trp	1.375	1.395	1.377	1.369	1.377	1.406

Except bond 1, bonds which are unique (in absence of branching) are designated by their number followed by the digit "1", i.e. with the column of the highest priority branch. This is so for economy of space, since it would necessitate to have an additional column for the non-branched. Entries in italics are not included in the statistics. Labelling of bonds is discussed in appendix 4.1 and indicated in figure 4.1.

Table 4.5 Selected side-chain bond angles (values in (°))

	1-2	1-21	1-22	2-3	2-31	2-32	3-4
Val	111.5	110.4	112.6				
Leu	115.2			111.0	109.4	112.5	
Ile	112.6	112.5	112.6				
Met	113.4			110.0			99.9
Cys	114.7						
Cystine	111.8			102.1			
Ser	113.9						
Thr	109.2	105.6	112.8				
$Lys^{(+)}$	114.2			111.3			111.1
$Arg^{(+)}$	114.6			110.8			109.8
His(I)	115.6				124.2	131.2	
His(II)	115.6				121.9	128.5	
$His^{(+)}$	117.3				124.3	130.0	
Asn(I)	116.5				120.0	117.1	
Asn(II)	112.5				121.0	116.2	
$Asp^{(-)}$	114.5				116.5	114.2	
$Glu^{(-)}$	115.4			114.5			
Gln	113.7			112.3			
Phe	113.3				121.5	120.0	
Tyr	113.6				121.9	120.5	
Tyr(DFT)	*113.7*				*121.8*	*120.4*	
Trp	116.1				129.5	124.6	
Pro	102.4			103.0			105.7

	3-41	3-42	31-41	32-42	32-43	4-5	41-5
$Lys^{(+)}$						110.2	
$Arg^{(+)}$						125.6	
His(I)			107.3	110.9			112.2
His(II)			105.9	105.7			112.0
$His^{(+)}$			110.1	107.1			108.1
$Glu^{(-)}$	116.7	114.1					
Gln	123.0	115.0					
Phe			120.7	121.0			
Tyr			121.5	121.6			
Tyr(DFT)			*121.4*	*121.5*			
Trp			110.4	107.2	133.6		

	42-43	42-5	41-51	42-52	43-53	51-52	5-61
$Arg^{(+)}$							120.7
His(I)		105.2					
His(II)		106.8					
$His^{(+)}$		109.1					
Phe			120.0	120.3		119.5	
Tyr			119.9	119.7		119.7	
Tyr(DFT)			*119.8*	*119.6*		*119.8*	
Trp	119.2		109.0	122.2	119.1	130.4	

	5-62	51-6	52-6	52-62	53-63	62-63	
$Arg^{(+)}$	120.2						
Tyr		117.5	122.7				
Tyr(DFT)		*117.4*	*122.8*				
Trp				117.5	120.8	121.3	

Table 4.6 Dihedral angles (°) defining the side-chain conformation

	χ11	χ12	χ21	χ22	χ31	χ32	χ41
Val	-65.8	58.8					
Leu	-58.8		175.2	-61.6			
Ile	63.6	168.3	162.9				
Met	-172.8		174.3		177.0		
Cys	-171.7						
Cystine	-173.3		-173.1		-86.7		-173.1
Ser	-161.5						
Thr	-58.4	63.9					
$Lys^{(+)}$	-168.5		-176.0		178.9		-178.0
$Arg^{(+)}$	-176.7		176.8		179.9		-178.9
His(I)	-170.2		-76.2	105.8	-178.7	178.6	0.2
His(II)	-174.6		-77.7	102.9	-179.6	179.6	0.0
$His^{(+)}$	-160.1		-67.1	116.8	-177.1	176.8	0.2
Asn(I)	-70.8		-145.6	35.3			
Asn(II)	-172.7		75.0	-105.6			
$Asp^{(-)}$	-62.3		-0.3	179.7			
$Glu^{(-)}$	-175.5		174.1		-2.9	177.3	
Gln	-172.5		179.2		18.6	-162.1	
Phe	-72.2		86.7	-93.3	-179.9	179.7	0.1
Tyr	-72.3		82.2	-97.8	-179.9	179.7	0.1
Tyr(DFT)	*-72.7*		*77.9*	*-101.8*	*-179.6*	*179.5*	*0.0*
Trp	-173.3		16.6	-166.2	177.6	-177.9	0.2
Pro	36.1		-36.0		23.3		-0.6

	χ42	χ43	χ511	χ512	χ521	χ522
Cystine			-173.3		64.1	
$Arg^{(+)}$			-0.2		179.8	
His(I)	-0.1		-0.0			
His(II)	-0.1		0.1			
$His^{(+)}$	0.0		-0.2			
Phe	0.3		-0.1		-0.1	
Tyr	0.3		-0.1	-179.9	-0.1	179.8
Tyr(DFT)	*0.2*		*-0.1*	*-0.0*	*-179.9*	*179.8*
Trp	0.2	179.9	-0.2	179.8	-0.1	-0.1
Pro			-22.3			

Dihedral angles are labeled according to the IUPAC-IUB rules. When a third digit is present the one with the smallest third digit is the dihedral formed with the bond with the smallest label. For example, χ521=∠ bond 42-bond 52 - bond 62, and χ522= ∠ bond 51 - bond 52 - bond 62. The values of χ53,χ62, and χ63 of tryptophan are all zero°.

Most C-C-C bond angles cluster around the ideal tetrahedral angle (109.5°) or the ideal sp^2 angle (120°). The departure of the angle **2-32** from the ideal sp^2 angle in Asn(I) and Asn(II) is likely due to strain induced by ring closure as a result of the hydrogen bonding linking Hδ2-1 to main-chain atoms in both cases. In the case of charged carboxylic groups [$Asp^{(-)}$ and $Glu^{(-)}$], the two O-C-C bond angles around the carboxyl carbon atom are smaller than the ideal sp^2 value of 120°, leading to a

positive departure of the O-C-O angle to *ca.* 129° in both molecules. The departures of the O-C-C bond angle are dependent on the stereochemistry with respect to the methylene group to which the carboxyl group is attached. The bond angles formed in a sulphide linking cluster around 100° with values of 98°, 100°, and 102° for C-S-H (Cys), C-S-C (Met), and C-S-S (CysS) respectively. The lengths of the perimeter bonds in the phenyl ring of Trp exhibit substantial alternation as do the associated bond angles, indicating a decrease in its aromatic character relative to benzene.

The optimized dihedral angles (table 4.6) depend on the starting geometries, and while describing the conformation of a minimum on the potential energy surface, the values do not necessarily describe the global minimum. It has been previously demonstrated, however, that rotations around side-chain dihedral angles seldom have major energetic consequences. Thorough *ab initio* conformational studies of glycine, (*Siam et al., 1984*) alanine, (*Cao et al., 1995; Gronert and O'Hair, 1995; Siam et al., 1984*) serine, (*Gronert and O'Hair, 1995; Van Alsenoy et al., 1988*) cysteine, (*Gronert and O'Hair, 1995; Schäfer et al., 1990b*) valine, (*Schäfer et al., 1990a; Shirazian and Gronert, 1997*) and threonine (*Schäfer et al., 1990a*) generally show little difference in the total energy of most conformers and the global minimum (within *ca.* 3 kcal/mol). This general conformational insensitivity of the total energy was also found in both neutral as well as zwitter-ionic leucine. (*Matta and Bader, 2000*) The conformational insensitivity of leucine is accompanied by a general transferability of equivalent atoms from one conformation to another as well as the transferability of bond distances, bond angles, and of the rest of the bond properties. (*Matta and Bader, 2000*) Since dihedral angles are the most sensitive parameters to crystal packing effects, these angles determined experimentally can differ widely for one and the same molecule from one crystal system to another and depending on the presence and on the nature of solvent or counter ion(s) which co-crystallise. (*Cody,Barrett, 1985*) Therefore a detailed discussion of trends in these angles is not warranted.

4.5.1.4 Effect of Hydrogen Bonding on Geometry

The topological theory of structure identifies all bonded interactions, (*Bader, 1998b*) including hydrogen bonds. The present study revealed the presence of nine hydrogen bonds formed between side-chain and main-chain atoms. The bond and atomic properties of these A-H···B interactions are discussed in the following chapter but their geometric parameters are listed here in table 4.7. In general, the strength of a hydrogen bond correlates with the value of the electron density at the bond critical point linking the hydrogen and base atoms (*Carroll and Bader, 1988; Cheeseman et al., 1988*) and this value, ρ_b, is also tabulated. The data in table 4.7 are listed in order of increasing values of ρ_b, an ordering that parallels the decrease in the A-B length, another measure of the bond strength, to within 0.001 au in the value of ρ_b with the exception of the bond in serine, the only AA with a O-H···O interaction. The hydrogen bonding results in the formation of 6- and 7-membered rings and the A-H···B angle is in all cases, considerably less than 180°, table 4.7.

Table 4.7 Hydrogen bonds identified between the side-chain and the main-chain atoms (distances in Å, angles in (°), all other values in au)

AA	B...H-A	B...H	H-A	B-A	∠B...H-A	ρ_b	$\nabla^2\rho_b$
Trp	O2...Hδ1-Cδ1	2.772	1.070	3.350	113.8	0.006	0.023
Asn(II)	O1...Hδ2(1)-O1	2.614	0.996	3.227	119.8	0.007	0.027
Ile	O2...Hγ1(2)-Cγ1	2.716	1.086	3.128	102.0	0.008	0.030
$Lys^{(+)}$	O1...Hγ(1)-Cγ1	2.500	1.083	3.133	116.2	0.011	0.041
His(I)	O1...Hδ1-Nδ1	2.260	0.996	2.936	124.0	0.013	0.054
Ser	O2...Hγ-Oγ	2.177	0.947	2.834	125.5	0.016	0.067
$Asp^{(-)}$	Od1...H1-N	2.220	1.004	2.930	126.4	0.017	0.062
Asn(I)	N...Hδ2(1)-Nδ2	2.183	0.997	2.927	130.3	0.019	0.072
$His^{(+)}$	O1...Hδ1-Nδ1	1.937	1.006	2.723	132.8	0.025	0.110

"A" stands for "Acid" and "B" stands for "Base".

The hydrogen bond with the shortest acid to base (A-B) length and the strongest in terms of the ρ_b value is found in $His^{(+)}$. The H bonded to Nδ1 of the protonated imidazole group in $His^{(+)}$ is bonded to O1, the keto oxygen of the main chain. The strength of this interaction causes a rotation about the Cα-C axis to give the largest of all the calculated deviations from planarity for the main

chain atoms, with $\psi = -143°$. The same three atoms form a hydrogen bond in neutral His(I), but the bond is longer and weaker than in the protonated species and the deviation from planarity is smaller with ψ = -108. The negative charge on Nδ1 is of greater magnitude in $His^{(+)}$ than in His; 1.55e compared to 1.53e, in line with the strength of the hydrogen bond.

The second largest ρ_b value is found for the hydrogen bond in Asn(I). The asparagine molecule forms two hydrogen bonded structures Asn(I) and Asn(II), figure 1, with the former more stable than the latter by 2.8 kcal/mol. The molecule folds back into one of these two compact forms even when energy optimizations are initiated with extended side-chain geometries. The relative energetic stability of these tautomers parallels the strength of their hydrogen bonds, as gauged from the data in table 4.7. The strongest of the two bonds is formed between the main chain N atom and the H of the terminal amino group in Asn(I), causing a moderate deviation from planarity with respect to the angle ψ = -17°. The bond in Asn(II) also involves the H of the terminal amino group but bonded to the keto oxygen O1. This is a relatively weak bond as judged by the long A-B length and low ρ_b value. However, it is sufficient to stabilize the conformation with ψ = -93°.

The hydrogen bond formed between the keto oxygen of the carboxylate group and a hydrogen atom of the main-chain amino group in $Asp^{(-)}$ is only slightly weaker than the bond in Asn(I). This interaction causes an 'inversion' of the amino group in this AA relative to its orientation in the remaining molecules. The ρ_b value for the hydrogen bond in serine is less than those for Asn(I) and $Asp^{(-)}$ even though its A-B length is less. This is the only hydrogen bond in this set that bridges two oxygens, the O-H group of the terminal hydroxyl group of the side chain serving as the acid and the hydroxyl oxygen O2 of the main chain as the base atom. The resulting 6-membered ring causes only a minimal departure from planarity of the main chain atoms, with $\psi = 4°$.

There are three relatively weak hydrogen bonds formed between a hydrogen bonded to a carbon of the side chain and a main chain oxygen atom. Two yield 6-membered rings with a methylenic hydrogen atom bonded to either a keto oxygen (O1), in $Lys^{(+)}$ or to a hydroxyl oxygen (O2)

in Ile. The weakest is found in a 7-membered ring formed by the H bonded to Cδ1 of the indole ring in tryptophan with the hydroxyl oxygen O2 of the main chain. While weak, their presence stabilises the resulting ring structures.

4.5.2 Comparison of Calculated and Experimental Geometries

Because of the low volatility of the solid phases of the amino acids, there are no available experimental geometries from gas phase electron diffraction or spectroscopic studies (*Vilkov et al., 1983*) for comparison with the computed AA geometries reported here. On the other hand, a large number of crystallographic determinations, mostly of amino acids in their zwitter-ionic state or of amino acid residues in small peptides, are available from the literature and can be downloaded from the crystallographic databases and this data shall be used in the following comparisons.

Most of the available geometric information for the amino acids, whether free or in small peptides, comes from either x-ray (X) or neutron (N) diffraction experiments or a combination of the two methods (X-N). Neutron diffraction improves the model used in the x-ray refinement by providing better coordinates for the location of light nuclei, especially hydrogens. X-ray diffraction experiments prior to 1978 used spherical atomic densities exclusively in the crystallographic refinement, a method still widely used as it provides density maps sufficiently accurate to determine the geometric parameters. However, as emphasized in chapter 1, such maps lack the resolution required to obtain experimental displays of the electron density. Since the advent of the MOLLY program, (*Hansen and Coppens, 1978*) aspherical multipolar refinement of the scattering data has been increasingly employed by crystallographers resulting in better agreement between calculated and observed structure factors. Multipolar refinement (*Coppens, 1997; Koritsanszky and Coppens, 2001*) has been increasingly used in x-ray diffraction refinement of data collected at very low temperatures, resulting in accurate electron densities which are now routinely used as a starting point for QT-AIM topological analyses. (*Housset et al., 2000; Koritsanszky and Coppens, 2001; Spackman, 1998; Volkov et al., 2000*) (See also section 1.3 of chapter 1). Some representative examples of the application of QT-AIM to experimental

densities are the use of the atomic moments to obtain an estimate of the electrostatic contribution to the intermolecular interaction energy of histidine, (*Coppens et al., 1999*) and the determination of bond and atomic properties and the Laplacian distributions of several crystallized amino acids (*Dahaoui et al., 1999; Destro et al., 1991; Destro et al., 2000; Espinosa et al., 1996; Flaig et al., 1998*) and related compounds. (*Benabicha et al., 2000; Gopalan R. S. et al., 2000*)

The experimental geometries have several possible sources of discrepancy with respect to those calculated for the free amino acids (AA): (i) The amino acids in the crystal are in their zwitter-ionic state or occur as residues in small peptides; (ii) Crystal packing effects; (iii) The presence of co-crystallized species such as solvent molecules or counter ions; (iv) Experimental error due to crystal imperfections, data collection, thermal and crystal averaging, reflected collectively in the *R*-factor. (Coppens remarks that x-ray diffraction geometries using spherical refinement incorporate an error he terms "aspherical shifts" that does not necessarily show in the *R*-factor, see p.50 of Ref. (*Coppens, 1997*)). On the other hand, the restricted HF method employed in the calculations is known to predict bond lengths that are generally shorter than those predicted by correlated methods or those estimated experimentally. For example, included in the tabulation of bond lengths and angles are values for the Gly and Tyr molecules calculated using the same basis set as in the RHF calculations, but using density functional theory (DFT). DFT includes electron correlation in an empirical manner and in general, leads to larger bond lengths than obtained in a RHF calculation. The DFT bond lengths for these two amino acids are indeed consistently longer than their RHF counter-parts. The DFT bond angles, on the other hand are remarkably similar to the RHF values. In the comparison that follows, one finds that the RHF bond lengths are in better agreement with the experimental crystal data than are the DFT values.

Tables 4.8-4.11 compare selected calculated and x-ray crystallographic geometric parameters of the amino acids. The experimental entries in these tables are either quoted directly from the cited reference, or represent a weighted average over two crystallographically-independent molecules within the same crystal, or a weighted average over two independent studies.

Table 4.8 Comparison of calculated and experimental geometries: bond lengths of the main-chain group (values in Å)

	Calc (c)	Exptl (x)	$\sigma\times10^3$	c-x	Sig	*	Ref
Cα-C							
Thr	1.527	1.537	1	-0.010	S		1
Trp	1.526	1.540	13	-0.014			2
$Arg^{(+)}$	1.526	1.534	1	-0.008	S		3
Asp	1.525	1.537	0	-0.012	S		4
Ile	1.525	1.527	7	-0.002			5
Met	1.525	1.519	6	0.006			6
Val	1.524	1.537	1	-0.013	S	a	7,8
Leu	1.524	1.532	1	-0.008	S	b	9
Ala	1.523	1.536	1	-0.013	S		10
Cys	1.523	1.533	1	-0.010	S	b	11
Tyr	1.523	1.519	4	0.004		b	12
Av	1.525	1.532	3	-0.007	64%		
$\sigma\times10^3$	1	7		0.007			
Av_w	1.525	1.536	0	-0.011	S		
Cα-N							
Met	1.456	1.447	5	0.009			
Cys	1.455	1.486	1	-0.031	S	b	11
Trp	1.452	1.505	14	-0.053	S		
Leu	1.449	1.493	1	-0.044	S	b	9
Ile	1.449	1.461	8	-0.012			
$Arg^{(+)}$	1.445	1.489	1	-0.044	S		
Val	1.445	1.488	1	-0.043	S	a	7,8
Tyr	1.444	1.480	4	-0.036	S	b	12
Asp	1.443	1.491	0	-0.048	S		4
Ala	1.443	1.489	1	-0.046	S		10
Thr	1.441	1.447	1	-0.006	S		1
Av	1.447	1.480	3	-0.032	82%		
$\sigma\times10^3$	5	18		0.019			
Av_w	1.447	1.485	0	-0.038	S		

(a) Indicates weighted averaging over two independent crystallographic determinations, (b) indicates weighted averaging (*Roe, 1992*) over two crystallographically independent entities within the same crystal. Av and Av_w denote a simple arithmetic average or a weighted average respectively. Statistically significant differences between the calculated parameter and the experimental parameter are indicated by "S".(See pages 404-405 of (*Stout and Jensen, 1989*)). References: 1 (*Benabicha et al., 2000*); 2 (*Takigawa et al., 1966*); 3 (*Espinosa et al., 1996*); 4 (*Flaig et al., 1998*); 5 (*Lalitha et al., 1984a*); 6 (*Chen and Parthasarathy, 1977*); 7 (*Dalhus and Görbitz, 1996*); 8 (*Lalitha et al., 1984b*); 9 (*Görbitz and Dalhus, 1996b*); 10 (*Destro and Marsh, 1988*); 11 (*Görbitz and Dalhus, 1996a*); 12 (*Subramanian et al., 1984*).
The same conventions and references apply to tables 4.9 – 4.11.

Table 4.9 Comparison of calculated and experimental geometries: bond lengths of the side-chains (values in Å)[#]

		Calc (c)	Exptl (x)	σ× 10^3	c-x	Sig	*	Ref			Calc (c)	Exptl (x)	σ× 10^3	c-x	Sig	*	Ref
C-S									C···C								
Cys	**2**	1.820	1.813	1	0.007	S	b	11	Trp	**32**	1.449	1.451	14	-0.002			2
Met	**3**	1.817	1.806	4	0.011	S		6	Trp	**62**	1.406	1.396	18	0.010			2
Met	**4**	1.808	1.807	6	0.001			6	Trp	**42**	1.400	1.382	14	0.018			2
Av		1.815	1.809	4	0.006	67 %			Tyr	**31**	1.395	1.379	5	0.016	S	b	12
σ×10^3		5	3		0.004				Tyr	**42**	1.390	1.403	5	-0.013	S	b	12
Av_w		1.815	1.812	1	0.003	S			Tyr	**51**	1.389	1.377	5	0.012		b	12
									Tyr	**32**	1.387	1.381	5	0.006		b	12
C-C									Tyr	**52**	1.384	1.381	5	0.003		b	12
Ile	**1**	1.552	1.547	6	0.005			5	Tyr	**41**	1.382	1.389	5	-0.007		b	12
Val	**1**	1.549	1.542	1	0.007	S	a	8,11	Trp	**52**	1.377	1.400	16	-0.023			2
Tyr	**1**	1.546	1.527	5	0.019	S	b	12	Trp	**62**	1.377	1.399	16	-0.022			2
Asp	**1**	1.541	1.527	0	0.014	S		4	Trp	**43**	1.375	1.412	14	-0.037	S		2
Ile	**21**	1.539	1.510	8	0.029	S		5	Trp	**53**	1.369	1.397	18	-0.028			2
Leu	**1**	1.539	1.535	1	0.004	S	b	9	Trp	**31**	1.353	1.344	15	0.009			2
Leu	**2**	1.537	1.530	1	0.007	S	b	9	Av		1.388	1.392	11	-0.004	21%		5
Arg(+)	**1**	1.536	1.531	1	0.005	S		3	σ×10^3		21	23		0.017			
Ile	**22**	1.533	1.528	9	0.005			5	Av_w		1.388	1.387	2	0.001			
Thr	**1**	1.533	1.544	1	-0.011	S		1									
Val	**22**	1.533	1.528	1	0.005	S	a	8,11	C-N								
Trp	**1**	1.532	1.539	13	-0.007			2	Arg(+)	**4**	1.472	1.458	1	0.014	S		3
Leu	**32**	1.532	1.528	2	0.004		b	9	Trp	**41**	1.400	1.377	13	0.023			2
Leu	**31**	1.532	1.523	1	0.009	S	b	9	Trp	**51**	1.395	1.391	14	0.004			2
Ala	**1**	1.532	1.527	1	0.005	S		10	Arg(+)	**62**	1.329	1.339	1	-0.010	S		3
Arg(+)	**2**	1.531	1.525	1	0.006	S		3	Arg(+)	**61**	1.325	1.328	1	-0.003	S		3
Ile	**31**	1.530	1.526	7	0.004			5	Arg(+)	**5**	1.317	1.331	1	-0.014	S		3
Val	**21**	1.530	1.521	1	0.009	S	a	8,11	Av		1.373	1.371	5	0.002	67%		
Met	**2**	1.529	1.513	6	0.016	S		6	σ×10^3		55	46		0.013			
Met	**1**	1.527	1.534	5	-0.007			6	Av_w		1.373	1.364	0	0.009	S		
Arg(+)	**3**	1.525	1.526	1	-0.001			3									
Cys	**1**	1.525	1.529	1	-0.004	S	b	11	C-O								
Thr	**22**	1.522	1.515	1	0.007	S		1	Thr	**21**	1.412	1.431	1	-0.019	S		
Tyr	**2**	1.513	1.519	5	-0.006		b	12	Tyr	**6**	1.354	1.379	4	-0.025	S	b	12
Asp	**2**	1.509	1.512	0	-0.003	S		4	Asp	**31**	1.330	1.313	0	0.017	S		4
Trp	**2**	1.506	1.530	15	-0.024			2	Av		1.365	1.375	1	-0.009	100 %		
Av		1.531	1.528	4	0.004	65 %			σ×10^3		34	48		0.019			
σ×10^3		11	9		0.010				Av_w		1.365	1.337	0	0.028	S		
Av_w		1.531	1.525	0	0.006	S											

For references, see footnote of table 4.8.

Table 4.10 Comparison of calculated and experimental geometries: Some bond angles around the α-carbon atom (values in (°))[#]

	C-Ca-N	Calc (c)	Exptl (x)	σ	c-x	Sig	N-Ca-R	Calc (c)	Exptl (x)	σ	c-x	Sig	*	Ref
Ala		113.3	110.0	0.3	3.3	S		110.2	109.8	0.3	0.4			10
Thr		113.0	111.9	0.4	1.1	S		111.1	110.5	0.4	0.6			1
Leu		112.9	108.6	0.6	4.3	S		111.0	108.7	0.6	2.3	S	b	9
Tyr		112.7	109.2	0.3	3.5	S		110.8	109.0	0.2	1.8	S	b	12
Val		112.6	110.0	0.2	2.6	S		112.1	110.0	0.2	2.1	S	a	8,11
Ile		111.7	111.5	0.4	0.2			110.5	109.9	0.4	0.6			5
Arg(+)		112.4	109.5	0.3	2.9	S		108.5	109.6	0.3	-1.1	S		3
Trp		111.1	107.0	0.8	4.1	S		108.9	109.7	0.8	-0.8			2
Cys		110.6	109.4	0.6	1.2			109.2	109.9	0.6	-0.7		b	11
Met		110.1	110.4	0.3	-0.3			109.7	113.3	0.3	-3.6	S		6
Av		112.0	109.8	0.4	2.3	70%		110.2	110.0	0.4	0.2	50%		
σ		1.0	1.3		1.5			1.1	1.2		1.7			
Av_w		112.0	110.0	0.1	2.0	S		110.2	110.1	0.1	0.1			

For references, see footnote of table 4.8.

Tables 4.8 and 4.9 compare bond lengths and tables 4.10 and 4.11 bond angles of the main-chain and the side-chain respectively. Statistics describing the central tendencies as well as spread are given for the data set in the tables. Each data set consists of the calculated geometric parameter (Calc); the corresponding crystallographic value or the weighted average of two crystallographic values (Exptl); the crystallographic estimated standard deviation (σ) or the weighted average of two crystallographic σ values; and an indication of the statistical significance of the difference between the calculated and experimental values (Calc-Exptl).

The bond lengths Cα-C and Cα-N compared in table 4.8 exhibit the largest discrepancies because they are most affected by the fact that one is comparing the geometries of zwitter-ions (experimental) with those of free non-zwitter-ionic amino acids. Thus for these bonds, one finds 64% and 82% of the entries respectively to be different in the experimental and the optimized in a statistical sense. The calculated Cα-C and Cα-N bonds are consistently shorter than the experimental counterparts as can be seen from the individual entries, with statistical significance, and from the average difference. Thus on the (simple) average the optimized Cα-C bond is 0.007Å shorter than the

Table 4.11 Comparison of calculated and experimental geometries: some bond angles of the side-chains[#]

	Angle	Calc (c)	Exptl (x)	σ	c-x	Sig		Ref		Angle	Calc (c)	Exptl (x)	σ	c-x	Sig		Ref
CCC									CNC								
Trp	32-43	133.6	131.2	1.0	2.4			2	Arg(+)	4-5	125.6	126.0	0.4	-0.4			3
Trp	2-31	129.5	128.0	1.0	1.5			2	Trp	41-51	109.0	107.4	0.9	1.6			2
Trp	2-32	124.6	126.5	0.9	-1.9			2	Av		117.3	116.7	0.7	0.6	0%		
Trp	42-52	122.2	123.2	1.0	-1.0			2									
Tyr	2-31	121.9	120.2	0.3	1.7	S		12	NCC								
Tyr	32-42	121.6	120.2	0.3	1.4	S	a	12	Trp	51-52	130.4	129.0	1.0	1.4			2
Tyr	31-41	121.5	121.3	0.3	0.2		a	12	Trp	31-41	110.4	111.5	0.9	-1.1			2
Trp	62-63	121.3	119.7	1.1	1.6		a	2	Arg(+)	3-4	109.8	111.4	0.3	-1.6	S		3
Trp	53-63	120.8	124.8	1.1	-4.0	S		2	Av		116.9	117.3	0.7	-0.4	33 %		
Tyr	2-32	120.5	120.8	0.3	-0.3		a	12									
Tyr	41-51	119.9	119.6	0.3	0.3		a	12	NCN								
Tyr	51-52	119.7	120.1	0.3	-0.4		a	12	Arg(+)	5-61	120.7	122.3	0.3	-1.6	S		3
Tyr	42-52	119.7	119.8	0.3	-0.1		a	12	Arg(+)	5-62	120.2	118.7	0.4	1.5	S		3
Trp	43-53	119.1	114.6	1.1	4.5	S		2	Av		120.5	120.5	0.4	-0.0	100 %		
Trp	52-62	117.5	116.4	1.0	1.1			2	σ		0.2	1.8		1.5			
Trp	1-2	116.1	114.2	0.9	1.9			2	Av_w		120.5	121.0	0.2	-0.5	S		
Leu	1-2	115.2	115.7	0.7	-0.5		a	9									
Arg(+)	1-2	114.6	112.7	0.3	1.9	S		3	OCC								
Ile	21-31	113.7	113.7	0.4	0.0			5	Tyr	52-6	122.7	121.1	0.2	1.6	S	a	12
Tyr	1-2	113.6	114.6	0.2	-1.0	S	a	12	Tyr	51-6	117.5	119.0	0.3	-1.5	S	a	12
Met	1-2	113.4	112.6	0.3	0.8	S		6	Thr	1-21	105.6	108.1	0.4	-2.5	S		1
Thr	1-22	112.8	111.9	0.5	0.9			1	Av		115.3	116.0	0.3	-0.8	100 %		
Ile	1-22	112.6	109.2	0.4	3.4	S		5									
Val	1-22	112.6	110.4	0.3	2.2	S	b	8,11	SCC								
Leu	2-32	112.5	110.0	0.9	2.5	S	a	9	Cys	1-2	114.7	114.5	0.6	0.2		a	11
Ile	1-21	112.5	112.3	0.4	0.2			5	Met	2-3	110.0	110.0	0.3	0.0			6
Arg(+)	2-3	110.8	112.3	0.3	-1.5	S		3	Av		112.4	112.2	0.4	0.1	0%		
Val	1-21	110.4	111.4	0.3	-1.0	S	b	8,11	σ		2.4	2.2		0.1			
Leu	2-31	109.4	110.8	0.8	-1.4		a	9	Av_w		112.4	110.9	0.3	1.5	S		
Trp	32-42	107.2	107.7	0.1	-0.5	S		2									
Av		117.4	116.9	0.6	0.5	43 %			CSC								
									Met	3-4	99.9	98.7	0.3	1.2	S		6

For references, see footnote of table 4.8.

experimental counterpart, this value being more pronounced in the case of Cα-N (0.032Å). These values change if one uses the weighted average for the experimental values, which gives more weight to the experimental values with the smaller estimated standard deviation, but the trends remain the same.

Table 4.9 compares bond lengths in the side-chains. Surprisingly, the optimized bond lengths for C-S bonds as well as for saturated C-C bonds appear to be consistently longer than their

experimental counterparts. The majority of the entries for both bonds have a statistically significant difference between the optimized and the experimental values (67% and 65% of the entries respectively). The weighted average of the C-S and C-C bond lengths are also statistically significantly different from the optimized counterparts, surprisingly favouring longer optimized bond lengths. The same behaviour is found in the case of the C-N bonds. The optimized C-N bonds are longer than the experimental counterparts, as is shown in the simple and weighted averages of (Calc-Exptl), the latter being statistically significant in 67% of the entries. These findings contradict the commonly held view that Hartree-Fock theory results in bonds that are shorter than experimental estimates.

The C-O bond is found in alcoholic (Ser, Thr), phenolic (Tyr), and carboxylic (Asp) groups. (The entry for Asp refers to the acid with nonionized side-chain). A simple average indicates that the calculated C-O bond is shorter by 0.009Å than experiment. However, the quality-weighted average indicates an opposite trend. The calculated alcoholic and phenolic C-O bonds are statistically significantly shorter than the experimental ones, but the trend is reversed for the carboxylic C-O bond.

There are no obvious trends in the case of aromatic bonds. One-half of the bonds listed in table 4.9 are longer in the calculated set and seven are longer in the experimental set. The simple average (Calc-Exptl) indicates a slight bond shortening of the optimized bonds with respect to the experimental values. However, when one factors in the quality of the experimental data in the weighted average, this trend is reversed. The results show that the experimental values for aromatic bond lengths are well reproduced by the restricted Hartree-Fock/6-31+G* optimization.

Table 4.10 compares the bond angles involving Cα. The optimized C-Cα-N angle is significantly wider (by 2°) than the experimental angle, while the optimized and the experimental C-Cα-R are statistically equal. The reason for this apparent inconsistency lies in the state of ionization of the main-chain atoms. Matta and Bader (*Matta and Bader, 2000*) find that the average optimized C-Cα-N angle over nine non-zwitter-ionic leucine conformers is 112.1°(±1.4°), while the average of the same optimized angle over the corresponding nine zwitter-ionic conformers is significantly smaller [104.1°(±0.5°)]. The change in the optimized value of this angle induced by tautomerization is possibly

due to the strain induced by the 5-membered ring closure via hydrogen bonding between a quaternized amine hydrogen and the proximal oxygen belonging to the anionic carboxyl group: $\overline{\text{C}_\alpha\text{-N-H}\cdots\text{O-C}}$, a strong hydrogen bond (as evidenced by its ρ_b for example) that appears to occur universally in zwitter-ionic amino acids. (*Matta and Bader, 2000*) Thus the difference between the optimized value of this angle and the corresponding experimental value is likely to arise primarily from the difference in the tautomeric forms of the optimized and experimental molecules. In the same work by Matta and Bader, (*Matta and Bader, 2000*) the average values of the optimized N-Cα-R angles for the nine non-zwitter-ionic and nine zwitter-ionic leucine conformers are respectively: 111.1°(±1.7°) and 111.2°(±2.5°), two statistically equal values, indicating that the value of the N-Cα-R angle is independent of tautomeric state. Therefore, the comparison of this angle between the optimized and experimental geometries provides a good indication of how well the Hartree-Fock method recovers bond angles. A comparison of this angle in table 4.10 reveals that the optimized and the experimental values are identical. This observation carries over to the rest of the angles of the side-chain (in table 4.11). From these observations it can be stated that Hartree-Fock optimization recovers the experimental bond angles more closely than bond lengths (on a percent discrepancy basis).

The last table in this geometric comparison, table 4.11, deals with bond angles in the side-chains. The C-C-C angles fall roughly into three types: (1) Angles clustering around 130° and representing outer junctions to rings, for which the calculated and experimental values are statistically identical; (2) Angles clustering around 120° (125° to 115°) representing angles where the central carbon is of sp^2 hybridization. Only 31% of the 13 entries falling in this range have statistically different calculated and experimental values with no obvious trends; (3) Angles clustering around 111° (107° to 115°) in which the central carbon atom is of sp^3 hybridization. In this set, 67% of the 12 entries exhibit significant differences between calculated and experimental values, even though the differences are rather small and show no systematic trend. From this analysis it appears that the Hartree-Fock method recovers the CCC bond angles well (arithmetic average difference of only 0.5°).

4.5.2 Transferability of the Calculated Geometries from the *Free* Amino Acid Molecules (AA) to the Amino Acid *Residues* in Peptides (Aa)

One of the primary goals of this work is to establish the extent of the transferability of the geometric parameters of the amino acids (especially with regards to the side chains) from their free form to their form as a residue in a peptide. Martín and Bader (*Bader and Martín, 1998; Martín, 2001*) optimized the geometries of the genetically-encoded amino acids in the Gly|Aa|Gly mold at the same level of theory used in the present work. The conformation of the side-chains in the residues are not necessarily the same as the ones for the free amino acids. Even though the main-chain group is dramatically altered in passing from AA to Aa because of the formation of two amide bonds, it is remarkable how little change occurs in the geometric parameters of the surviving main chain functional groups. One sees from table 4.12 how well the geometries of the side-chains calculated in AA reproduce those in Aa. The overall average deviation [Δ_{av} = (parameter calculated for AA - parameter calculated for Aa)$_{av}$] for all the bond lengths is -0.001Å and for all the angles is 0.5°. This shows negligibly small systematic changes in the bond lengths and angles between AA and Aa. The overall average absolute deviation ($|\Delta|_{av}$ = |parameter calculated for AA - parameter calculated for Aa $|_{av}$) for all the angles is 0.7° and for all the bond lengths is -0.002Å, which demonstrates an excellent general agreement between the two sets of geometric parameters across the entire family of amino acids. The same conclusions can be drawn by examining the individual entries in the table and the Δ_{av} and $|\Delta|_{av}$ given for each type of bond or functional groups. The table also allows for cross comparison of functional groups in different AA and Aa. For example, entries for the carboxylate group ($COO^{(-)}$) show that the average C-O bond distance is the same in $Asp^{(-)}$ and in $Glu^{(-)}$ whether in the free form or in a peptide. The same can be said for the formyl group, which retains its essential geometric features in the two forms in both Asn and Gln.

The final set of entries under ring systems in table 4.12 compares the deviation of the C-C-C ring angle in the phenyl group from 120°. The same pattern of increasing deviation from perfect hexagonal symmetry in passing from Phe to Tyr to Trp is found for both the AA and Aa.

The general conclusion is that with remarkably few exceptions, and these not serious, the AA geometric parameters recover those in the peptide residues.

4.6. Conclusions

The transferability of the bond lengths and bond angles found in the main-chains and in corresponding functional groups in the side-chains of the genetically-encoded amino acids has been quantified and found to agree with chemical expectations regarding the characteristic properties exhibited by amino acids, contributors to nature's set of building blocks. The transferability of the geometric parameters of the side chains is found to extend to the corresponding amino acid residues bound in a tripeptide. The demonstrated transferability of conformation-independent geometric parameters leads one to anticipate a parallel high degree of transferability of bond and atomic properties which is explored in the next chapter. It is demonstrated that restricted Hartree-Fock theory, using a basis set containing polarizing and diffuse functions, recovers the experimental values of the geometric parameters for the side chains of the amino acids with an acceptable degree of accuracy. The agreement between computed and experimental geometric parameters yields the following generalities: bond angles are better determined than bond lengths; unsaturated geometries are better determined than saturated ones and bond lengths are better determined for bonds with small rather than large charge separations.

Table 4.12 Comparison of calculated *free* amino acid (AA) side-chain geometries with their counterparts in amino acid *residues* (Aa) within tripeptide molds of the type Gly|Aa|Gly (distances in Å and angles in (°))

Group	Side-Chain	AA	Aa*
$COO^{(-)}$			
C-O(av)	$Asp^{(-)}$	1.235	1.235
C-O(av)	$Glu^{(-)}$	1.236	1.236
$\Delta_{av} = 0.000$; $\lvert\Delta\rvert_{av} = 0.000$			
∠O-C-O	$Asp^{(-)}$	129.3	128.8
∠O-C-O	$Glu^{(-)}$	129.2	129.1
$\angle\Delta_{av} = 0.3$; $\lvert\Delta\rvert_{av} = 0.3$			
Formyl $C(=O)NH_2$			
C-O	Asn	1.202	1.202
C-O	Gln	1.201	1.202
C-N	Asn	1.350	1.349
C-N	Gln	1.355	1.351
$\Delta_{av} = 0.001$; $\lvert\Delta\rvert_{av} = 0.002$			
∠O-C-N	Asn	122.8	122.7
∠O-C-N	Gln	122.0	122.1
$\angle\Delta_{av} = 0.0$; $\lvert\Delta\rvert_{av} = 0.1$			
C-X-H, C-X-C (X =O or S)			
$CH_2(\beta)$-O(γ)	Ser	1.395	1.401
CH(β)-O(γ1)	Thr	1.412	1.413
$CH_3(\epsilon)$-S(δ)	Met	1.808	1.809
$CH_2(\gamma)$-S(δ)	Met	1.817	1.818
$CH_2(\beta)$-S(γ)	Cys	1.820	1.828
$\Delta_{av} = -0.003$; $\lvert\Delta\rvert_{av} = 0.003$			
∠$CH_2(\gamma)$-S(δ)-$CH_3(\epsilon)$	Met	99.9	99.9
∠Cα-$CH_2(\beta)$-O(γ)	Thr	105.6	106.8
∠Cα-$CH_2(\beta)$-O(γ)	Ser	113.9	108.1
∠Cα-$CH_2(\beta)$-S(γ)	Cys	114.7	110.5
$\angle\Delta_{av} = 2.2$; $\lvert\Delta\rvert_{av} = 2.8$			
Guanidino-group			
C(ζ)-N(ϵ)	$Arg^{(+)}$	1.317	1.318
C(ζ)-N(η)	$Arg^{(+)}$	1.327	1.329
$\Delta_{av} = -0.001$; $\lvert\Delta\rvert_{av} = 0.001$			
∠N(ϵ)-C(ζ)-N(η,av)	$Arg^{(+)}$	120.5	120.4
∠N(η1)-C(ζ)-N(η2)	$Arg^{(+)}$	119.1	119.2
$\angle\Delta_{av} = -0.0$; $\lvert\Delta\rvert_{av} = 0.1$			

Group	Side-Chain	AA	Aa*
Hydrocarbon side-chains			
$CH_2(\gamma 1)$-$CH_3(\delta 1)$	Ile	1.530	1.530
CH(β)-CH_3(γ,av)	Val	1.532	1.533
CH(γ)-CH_3(δ,av)	leu	1.532	1.532
CH(β)-CH_3(γ2)	Thr	1.522	1.524
CH(β)-CH_3(γ2)	Ile	1.533	1.535
CH(γ)-$CH_2(\beta)$	leu	1.537	1.543
CH(β)-$CH_2(\gamma 1)$	Ile	1.539	1.539
$\Delta_{av} = -0.002$; $\lvert\Delta\rvert_{av} = 0.002$			
∠$CH_3(\delta 1)$-CH(γ)-$CH_3(\delta 2)$	leu	110.5	110.9
∠$CH_3(\gamma 1)$-CH(β)-$CH_3(\gamma 2)$	Val	110.9	110.4
∠$CH_2(\gamma 1)$-CH(β)-$CH_3(\gamma 2)$	Ile	112.3	111.8
∠CH(β)-$CH_2(\gamma 1)$-$CH_3(\delta 1)$	Ile	113.7	113.9
$\angle\Delta_{av} = 0.1$; $\lvert\Delta\rvert_{av} = 0.4$			
Ring systems			
Imidazole CH($\epsilon 1$)-N($\delta 1$)	His(II)	1.290	1.290
Imidazole CH($\epsilon 1$)-NH($\epsilon 2$)	His(II)	1.347	1.346
Imidazole C(γ)-CH($\delta 2$)	His(II)	1.355	1.354
Imidazole CH($\delta 2$)-NH($\epsilon 2$)	His(II)	1.373	1.374
Imidazole C(γ)-N($\delta 1$)	His(II)	1.375	1.376
Phenyl CH-CH(av)	Tyr	1.388	1.387
Phenyl CH-CH(av)	Phen	1.389	1.379
Phenyl CH-CH(av)	Trp	1.393	1.394
Pyrrolidine $CH_2(\delta)$-NH	Pro	1.470	1.475
Pyrrolidine $CH_2(\beta)$-$CH_2(\gamma)$	Pro	1.531	1.528
Pyrrolidine $CH_2(\gamma)$-$CH_2(\delta)$	Pro	1.534	1.528
Pyrrolidine Cα-$CH_2(\beta)$	Pro	1.539	1.536
$\Delta_{av} = 0.001$; $\lvert\Delta\rvert_{av} = 0.003$			
∠Phenyl |120 - CH-CH-CH|(av)	Phen	0.7	0.4
∠Phenyl |120 - CH-CH-CH|(av)	Tyr	1.0	0.6
∠Phenyl |120 - CH-CH-CH|(av)	Trp	1.4	1.5
∠Pyrrolid.Cα-$CH_2(\beta)$-$CH_2(\gamma)$	Pro	102.4	103.2
∠Pyrrolid.$CH_2(\beta)$-$CH_2(\gamma)$-$CH_2(\delta)$	Pro	103.0	103.5
∠Pyrrolid.$CH_2(\gamma)$-$CH_2(\delta)$-N	Pro	105.7	103.7
∠Imidazole C(γ)-N($\delta 1$)-C($\epsilon 1$)	His(II)	105.9	105.8
∠Imidazole C($\epsilon 1$)-N($\epsilon 2$)-CH($\delta 2$)	His(II)	106.8	107.1
∠Imidazole N($\delta 1$)-C($\epsilon 1$)-N($\epsilon 2$)	His(II)	112.0	111.9
$\angle\Delta_{av} = 0.1$; $\lvert\Delta\rvert_{av} = 0.5$			

Δ_{av} stands for the average of the difference (parameter value for AA - parameter value for Aa), and $\lvert\Delta\rvert_{av}$ stands for the average of the magnitude of the difference (| parameter value for AA - parameter value for Aa |).

* Data for the Aa obtained from (*Martín, 2001*).

APPENDIX 4.1: Proposed Supplementary Labeling Conventions for the Amino Acid Atoms and Bonds

The IUPAC-IUB conventions (*IUPAC-IUB Commission on Biochemical Nomenclature, 1970a; IUPAC-IUB Commission on Biochemical Nomenclature, 1970b*) are adopted to label atoms and bonds and supplemented with additional rules to uniquely identify atoms based on their stereochemistry.

(i) Angles have absolute values between 0 and 180°, with positive and negative values rather than from 0 to 360°. (ii) **n**-Aa refers to the *non-zwitter ionic* amino acid molecule (AA) while **z**-Aa refers to AA in the *zwitter ionic form*, whether it has a neutral or ionized side-chain. (iii) ***Main-chain atoms:*** O1 = oxygen atom bonded to the carbon with a double bond; O2 = oxygen atom bonded to the carbon and to the acidic hydrogen atom; H = the acidic hydrogen of the main-chain group; N = main-chain nitrogen atom; H1 = the hydrogen atom bonded to N exhibiting the smallest magnitude of the dihedral angle H-N-Cα-Hα. (iv) ***Side-chain atoms:*** If only one hydrogen atom is bonded to a second-row atom, this hydrogen is labeled with the same label assigned to the second row atom. If two hydrogen atoms are bonded to the same atom, the hydrogen atom with the smallest absolute value of the dihedral angle to towards the main-chain group is (1) and the other is (2). If three hydrogen atoms are bonded to the same atom, the hydrogen atom with the smallest absolute value of the dihedral angle towards the main-chain group is (1), the two others are (2) and (3) in anticlockwise order when the hydrogen atoms point towards the viewer. (v) ***Bond numbers involving main-chain atoms CαHα(NH_2)COOH:*** 0.*n* (zero point one digit (*n*)): C-O1 = 0.1; C-O2 = 0.2, O2-H = 0.3; Cα-C = 0.4; Cα-N = 0.5; N-H1 = 0.6; N-H2 = 0.7; Cα-Hα = 0.8. (vi) ***Bond numbers involving hydrogen atoms of the side-chain:*** The bond linking an atom (A) of the side-chain to a hydrogen is designated by the number of the bond immediately linking (A) to the atom preceding it in the side-chain (to the side of the main-chain group), supplemented with a decimal digit indicating the atom numbering of the hydrogen atom in question as given by rules 2 and 3 above. Bond numbers of the side-chain skeleton bonds follow the IUPAC-IUB conventions.

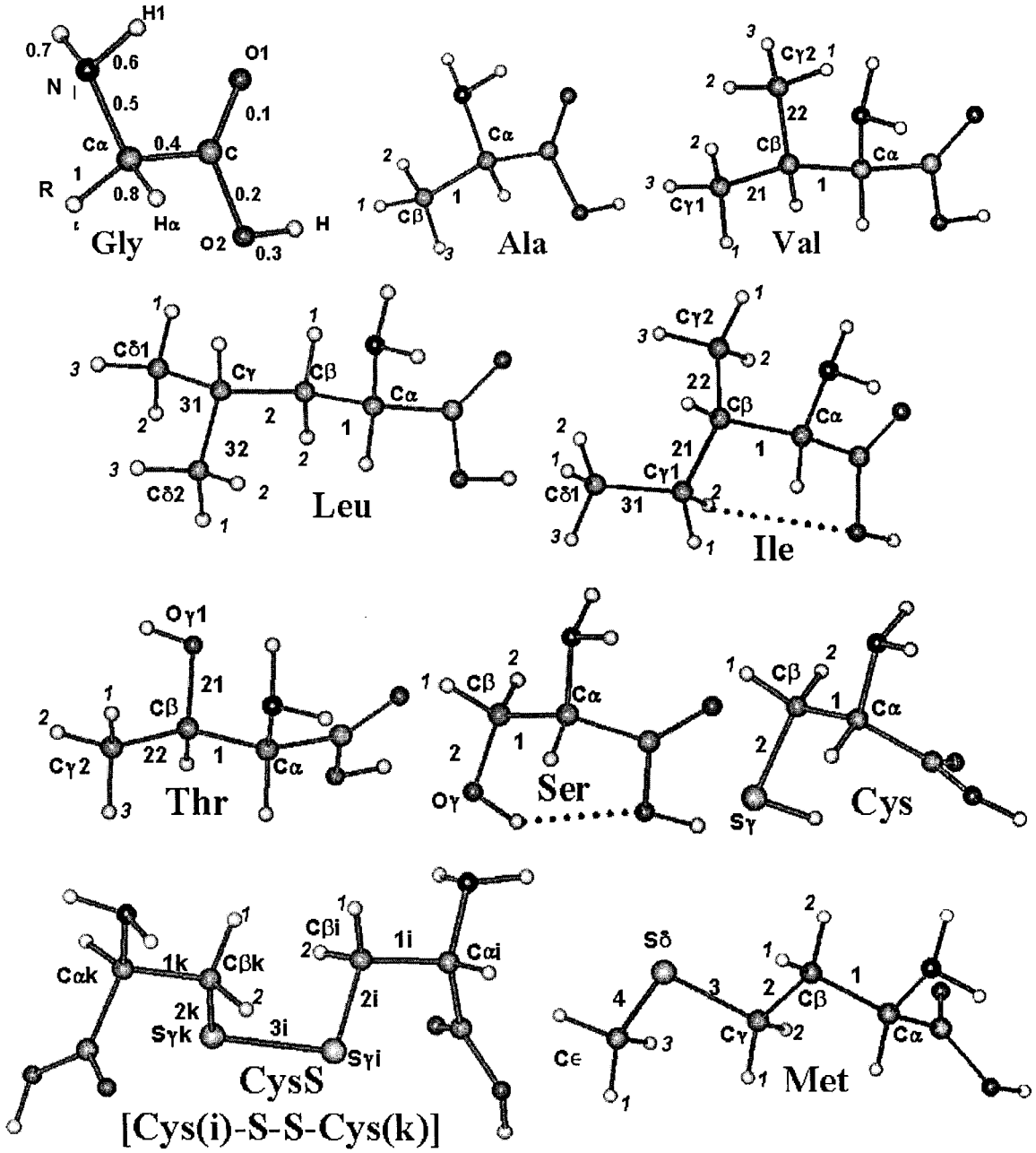


Figure 4.1 A view of the optimized geometries of the amino acids in approximately the same orientation. The IUPAC-IUB labeling conventions supplanted with additional conventions described in appendix 4.1 are indicated. The bond and atom labeling for the common main-chain atoms is only indicated for the simplest amino acid (glycine) but was omitted for clarity in the rest of the structures. Hydrogen bonds are depicted by dotted lines.

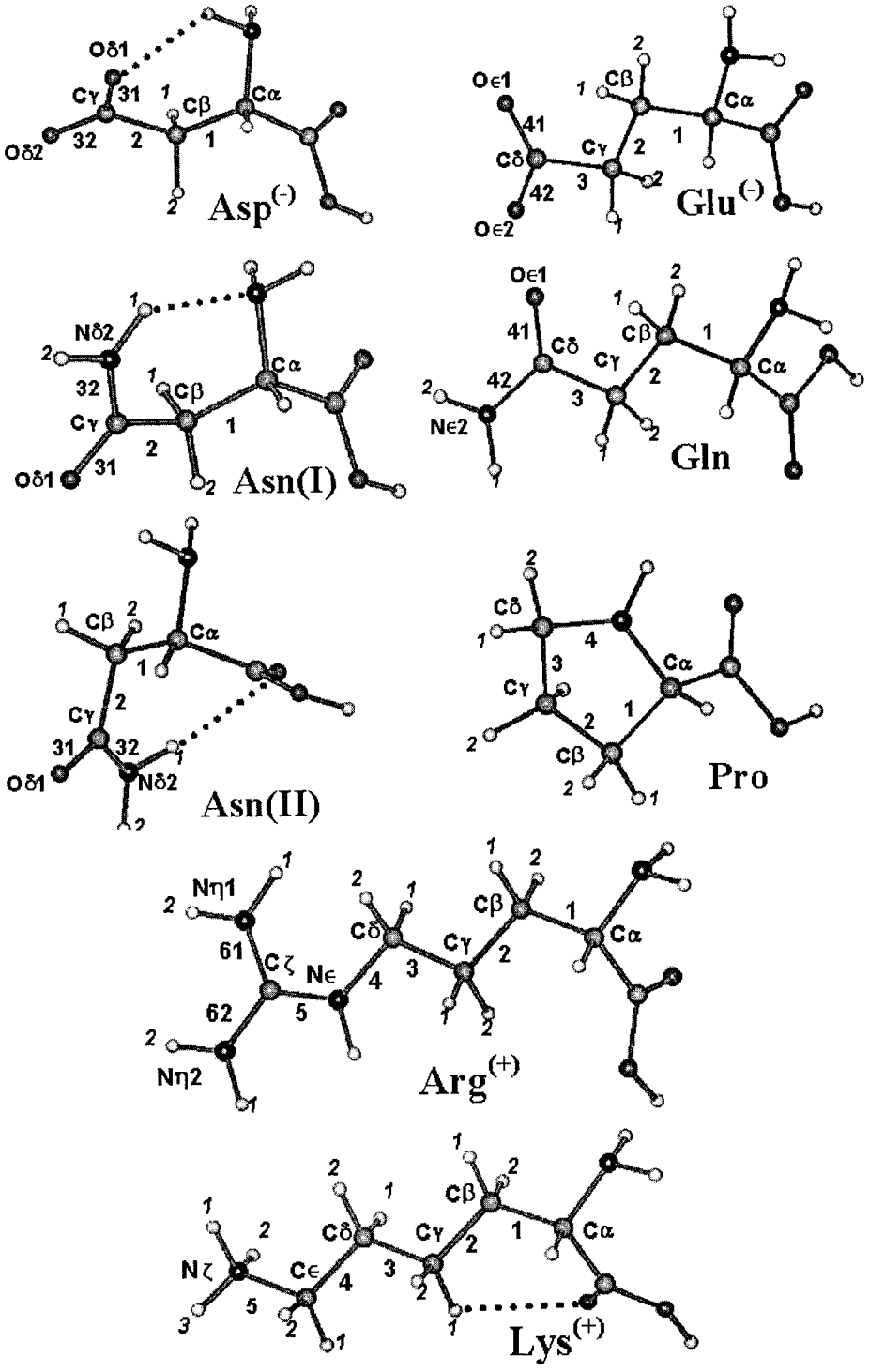


Figure 4.1 (Continued)

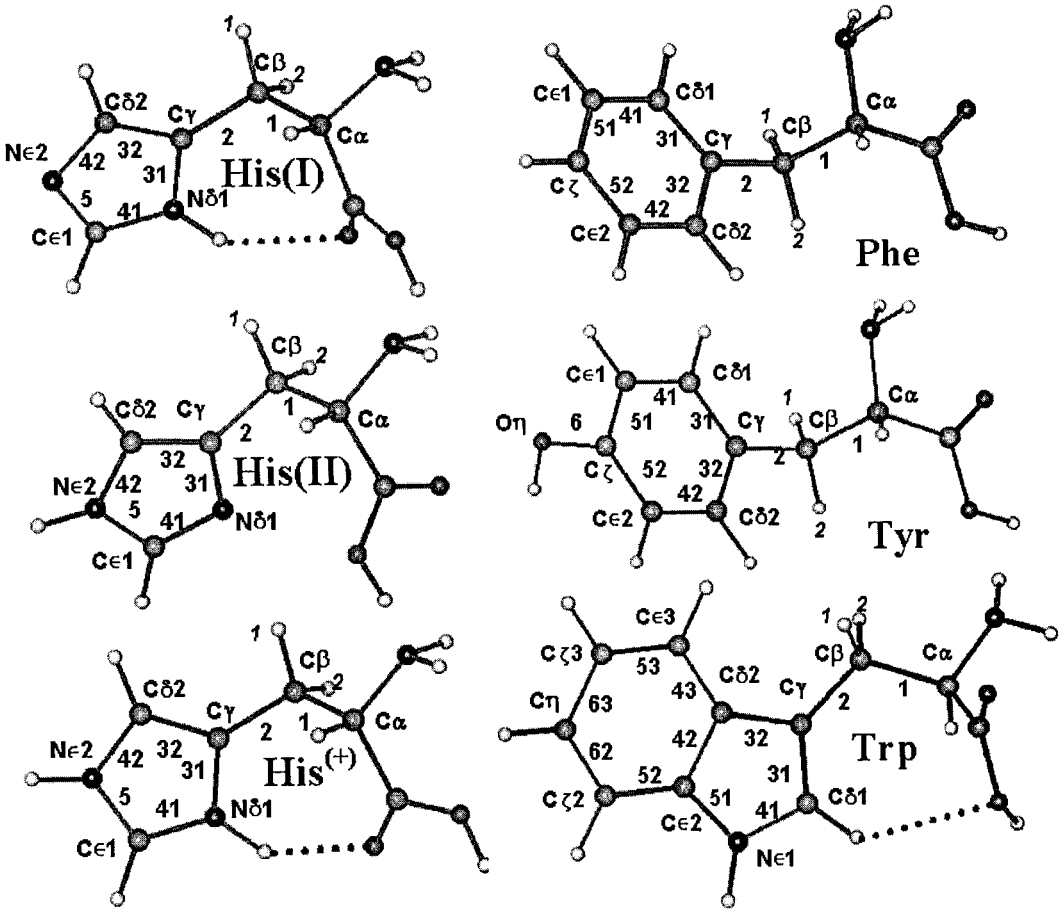


Figure 4.1 (Continued)

Chapter 5

Atomic and Bond Properties of the Genetically-Encoded Amino Acids and their Correlations to Physicochemical and Biological Properties

5.1 Statement of the Problem

The present chapter concludes this part of the thesis and aims at filling an important void in the literature. The properties of every bond and every atom in the entire family of the genetically-encoded amino acids are tabulated and studied in an effort to understand and summarize their chemistry. This chapter describes how one can use the calculated atomic and functional group properties to develop predictive models for the experimental physicochemical and biological properties of this series of molecules. For example, the statistical fit of atomic properties to partial molar volumes and to hydrophobicity measures are explored. A strong correlation is also found between the genetic code of an amino acid and the partial atomic charges as well as the dipole moment of its side chain. The coincidence of a regularity in the genetic code with another regularity in the electrostatic properties of the side-chains adds a quantitative dimension to a known phenomenon described so far in qualitative terms of "polar" or "non-polar" side-chains. These and other correlations described here and elsewhere, (*Matta and Bader, 2002b*) set the stage for future utilization of QT-AIM as a basis in the

development of quantitative structure-to-activity relationships (QSARs) used in drugs and materials design. (*Bader et al.,Alber and Carloni, 2002*).

Another goal of this chapter is to show how these properties - as they relate to a specific functional grouping of atoms (functional group) - are transferable from one molecule to another. This transferability is a prerequisite to develop computational methods which use properties calculated on small model systems to approximate properties of large molecules, not amenable to direct computation, (*Matta, 2001c*) the focus of chapter 6 in the following part of the thesis. This chapter is based on the manuscript (*Matta and Bader, 2002b*).

5.2 Introduction

It is surprising how little has been published about the electron density distribution of the genetically-encoded amino acids, a class of molecules the importance of which can hardly be overemphasised. A detailed study of the electron density of the entire class of amino acids is simply missing from the literature. Such a study is important since it is the density that determines the behaviour of a protein, and hence its biological function. In the process of transcription, the genetic information is converted into physicochemical properties through the electron density of the amino acid residues.

The time is ripe for a comprehensive study of the electron density of these molecules since computational resources, experimental data, and a theory capable of decoding the chemistry in the electron density (QT-AIM) have all become available. The atoms of theory possess the characteristics that are essential to the use of the atomic properties in the understanding and prediction of molecular properties, which were already described in previous chapters and which we would like to re-emphasize briefly: (i) Atomic properties are additive. It is important to emphasise that every measurable property of a molecule exhibits this additive behaviour, including for example, the dipole moment. (ii) The atoms and functional groups defined within QT-AIM maximize the transferability of their properties from one molecule to another. What is remarkable as demonstrated by this theory, is

the exceptional degree of transferability that the electron distribution and hence the properties of an atom or a functional group can exhibit. While such a finding must come as no surprise to chemists used to understanding the properties of a protein in terms of its amino acid residues, it requires a theory of atoms in molecules to implement and make quantitative use of this knowledge.

Chapter 3 establishes the *conformational transferability* of geometric, atomic and bond properties amongst the different conformers of an amino acid as well as the short-range nature of the perturbations induced by such changes as tautomerization. (*Matta and Bader, 2000*) Chapter 4 demonstrates the geometric *transferability* of a functional group from its environment in one amino acid to a different environment in another amino acid, leading one to anticipate an underlying transferability of the electron density. (*Matta and Bader, 2002a*) In the present chapter, we seek to explore the transferability of the bond and of the atomic properties and to present an in-depth study of the electron density of the building blocks of life: the amino acids. (*Matta and Bader, 2002b*)

5.3 Computational Strategy and Methods

The optimized geometries were used in single point calculations to obtain restricted Hartree-Fock (RHF) wavefunctions using a 6-311++G** basis set. The computational details of the geometry optimizations are described in chapter 4. Figure 4.1 and the appendix of chapter 4 serve to display the optimized geometries as well as to label the atoms and bonds of these molecules. The reasons for selecting the RHF/6-311++G**/6-31+G* level of theory are described in the previous chapter. All electronic structure calculations were performed using the Gaussian94 program. (*Frisch et al., 1995*) The resulting electron densities were analysed and atomic integrations were performed using the AIMPAC suite of programs. (*Biegler-König et al., 1982*) Group contributions to the molecular dipole moment were calculated using the program FRAGDIP which is listed in appendix 7.2. (*Matta, 2001b*) Statistical analysis was performed using the MINITAB package. (*Minitab, 2000*)

The satisfaction of the Poincaré-Hopf relationship (Eq. 2.5) was verified for every molecule to ensure that no critical points have been missed especially in cases exhibiting hydrogen bonding. As discussed in chapter 2, the deviation from zero of the integrated Laplacian density measures the quality of atomic integrations, since for a proper open systemd $L(\Omega)=0$ (see Eq. 2.66). In this work, 491 atomic integrations were carried out with an overall average of the magnitude of $L(\Omega)$ equal to 0.0005(± 0.0009) au, where the value in parenthesis is the standard deviation (σ). On the average, the sum of the integrated atomic energies (Eq. 2.67) deviates from the total molecular energy obtained directly from the SCF-calculation by 0.8(±1.0) kcal/mol per amino acid molecule. The sum of the integrated atomic electron populations (Eq. 2.62) deviates from the total molecular electron population by an average of 0.006(±0.003) electrons per molecule. The very small magnitudes of these errors reflect the accuracy of the reported results.

Only those ionization states of the side-chains which predominate at the biological pH are included in the statistical correlations. To reflect the partial ionization of histidine at physiological pH, the contribution of this amino acid in the statistical analysis is 0.934×[Value for His(I)]+0.066×[Value for $His^{(+)}$], where His(I) refers to the more stable of the two tautomers of histidine and $His^{(+)}$ refers to histidine with an ionized side chain.

5.4 Results and Discussion

5.4.1 Bond Properties

Table A5.1 in the appendix lists the properties of each bond in this series of molecules. Table 5.1 emphasizes the transferability of bond properties in the series by listing the average values and the statistical spread for each type of bond present in the amino acids.

Generally bond properties exhibit a small spread, despite the different immediate environments for each type of bonding (see table 5.1). The properties that are the most transferable

include the bonded radii ($r(A)$ and $r(B)$), the density (ρ_b) at the bond critical point (bcp), the kinetic energy density at the bcp (G_b), and the sum of the kinetic and of the potential (V_b) energy densities (the energy density H_b) at the bcp. The Laplacian ($\nabla^2\rho$) and the ellipticity (ϵ), while also generally transferable, can show significant spread about the mean, being more sensitive to the immediate electronic environment. However, if one compares only bonds with the same immediate electronic environment, all bond properties, including $\nabla^2\rho$ and ϵ, exhibit a remarkable transferability. The reader can examine table A5.1 and compare the values for the bond properties of equivalent bonds in different amino acids. Table (A5.1) has been organized to ease such comparison by sorting the bonds in every amino acid in the same order. (Numerical labels of bonds and alphanumeric labeling of atoms was introduced in the previous chapter – see appendix of chapter 4 and figure 4.1). For instance, bond 0.1, the bond between the main-chain atoms C and O1, reveals an almost constant value for every bond property including $\nabla^2\rho$ and ϵ. This transferability is maintained in different conformers e.g. Asn(I) and Asn(II), and even in amino acids carrying a net charge regardless of its sign, positive or negative. The average bond properties summarize, classify, and provide a physical characterization of each type of bonding in a quantitative manner.

Table 5.1 Transferable average bond properties grouped by bond type in the genetically-encoded amino acids (values in au except for bonded radii which are given in Å)

Bond (A-B)	Statistics	r_b(A)	r_b(B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
C-C (Saturated)	Av, n=77	0.765	0.765	0.259	-0.722	0.045	0.054	-0.234
	σ	0.049	0.049	0.008	0.058	0.030	0.004	0.017
	max	0.860	0.860	0.272	-0.627	0.115	0.067	-0.210
	min	0.655	0.655	0.241	-0.836	0.003	0.047	-0.276
	max-min	0.205	0.205	0.031	0.209	0.112	0.020	0.066
	%σ	6%	6%	3%	8%	67%	7%	7%
C-C (aromatic)	Av, n=28	0.694	0.694	0.323	-1.018	0.261	0.101	-0.355
	σ	0.023	0.023	0.010	0.049	0.072	0.011	0.022
	max	0.745	0.745	0.350	-0.826	0.476	0.133	-0.284
	min	0.629	0.629	0.288	-1.130	0.170	0.078	-0.412
	max-min	0.116	0.116	0.062	0.304	0.307	0.056	0.128
	%σ	3%	3%	3%	5%	28%	11%	6%

Table 5.1 (Continued)

Bond (A-B)	Statistics	r_b(A)	r_b(B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
C-H (aromatic)	Av, n=23	0.685	0.389	0.294	-1.105	0.026	0.036	-0.313
	σ	0.009	0.011	0.004	0.033	0.009	0.004	0.004
	max	0.716	0.401	0.305	-1.058	0.047	0.042	-0.306
	min	0.673	0.354	0.288	-1.215	0.013	0.023	-0.327
	max-min	0.043	0.048	0.017	0.157	0.034	0.019	0.021
	%σ	1%	3%	1%	3%	35%	11%	1%
C-H (CH saturated)	Av, n=27	0.688	0.396	0.294	-1.099	0.024	0.039	-0.313
	σ	0.006	0.007	0.003	0.021	0.011	0.002	0.004
	max	0.711	0.414	0.299	-1.057	0.040	0.044	-0.307
	min	0.672	0.373	0.289	-1.141	0.001	0.033	-0.327
	max-min	0.039	0.041	0.009	0.084	0.039	0.011	0.020
	%σ	1%	2%	1%	2%	46%	5%	1%
C-H (CH2 saturated)	Av, n=61	0.682	0.403	0.289	-1.061	0.016	0.042	-0.307
	σ	0.008	0.009	0.004	0.039	0.012	0.003	0.008
	max	0.698	0.426	0.303	-0.999	0.046	0.048	-0.297
	min	0.662	0.386	0.280	-1.215	0.001	0.035	-0.340
	max-min	0.036	0.040	0.023	0.215	0.044	0.013	0.043
	%σ	1%	2%	1%	4%	75%	7%	3%
C-H (CH3 saturated)	Av, n=27	0.678	0.407	0.284	-1.025	0.009	0.044	-0.300
	σ	0.004	0.005	0.002	0.013	0.002	0.002	0.002
	max	0.687	0.413	0.288	-1.005	0.014	0.046	-0.297
	min	0.675	0.395	0.281	-1.055	0.006	0.041	-0.305
	max-min	0.012	0.018	0.008	0.049	0.008	0.005	0.008
	%σ	1%	1%	1%	1%	22%	5%	1%
C-N	Av, n=33	0.519	0.913	0.288	-0.845	0.026	0.181	-0.393
	σ	0.035	0.035	0.031	0.229	0.027	0.069	0.072
	max	0.560	1.038	0.365	-0.052	0.130	0.338	-0.299
	min	0.454	0.854	0.213	-1.166	0.002	0.123	-0.581
	max-min	0.106	0.184	0.151	1.114	0.128	0.215	0.282
	%σ	7%	4%	11%	27%	104%	38%	18%
C-N (aromatic)	Av, n=10	0.450	0.902	0.323	-0.527	0.123	0.369	-0.501
	σ	0.008	0.028	0.030	0.217	0.064	0.037	0.066
	max	0.470	0.935	0.384	-0.216	0.273	0.414	-0.430
	min	0.440	0.846	0.290	-0.940	0.041	0.288	-0.639
	max-min	0.030	0.088	0.094	0.723	0.233	0.127	0.209
	%σ	2%	3%	9%	41%	52%	10%	13%
C-O(H)	Av, n=27	0.431	0.904	0.304	-0.076	0.147	0.437	-0.456
	σ	0.006	0.015	0.014	0.046	0.037	0.030	0.030
	max	0.451	0.961	0.324	-0.010	0.195	0.486	-0.351
	min	0.421	0.886	0.252	-0.293	0.042	0.339	-0.499
	max-min	0.030	0.075	0.072	0.283	0.153	0.147	0.147
	%σ	1%	2%	5%	61%	25%	7%	7%

Table 5.1 (Continued)

Bond (A-B)	Statistics	r_b(A)	r_b(B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
C-S(C)	Av, n=4	0.857	0.961	0.188	-0.367	0.102	0.043	-0.135
	σ	0.003	0.010	0.002	0.011	0.010	0.001	0.003
	max	0.860	0.976	0.191	-0.353	0.112	0.044	-0.131
	min	0.852	0.948	0.186	-0.383	0.086	0.043	-0.140
	max-min	0.008	0.028	0.005	0.030	0.026	0.002	0.009
	%σ	0%	1%	1%	3%	10%	2%	2%
S-S	n=1	1.026	1.026	0.152	-0.176	0.007	0.043	-0.087
S-H	n=1	0.651	0.675	0.226	-0.393	0.388	0.197	-0.295
C=O	Av, n=27	0.397	0.795	0.434	0.100	0.060	0.787	-0.762
	σ	0.002	0.003	0.004	0.073	0.012	0.024	0.010
	max	0.401	0.802	0.443	0.393	0.098	0.850	-0.740
	min	0.389	0.788	0.426	-0.031	0.035	0.735	-0.782
	max-min	0.012	0.014	0.017	0.424	0.063	0.115	0.042
	%σ	1%	0%	1%	73%	20%	3%	1%
C=O(-ve)	Av, n=4	0.410	0.825	0.395	-0.186	0.016	0.620	-0.667
	σ	0.001	0.002	0.003	0.012	0.003	0.008	0.007
	max	0.411	0.828	0.399	-0.172	0.021	0.632	-0.659
	min	0.409	0.822	0.391	-0.204	0.012	0.611	-0.677
	max-min	0.002	0.005	0.007	0.032	0.009	0.021	0.018
	%σ	0%	0%	1%	6%	19%	1%	1%
N-H	Av, n=60	0.745	0.256	0.354	-1.872	0.048	0.055	-0.523
	σ	0.013	0.013	0.001	0.133	0.014	0.006	0.027
	max	0.786	0.271	0.360	-1.719	0.063	0.062	-0.491
	min	0.731	0.221	0.351	-2.349	0.001	0.036	-0.623
	max-min	0.054	0.050	0.008	0.630	0.062	0.026	0.131
	%σ	2%	5%	0%	7%	29%	11%	5%
N-H (aromatic)	Av, n=4	0.766	0.233	0.351	-2.142	0.040	0.042	-0.577
	σ	0.017	0.012	0.006	0.105	0.011	0.006	0.020
	max	0.792	0.247	0.356	-1.996	0.057	0.050	-0.549
	min	0.745	0.214	0.342	-2.286	0.028	0.034	-0.606
	max-min	0.046	0.033	0.015	0.290	0.029	0.016	0.056
	%σ	2%	5%	2%	5%	28%	14%	3%
O-H (acid)	Av, n=24	0.783	0.170	0.375	-2.817	0.017	0.061	-0.765
	σ	0.002	0.002	0.002	0.065	0.001	0.002	0.015
	max	0.789	0.174	0.382	-2.515	0.019	0.067	-0.696
	min	0.778	0.163	0.370	-2.896	0.015	0.056	-0.787
	max-min	0.011	0.011	0.012	0.381	0.003	0.011	0.091
	%σ	0%	1%	1%	2%	6%	3%	2%
O-H (2 alc.+1 phen)	Av, n=3	0.773	0.174	0.384	-2.844	0.023	0.069	-0.780
	σ	0.002	0.002	0.000	0.018	0.001	0.001	0.003
	max	0.775	0.177	0.384	-2.819	0.024	0.071	-0.775
	min	0.770	0.172	0.384	-2.860	0.022	0.067	-0.783
	max-min	0.005	0.004	0.000	0.041	0.002	0.003	0.008
	%σ	0%	1%	0%	1%	4%	1%	0%

Table 5.1 (Continued)

Bond (A-B)	Statistics	r_b(A)	r_b(B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
O...H-C	Av, n=3	1.524	1.164	0.008	0.031	0.399	0.007	0.001
	σ	0.051	0.079	0.002	0.008	0.130	0.002	0.000
	max	1.588	1.221	0.011	0.041	0.572	0.009	0.001
	min	1.464	1.053	0.006	0.023	0.260	0.005	0.001
	max-min	0.124	0.168	0.005	0.019	0.312	0.004	0.000
	%σ	3%	7%	25%	26%	33%	29%	0%
O...H-N	Av, n=4	1.384	0.897	0.016	0.063	0.152	0.014	0.002
	σ	0.105	0.147	0.006	0.030	0.109	0.007	0.001
	max	1.541	1.117	0.025	0.110	0.330	0.024	0.003
	min	1.246	0.705	0.007	0.027	0.035	0.006	0.001
	max-min	0.295	0.413	0.018	0.083	0.295	0.018	0.002
	%σ	8%	16%	38%	48%	72%	50%	50%
O...H-O	n=1	1.330	0.856	0.016	0.067	0.022	0.015	0.002
N...H-N	n=1	1.395	0.790	0.019	0.072	0.053	0.016	0.002

Row Headers: r_b(A) and r_B(B) denote the bonded radii of the two bonded atoms A and B, ρ_b is the density at the bcp, ϵ is the ellipticity, G_b and H_b are the kinetic energy and the total energy densities at the bcp respectively. *Statistical Parameters: n* refers to the number of bonds included in the average (Av), the statistical number of degrees of freedom, σ is the standard deviation, max and min the bounding values defining the range of the data, and %σ = |(std/Av)|×100. When the two bonded atoms are the same element, r_b(A)=r_b(B) = [r_b(A)+r_b(B)]/2.

5.4.2 Atomic and Group Properties

A listing of selected atomic properties of all the atoms in the amino acid series is given in table A5.2 of appendix 5.1. The listed integrated properties include the atomic energy ($E(\Omega)$) – Eq. 2.67; the atomic volume integrated to the 0.001 au external isodensity envelope (v(Ω)); the net atomic charge ($q(\Omega)$) – Eq. 2.63; the components and magnitude of the dipole moment vector ($\mu_x,\mu_y,\mu_z,|\boldsymbol{\mu}|$) – Eqs. 2.68 and 2.69 respectively; two of the components of the diagonalized quadrupole moment tensor ($\mathcal{Q}_{xx},\mathcal{Q}_{yy}$) – Eq.2.71 - the third component being redundant as discussed in chapter 2; the quadrupolar polarization magnitude ($|\mathbf{Q}|$); the five independent components of the original (non-diagonalized) quadrupole tensor Q_{ij}; and the Cartesian coordinates (x, y, z) of the nuclei at the equilibrium geometries optimized at the RHF/6-31+G* level. The listed electric multipoles together with the atomic coordinates can be used to expand the molecular electrostatic potential.

The diagonalized (rotated) quadrupolar polarization tensor components reflect the quadrupolar polarization within the atomic basin. (See section 2.5.2). These three components reflect the transferability of an atom from one amino acid molecule to another, a transferability which is not apparent in the non-diagonalized quadrupole tensor. It is to be noted when comparing individual components of the diagonalized quadrupole tensor listed in table A5.2 of appendix 5.1 that the principal axes can be interchanged from one molecule to another. For example, comparing $\mathcal{Q}_{xx}$ and $\mathcal{Q}_{yy}$ of the atom labelled "C" in the table and belonging to Gly with that of Ala, one realizes that the axes are interchanged so that one finds Gly($\mathcal{Q}_{yy}$) ≈ Ala($\mathcal{Q}_{xx}$) ≈ -0.65 au, Gly($\mathcal{Q}_{xx}$) ≈ Ala($\mathcal{Q}_{zz}$) ≈+ 1.53 au (values for $\mathcal{Q}_{zz}$ are found using the traceless property – Eq. 2.73), and Gly($\mathcal{Q}_{zz}$) ≈ Ala($\mathcal{Q}_{yy}$) ≈ -0.89 au. Thus, from this table one can witness not only the transferability of the magnitudes of the dipolar and quadrupolar polarizations, but one can also follow the transferability of individual components as well, taking into consideration the possible interchanges of the principal axes.

Table 5.2 illustrates the transferability of some functional group properties. From the table, the group properties show little spread, which again demonstrates the transferability of the electron density despite changes in the environment. As for the bond properties, the reader is invited to compare individual atomic properties of equivalent atoms in different amino acids, properties which are organized and sorted in table A5.2 to facilitate such a comparison. The marked transferability of bond properties is parallelled by a marked transferability of the integrated atomic properties. The transferability of the density which is anticipated from the transferability of the geometric parameters (chapter 4) is made quantitative in terms of both bond and atomic properties in the present chapter.

Table 5.2 Transferability of some functional group properties in the amino acid molecules*

Group	Statistics	$q(\Omega)$	$N(\Omega)$	$E(\Omega)$	$v(\Omega)$
CαH	Av, n=23	0.5759	6.4241	-38.13645	87.91
	σ	0.0209	0.0209	0.01808	2.18
	max	0.6259	6.4611	-38.09521	96.53
	min	0.5389	6.3741	-38.19050	85.55
	|max-min|	0.0870	0.0870	0.09529	10.98
	%σ	4%	0%	-0%	2%

Table 5.2 (Continued)

Group	Statistics	q(Ω)	N(Ω)	E(Ω)	v(Ω)
C=O	Av, n=3	0.4112	13.5888	-112.36040	170.05
amidic	σ	0.0144	0.0144	0.01276	1.11
	max	0.4311	13.6025	-112.34236	171.02
	min	0.3975	13.5689	-112.36951	168.49
	|max-min|	0.0336	0.0336	0.02715	2.53
	%σ	4%	0%	-0%	1%
$COO^{(-)}$	Av, n=2	-0.8413	22.8413	-187.64730	318.08
	σ	0.0057	0.0057	0.00672	2.92
	max	-0.8356	22.8469	-187.64058	321.00
	min	-0.8469	22.8356	-187.65402	315.15
	|max-min|	0.0114	0.0114	0.01344	5.84
	%σ	-1%	0%	-0%	1%
COOH	Av, n=23	-0.2359	23.2359	-188.34045	303.89
	σ	0.0265	0.0265	0.05167	3.20
	max	-0.1945	23.3096	-188.12045	309.08
	min	-0.3096	23.1945	-188.37466	292.64
	|max-min|	0.1151	0.1151	0.25422	16.44
	%σ	-11%	0%	-0%	1%
NH_2	Av, n=27	-0.4213	9.4213	-55.89902	173.56
	σ	0.0354	0.0354	0.07354	3.19
	max	-0.3559	9.5182	-55.79929	178.05
	min	-0.5182	9.3559	-56.06892	165.94
	|max-min|	0.1623	0.1623	0.26963	12.11
	%σ	-8%	0%	-0%	2%
OH	Av, n=3	-0.6416	9.6416	-75.85538	138.52
2 alc.,1phen.	σ	0.0073	0.0073	0.02539	1.99
	max	-0.6317	9.6489	-75.83637	141.25
	min	-0.6489	9.6317	-75.89126	136.57
	|max-min|	0.0172	0.0172	0.05489	4.68
	%σ	-1%	0%	-0%	1%

* See footnote of table 5.1 for the definitions of the statistical parameters

5.4.3 Atomic Properties Classification of the Genetically-Encoded Amino Acids

All genetically-encoded amino acids have the main-chain group (MCG), -CαH(NH_2)COOH, in common. This group has a population of 39.07 au and thus generally carries a slight negative charge (-0.07 au) resulting from the near cancellation of the negative charges carried by the NH_2 group (-0.41 au) and the COOH group (-0.24 au) with the net positive charge carried by CαHα (+0.58 au). (See table 5.3).

Table 5.3 Some properties of the main chain group (MCG), -CαH(NH_2)COOH, (in au)

	Total(MCG)		COOH		CH		NH_2	
	q(MCG)	v(MCG)	q(COOH)	v(COOH)	q(CH)	v(CH)	q(NH_2)	v(NH_2)
$Glu^{(-)}$	-0.208	573.0	-0.291	306.7	0.539	88.3	-0.455	178.0
Pro*	-0.172	552.1	-0.242	305.3	0.544	90.6	-0.474	156.1
Ile	-0.138	563.5	-0.261	303.8	0.546	86.9	-0.422	172.9
$Asp^{(-)}$	-0.132	568.4	-0.310	309.1	0.587	85.6	-0.410	173.7
Val	-0.116	562.7	-0.245	304.3	0.555	87.3	-0.426	171.1
Leu	-0.100	565.1	-0.246	303.9	0.565	85.6	-0.419	175.6
Gln	-0.096	568.2	-0.250	305.5	0.575	87.6	-0.421	175.1
Met	-0.087	569.0	-0.237	304.3	0.558	87.8	-0.407	176.9
Ser	-0.082	565.2	-0.250	302.6	0.578	88.2	-0.411	174.3
Ala	-0.080	570.9	-0.240	306.5	0.584	88.7	-0.424	175.6
Tyr	-0.079	566.5	-0.239	305.6	0.582	88.5	-0.422	172.3
Trp	-0.079	566.3	-0.234	304.2	0.567	85.9	-0.411	176.1
Phe	-0.074	565.0	-0.240	304.4	0.582	88.4	-0.416	172.2
His(I)	-0.073	563.8	-0.249	300.1	0.583	87.9	-0.407	175.9
His(II)	-0.067	567.2	-0.215	303.3	0.568	88.0	-0.420	175.9
Asn(I)	-0.055	558.7	-0.202	303.9	0.567	88.0	-0.420	166.7
Thr	-0.043	560.8	-0.215	303.4	0.570	87.7	-0.398	169.6
Cys	-0.037	565.8	-0.226	303.4	0.597	85.8	-0.408	176.6
$Lys^{(+)}$	-0.031	562.1	-0.217	300.1	0.561	88.6	-0.376	173.4
CysS	-0.017	567.0	-0.205	304.6	0.585	87.4	-0.397	174.9
$Arg^{(+)}$	-0.014	565.1	-0.214	303.6	0.586	87.0	-0.386	174.5
Gly	-0.006	581.8	-0.218	307.9	0.626	96.5	-0.414	177.4
Asn(II)	0.010	564.6	-0.194	301.3	0.610	86.2	-0.406	177.0
$His^{(+)}$	0.034	551.3	-0.209	292.6	0.600	87.8	-0.356	170.9
av	-0.073	565.2	-0.235	303.8	0.576	87.9	-0.413	173.4
$\sigma \times 10^3$	0.055	6.0	0.026	3.1	0.020	2.1	0.022	4.5
max	0.034	581.8	-0.194	309.1	0.626	96.5	-0.356	178.0
min	-0.208	551.3	-0.310	292.6	0.539	85.6	-0.474	156.1
|max-min|	0.242	30.5	0.116	16.5	0.087	10.9	0.118	21.9

* Values for Pro equal twice the contribution of the amine H atom and are excluded from the averages for MCG and for NH_2.

The classification of the -CαH(NH_2)COOH group based on total charge shows interesting trends (table 5.3). Among the ones carrying the most negative charges (from -0.208 to -0.100 au) are the amino acids with negatively charged side chains ($Glu^{(-)}$, and $Asp^{(-)}$), and those with long saturated aliphatic hydrocarbon chains (Ile, Val, and Leu). The imino acid Pro is included in the comparison by multiplying the contribution of its single amino H atom by two. These amino acids possess the most electropositive R groups in the series. On the other extreme, one finds the positively charged $His^{(+)}$, which is the only amino acid with a slight net positive charge (+0.034 e) on the main-chain group and hence one of only two R groups which is more electronegative than the main-chain group. This is due

to the close proximity of the positively charged imidazole ring to the main-chain group, being separated from the latter by one methylene group only. The other exception is exhibited by Asn(II). The small magnitude of the charge carried by the main-chain group of Asn(II), which is 0.06 au more deficient than the more stable conformer Asn(I), can be traced to its atomic origin by inspection of table A5.2 in the appendix. For example, the atom that shows maximum difference between these two conformers is N. Besides being bonded to Cα and its two hydrogen atoms (H1 and H2), in Asn(I) N is hydrogen bonded to Hδ2-1 of the side chain from which it withdraws electrons, explaining why this atom is slightly more negative than its counterpart in Asn(II). (See figure 4.1).

Other than $His^{(+)}$, the least electronegative R group include the two other positively-charged ones ($Lys^{(+)}$ and $Arg^{(+)}$), as well as Cys, CysS, and Gly. The intermediate range between these two extremes is represented by the aromatic side-chains (Tyr, Trp, Phe, and His). These induce a charge between -0.079 and -0.073 au in the main-chain group, and have very similar properties. Upon ionization of the histidine R group from His to $His^{(+)}$, the main-chain group loses only 0.107 au and the positive charge acquired upon ionization is mainly localized in the side-chain.

The main-chain group of glycine, having only a hydrogen atom as its side-chain, has no reservoir of electrons to donate or accept. The main-chain group of this amino acid is the closest to neutrality (q(R)=-0.006 au). The glycine main-chain group has the largest volume in the series (15.8 au larger than the average), more than half of the difference being due to the CH group alone.

5.4.4 Correlation Between the Genetic Code and the Side-Chains Group Dipole

It is possible to express the dipole moment of a molecule in terms of atomic or group contributions. The manner in which this is achieved is discussed in detail, together with the underlying theory, in chapter 7. In this section, we wish to discuss the contributions of the side-chains to the molecular dipole moments of the amino acids. We are fortunate in this regard that the electrically-neutral amino acids generally carry a small negative charge on the main-chain group ($-0.138 \leq q(\mathrm{MCG}) \leq 0.010$ au) and, hence, a similar small and opposite charge on the side chains. The approximate electrical

neutrality of the side-chains of uncharged amino acids implies an almost origin-independent group dipole moment for these side-chains. The calculation of the dipole moment of the side chains determined as a proper open system is made possible using the program FRAGDIP. (*Matta, 2001b*)

Table 5.4 compares the molecular dipoles obtained from the group contributions with the molecular dipoles generated directly in the SCF calculation. The table shows an excellent agreement between the reconstructed and direct dipoles not only in magnitude but in individual components as well. The dipole group contributions of the side chains and those of the main-chain group are given in table 5.5. Both tables 5.4 and 5.5 are sorted in order of increasing side-chain group dipole magnitude, an important physical property which is frequently discussed in qualitative terms and which is hereby given a quantitative basis. When sorted in this manner, and upon listing of the second codon-letter in the genetic code encoding the amino acid in question in table 5.5, a pattern emerges: the upper part of the table is dominated by amino acids having a small side-chain dipole moment magnitude with a pyrimidine base as their second codon-letter, while the lower part is dominated by amino acids having large side-chain dipole magnitudes with a purine as the second codon-letter.

The importance of the second letter in the triplet genetic code in determining the physical properties of the encoded amino acids has been recognized for quite some time. (*Pelc, 1965; Volkenstein, 1965; Volkenstein, 1966; Volkenstein, 1977; Woese, 1965*) More recently, Wolfenden and co-workers (*Wolfenden et al., 1981*) used the enthalpies of transfer from the gas phase to the aqueous phase of a series of model molecules representing the side-chains (where the main-chain group in replaced by a hydrogen atom) to classify the amino acids. This group was able to show that this classification reveals regularities in the genetic code in which the second mRNA codon-letter is predominantly a pyrimidine base (either uracil (U) or cytosine(C)) for hydrophobic amino acids and predominantly a purine base (either adenine (A) or guanine (G)) for hydrophilic amino acids.

Since the experimental physicochemical and biological properties are reflections of the underlying electron density, the empirical success of this type of classification leads to the question of

whether the regularity holds at a more fundamental level. In other words, *is the genetic message translated into a regularity in the electron densities of the encoded amino acids*?

Table 5.4 Comparison of the molecular dipole moment obtained from the side chain and the main chain group contributions with that obtained directly from the SCF calculation (entries have the same sorting as in table 5.5, and all values are in au)

		μ_x	μ_y	μ_z	$\|\mu\|$
Gly	Rec	0.409	-0.307	-0.002	0.512
	SCF	0.402	-0.317	-0.001	0.512
Ile	Rec	-0.233	0.533	-0.341	0.674
	SCF	-0.207	0.537	-0.269	0.635
Ala	Rec	0.251	0.481	-0.193	0.576
	SCF	0.252	0.480	-0.193	0.576
Val	Rec	-0.006	-0.473	-0.328	0.575
	SCF	-0.013	-0.473	-0.319	0.571
Leu[#]	Rec	-0.127	-0.424	0.244	0.506
	SCF	-0.119	-0.412	0.264	0.503
Phe	Rec	0.118	0.501	-0.175	0.543
	SCF	0.077	0.495	-0.187	0.535
Tyr	Rec	-0.154	-1.024	-0.266	1.069
	SCF	-0.137	-1.026	-0.272	1.071
Thr	Rec	0.514	-0.904	-0.483	1.147
	SCF	0.515	-0.904	-0.479	1.146
Met	Rec	-0.665	-0.642	-0.709	1.165
	SCF	-0.661	-0.647	-0.695	1.157
Trp	Rec	0.270	1.225	-0.313	1.293
	SCF	0.245	1.233	-0.315	1.296
Cys	Rec	-1.058	0.107	-0.029	1.064
	SCF	-1.047	0.118	-0.033	1.055
Ser	Rec	-0.433	-0.519	0.127	0.688
	SCF	-0.441	-0.528	0.130	0.700
CysS	Rec[a]	0.057	0.381	0.012	0.385
	Rec[b]	0.045	0.382	0.007	0.384
	SCF	-0.000	0.364	-0.001	0.364
Gln	Rec	-0.261	-0.860	-0.335	0.959
	SCF	-0.255	-0.894	-0.339	0.989
Asn(II)	Rec	0.536	-0.162	-0.796	0.973
	SCF	0.534	-0.169	-0.793	0.971
His	Rec	-1.890	-0.625	0.295	2.013
	SCF	-1.860	-0.637	0.303	1.989
Asn(I)	Rec	-0.982	0.980	-0.446	1.457
	SCF	-0.983	0.988	-0.458	1.467

The entry for leucine is that of rotamer Leu(**A**), the global minimum. (See chapter 3 and figure 3.1(b)). (a) Indicates the Cys-Cys dimer where the side chain includes the k-main-chain group, while in (b) the side chain includes only atoms between the i-main-chain group and the k-main-chain groups.

Table 5.5 Group-dipole moments of the side-chains (R) and of the main-chain groups (MCG), $C\alpha H(NH_2)COOH$, sorted in order of increasing side-chain dipole magnitude (in au)*

	m-RNA	m-RNA	R				MCG			
			μ_x	μ_y	μ_z	$\|\mu\|$	μ_x	μ_y	μ_z	$\|\mu\|$
Gly	Purine	G	-0.006	0.068	-0.098	**0.119**	0.415	-0.376	0.095	0.568
Ile	Pyrimidine	U	0.130	0.012	-0.074	**0.150**	-0.363	0.521	-0.267	0.689
Ala	Pyrimidine	C	0.034	-0.108	-0.120	**0.165**	0.217	0.590	-0.073	0.632
Val	Pyrimidine	U	-0.103	0.149	-0.062	**0.191**	0.097	-0.621	-0.266	0.683
LeuA	Pyrimidine	U	0.069	0.182	0.081	**0.211**	-0.195	-0.607	0.163	0.658
Phe	Pyrimidine	U	0.217	-0.142	-0.118	**0.285**	-0.099	0.642	-0.057	0.652
Tyr	Purine	A	-0.292	-0.401	-0.201	**0.536**	0.138	-0.622	-0.065	0.641
Thr	Pyrimidine	C	0.399	-0.259	-0.548	**0.726**	0.114	-0.645	0.065	0.658
Met	Pyrimidine	U	-0.377	-0.636	-0.327	**0.809**	-0.288	-0.005	-0.382	0.479
Trp	Purine	G	0.354	0.727	-0.140	**0.820**	-0.084	0.498	-0.173	0.534
Cys	Purine	G	-0.692	0.090	0.450	**0.830**	-0.366	0.017	-0.479	0.603
Ser	Pyrimidine	C	-0.823	-0.091	0.174	**0.847**	0.390	-0.428	-0.047	0.581
Ser	Purine	G	-0.823	-0.091	0.174	**0.847**	0.390	-0.428	-0.047	0.581
CysS[a]	Purine	G	0.067	0.704	-0.568	**0.907**	-0.010	-0.323	0.580	0.663
CysS[b]	Purine	G	0.043	1.024	-0.005	**1.025**	-0.010	-0.323	0.580	0.663
Gln	Purine	A	-0.206	-1.296	-1.030	**1.668**	-0.056	0.436	0.695	0.823
Asn(II)	Purine	A	0.918	-0.951	-1.248	**1.818**	-0.382	0.789	0.452	0.986
His	Purine	A	-1.510	-0.722	0.782	**1.848**	-0.380	0.097	-0.487	0.625
Asn(I)	Purine	A	-0.963	1.631	-0.197	**1.904**	-0.019	-0.651	-0.249	0.697
av										0.657
$\sigma\times10^3$										0.110
max										0.986
min										0.479

* The column labeled "m-RNA" lists the second codon-letter as it appear in messanger-RNA. The three dipole moment components are labelled μ_x, μ_y, and μ_z, respectively and $|\mu|$ indicates the dipole moment magnitude of the side-chain (R) and of the common main-chain group (MCG). (a) Indicates the Cys-Cys dimer where the side chain includes the k-main-chain group, while in (b) the side chain includes only atoms between the i-main-chain group and the k-main-chain groups.

A property which is likely to be meaningfully correlated to the genetic code is the group dipole moment of the amino acid side-chains. In the study of this correlation, the side-chain dipole contributions of only those amino acids bearing no net charge are considered, since a dipole moment cannot be assigned uniquely to a charged species. When the amino acids are sorted in increasing magnitudes of the side-chain group contributions to the molecular dipole moment, and juxtaposing the second mRNA codon-letter of each amino acid, a pattern emerges which is not dissimilar to the one discovered by Wolfenden et al. (*Wolfenden et al., 1981*) Table 5.5 shows that amino acid side-chains with very small dipoles ($0.090\leq|\mu|\leq0.285$ au), with the exception of glycine, all have a pyrimidine

(mostly uracil) as the second codon-letter. These include the side chains of Leu, Ile, Ala, Val, and Phe, all of which are hydrocarbons with no hetero-atoms. The amino acids with the largest dipole magnitude ($1.668 \leq |\mu| \leq 1.848$ au), which include Gln, Asn, and His, all have the purine base adenine as the second codon-letter and all have two hetero atoms. The intermediate range ($0.536 \leq |\mu| \leq 0.907$ au) include amino acids having a purine (Tyr, Met, Trp, and Cys) and others a pyrimidine (Thr, and Ser). All side chains in the intermediate range have one hetero atom. It is also notable how this sorting lists consecutively the two amino acids with an aromatic benzene ring (Phe, Tyr) as well as the bioisosteric amino acids with a thioalcohol (Cys) and alcohol (Ser) groups.

The regularities in table 5.5 can be slightly sensitive to conformation since the dipole moment depends on the conformation: for example, the two asparagine conformers Asn(I) and Asn(II) which possess side-chain dipole magnitudes that differ by less than 0.1 au. This example shows that conformational freedom can change the ordering slightly, yet it is expected that the general trends in the table will essentially be the same.

Table 5.5 also shows the high degree of transferability of the main-chain group dipole, the magnitude of which clusters around 0.6 au. Notable exceptions include Met, Gln, and Asn, three amino acids which adopt conformations that are significantly different from the other members of the amino acid series by having the -COOH group twisted with respect to the N-Cα-C(OOH) plane. (See figure 4.1).

5.4.5 Correlation Between the Genetic Code and Side-Chain Polarity

Even though conformational changes are not likely to cause major changes in the dipole magnitudes or in the resulting classification, one can seek to investigate correlations with properties that are not conformation-dependent. Such properties include the integrated atomic charges of the side-chains, which were shown to be rather insensitive to conformation in chapter 3. (*Matta and Bader, 2000*) These charges can provide the basis for the definition of a local polarity index or a *charge separation*

index which – as suggested by Collantes and Dunn III (*Collantes and Dunn, 1995*) – is defined as the sum of the magnitudes of the atomic charges. Mathematically, the CSI is given by:

$$CSI = \sum_i |q_i| \tag{5.1}$$

where q_i is the charge on atom i and the sum runs over all the atoms in the functional group or in the residue under consideration (here, the sum runs over all the atoms in the side chains).

An added advantage of the CSI over the dipole moments is their origin independence on whether the side-chain carries a net charge or not. Thus, the CSI indices were calculated for all the amino acids in their predominant ionization state. Table 5.6 lists the CSI's of the main-chain groups, of the side-chains, and of the whole amino acid molecules. The table is sorted in order of increasing CSI of the side-chains. This sorting clearly shows an even more striking pattern than the one based on the dipoles in table 5.5. All amino acids with $0.22 < \text{CSI} \leq 2.71$ au have a pyrimidine base as their second codon-letter, with only one exception: cysteine. This group includes the amino acids with mainly hydrophobic and non-polar side-chains. Neither glycine nor cystine are considered to be part of this classification - and thus do not represent exceptions - since the former lacks a side-chain while the latter is not directly encoded by DNA (only cysteine is encoded which later dimerizes - after protein synthesis - to form cystine). The second group, possessing side-chains with $2.71 \leq \text{CSI} \leq 9.68$ au, consists mainly of the hydrophilic / polar amino acids. All the members of this group possess a purine base as their second codon-letter with the exception of threonine. Serine falls on the border-line between the two groups and is the only amino acid which is coded by two synonomous triplet codons, one having a purine (G) and one having a pyrimidine (C) at the second position. In contrast, every other amino acid is encoded by codons having only one possibility per amino acid in the second position, synonyms being formed by changing the first and/or the last codon-letter but not the middle one.

These findings add significant weight to the hypothesis that the second codon-letter of a DNA triplet is important in the determination of the physical properties of the encoded amino acids.

Table 5.6 Charge separation index CSI of the amino acids (AA), of the main-chain group (MCG), and of the side-chains (R) sorted by side-chain CSI, (values in au)

	AA	MCG	R	mRNA 2nd letter	Chemical nature
Gly	7.6259	7.6168	0.0091	G	Purine
Ala	7.7757	7.5544	0.2213	C	Pyrimidine
Met	7.8043	7.5293	0.2750	U	Pyrimidine
Phe	7.8684	7.5480	0.3205	U	Pyrimidine
Val	8.2807	7.5442	0.7365	U	Pyrimidine
Cys	8.3142	7.5734	0.7408	G	Purine
Ile	8.5306	7.5147	1.0159	U	Pyrimidine
Leu	8.5625	7.5351	1.0274	U	Pyrimidine
Pro	8.3177	7.2551	1.0626	C	Pyrimidine
Ser	10.2346	7.5277	2.7068	C	Pyrimidine
Ser	10.2346	7.5277	2.7068	G	Purine
Thr	10.4002	7.5948	2.8054	C	Pyrimidine
Tyr	10.3741	7.5481	2.8260	A	Purine
Trp	10.7931	7.5547	3.2384	G	Purine
$Lys^{(+)}$	12.1032	7.9284	4.1747	A	Purine
$Asp^{(-)}$	12.6358	7.5525	5.0834	A	Purine
$Glu^{(-)}$	12.8323	7.4824	5.3498	A	Purine
Asn(II)	13.0948	7.6077	5.4871	A	Purine
Asn(I)	13.2189	7.5900	5.6289	A	Purine
Gln	13.2066	7.5415	5.6651	A	Purine
His(II)	13.6557	7.5640	6.0918	A	Purine
His(I)	13.6781	7.5537	6.1244	A	Purine
$His^{(+)}$	14.7232	7.6560	7.0671	A	Purine
$Arg^{(+)}$	17.2692	7.5902	9.6790	G	Purine

5.4.6 Correlations Between Atomic and Group Properties and Experimental Physicochemical Properties

5.4.6.1 Partial Molar Volumes of the Amino Acid Side Chains

The partial molar volume of a molecular species is defined as the volume increment resulting from the dissolution of one mole of this species into a specific solvent at the limit of infinite dilution. Group additivity of partial molar volume at infinite dilution v^0 is well documented. (*Hinz, 1986; Millero et al., 1978*) Group contributions to v^0 are the result of the sum of two contributions: (*Millero et al., 1978*)

$$v^0 = v^0(\text{int}) + v^0(\text{elect}) \quad (5.2)$$

where v^0(int) and v^0(elect) denote the intrinsic partial molar volume of the group (due to the space occupied by the group) and the electrostriction partial molar volume (which is due to the electrostatic pull the group exerts on the surrounding water of hydration (*Desnoyers et al., 1965*)). The v^0(int) in turn is due to the physical space occupied by the electron density of the group in addition to any solvent-inaccessible volume caused by packing effects.

In the search for statistical correlations between theoretically computed atomic properties and the experimental partial molar volumes at infinite dilution of amino acids and of functional groups, one has to select properties that represent the two major contributions to v^0, namely, the intrinsic and the electrostriction volumes. The intrinsic volume was taken to equal the van der Waals volume v_{vdw}, the integrated atomic volume up to the 0.001 au envelope. (*Bader, 1990*) No corrections for dead volume were made, which can partly account for the small discrepancies between the experimental and calculated v^0. The charge separation index (CSI) defined in Eq. 5.1 is taken as the term corresponding to the electrostriction term in the regression equation.

A regression equation was obtained between the experimental aqueous v^0 of the whole free amino acids, the side-chain van der Waals' volume v_{vdw}, and the CSI of the side chains:

$$v^0_{(\text{amino acid})} = 37.250 + 0.098\ v_{vdw(\text{side chain})} - 0.884\ \text{CSI}_{(\text{side chain})} \qquad [r^2 = 0.978,\ s = 3.887,\ \text{n}=20] \quad (5.3)$$

Experimental versus calculated values using this equation are given in table 5.7 and plotted in figure 5.1, these show that the linear regression of v^0 on the total integrated group volume up to the 0.001 au isodensity envelope and simultaneously on the CSI accounts for 97.8% of the variance in v^0. It is important to note that in Eq. 5.3 the CSI is found to have negative correlation to v^0 as expected from the increased electrostriction of water molecules with increasing the polarity of the solute. Eq. 5.3 is term-to-term equivalent to the empirical equation (Eq. 5.2), where the positive contribution of the intrinsic volume v_{vdw} is added to a negative contribution due to electrostriction. (*Millero et al., 1978*)

The correlation holds not only for the side-chains taken in their entirety, but also for groups that make them up. Thus, the experimentally-derived functional group contributions to v^0 can also be fitted to a similar regression equation:

$$v^0 = -0.925 + 0.127\, v_{vdw} - 2.456\ \mathrm{CSI}, \quad [r^2 = 0.983, s = 1.122, n=8] \tag{5.4}$$

which indicates an almost perfect linear correlation accounting for 98.3% of the variance in v^0. A linear regression to the v_{vdw} alone (without the CSI term) is characterized by an r^2 of only 0.806, which shows the marked improvement in the fit when the CSI/electrostriction term is included. Experimental and calculated (Eq. 5.4) functional group contributions to v^0 are compared in table 5.8. The calculated CSI and v_{vdw} are averages taken over all amino acids possessing the tabulated functional group.

Thus the partial molar volumes constitute another example of a transferable and additive property that can be expressed in terms of functional group contributions both experimentally as well as theoretically.

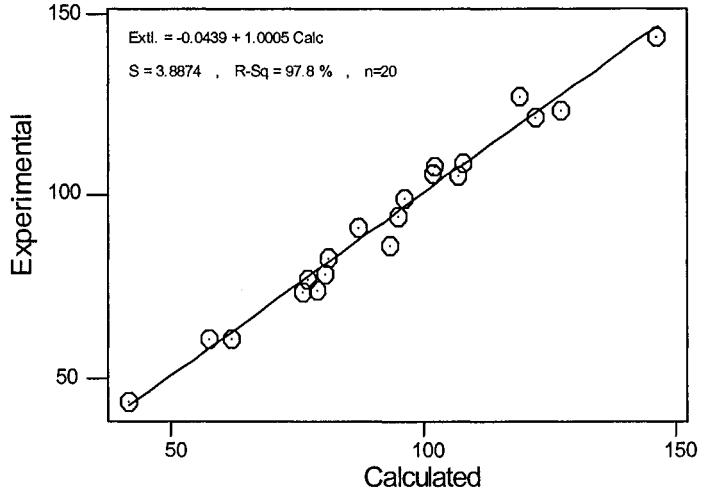


Figure 5.1 A scatter plot of the calculated (Eq.5.4) versus experimental partial molar volumes of the genetically encoded amino acids. The linear regression fitted line is superimposed on the data points.

Table 5.7 Experimental and calculated aqueous partial molar volumes at infinite dilutions (v^0) of the free amino acids (AA)*

	Expt. $v^0_{(amino\ acid)}$	Calc. $v^0_{(amino\ acid)}$	$V_{vdv(R)}$	$CSI_{(R)}$
Gly	43.3	41.9	47.25	0.009
Ala	60.5	57.8	212.22	0.221
Ser	60.6	62.0	277.93	2.707
Cys	73.4	76.3	406.30	0.741
$Asp^{(-)}$	73.8	79.0	473.10	5.083
Thr	76.9	77.3	435.30	2.805
Asn(II)	78.0	80.5	493.10	5.487
Pro	82.8	81.4	461.70	1.070
$Glu^{(-)}$	85.9	93.6	625.23	5.350
Val	90.8	87.3	519.70	0.737
Gln	93.9	95.2	644.87	5.665
His	98.8	96.2	666.83	6.982
Met	105.4	106.9	715.79	0.275
Ile	105.8	101.9	670.83	1.010
Leu	107.8	102.2	674.48	1.019
$Lys^{(+)}$	108.5	107.8	760.30	4.175
Phe	121.5	122.5	879.93	0.698
Tyr	123.6	127.3	947.65	2.826
$Arg^{(+)}$	127.3	119.1	926.27	9.679
Trp	143.9	146.3	1146.63	3.238

* Experimental values are determined at 25°C in water and were taken from (*Hinz, 1986*). Partial molar volumes are in $cm^3\ mol^{-1}$, Van der Waals volume and CSI are in au.

Table 5.8 Experimental and calculated group contributions to the partial molar volume at infinite dilution*

	Expt.	Calc.	V_{vdw}	CSI
$-CH_3$	26.5	25.5	213.79	0.277
$-CH_2OH$	28.2	27.7	277.93	2.707
$-CONH_2$	28.8	29.6	344.49	5.363
-COOH	25.8	25.5	306.26	5.077
$-NH_2$	15.4	16.4	175.63	2.013
-C=O	13.1	12.7	167.78	3.119
-NH-	11.6	10.7	126.83	1.826
$-CH_2-$	15.9	17.4	150.49	0.309

* Experimental values are determined at 25°C in water and were taken from (*Hinz, 1986*). Partial molar volumes are in $cm^3\ mol^{-1}$, Van der Waals volume and CSI are in au.

5.4.6.2 The Hydrophobicity / Hydrophilicity of the Amino Acid Side-Chains

An important property of side-chains is their relative water affinities known as "hydrophobicity / hydrophilicity", since proteins exist universally in aqueous media in the biological environment (with few exceptions such as those parts of a protein spanning the hydrophobic interior of a biological membrane (*Engelman and Steitz,Wetlaufer, 1984; White and Wimley, 1999*)). Hydrophobicity / hydrophilicity is crucial for protein folding (*Creighton, 1983*) since hydrophobic side-chains tend to cluster in the interior of the protein while hydrophilic residues tend to maximize their exposure to the aqueous phase on the surface of the folded protein (See for example Ref. (*Volkenstein, 1977*)).

Hydrophobicity, which is mainly an entropy-driven phenomenon, (*Némethy, 1967*) is related to the polarity of the side chains. Polar side-chains tend to disrupt the local water structure as opposed to non-polar side-chains which preserve the aqueous hydrogen-bonded network. (*Head-Gordon et al., 1997; Ide et al., 1997*) The charge separation index, being a conformationally-independent measure of the local polarity of a side-chain, is therefore a candidate for correlation with experimental estimates of hydrophobicity / hydrophilicity.

The free energy of hydration is defined to be the energy required to transfer the molecule from the gas phase to the aqueous phase.[1] In an attempt to classify the genetically-encoded amino acids based on the water affinities of the side chains, Wolfenden *et al.* measured $\Delta G_{hydr.}$ of a series of model molecules representing the side-chains capped with a hydrogen (instead of the main-chain group). (*Wolfenden et al., 1981*) The reason for using model side-chain molecules instead of the intact amino acids is the low volatility of the latter due to zwitter-ionization in solution resulting in very low vapor pressures which cannot be measured accurately at the time of writing. An additional reason for choosing the model molecules is the better representation of the *solvation of the side-chain in a*

---

[1] Free energy of transfer $\equiv \Delta G_{hydr.} = \Delta G^0 + RT \ln K_d$, where K_d is the partition coefficient (the ratio between the concentration in the gas phase to the concentration in the aqueous phase.

peptide, since the peptide group solvation energy is significantly lower than that of the zwitter-ionic main-chain group.

The experimental free energies of hydration of the model side-chains were also found to be highly correlated with the CSI. Figure 5.2 displays the scatter plot as well as the fitted line corresponding to the regression equation:

$$\Delta G_{hydr.} = 1.8319 - 2.2366\ \mathrm{CSI}_{(\text{side chain})} \qquad [r^2 = 0.936,\ s = 1.5985,\ \mathrm{n}=19]\ . \qquad (5.5)$$

Table 5.9 shows a linear correlation between the experimental and the calculated values. The possible sources of the small fluctuations around the regression line can be summarized as follows: (i) errors in the experimental data themselves; (ii) discrepancy between the experimental model of the side-chains which are capped with a hydrogen atom and the theoretical calculations which are performed on the intact amino acid molecules.

The previous examples of correlations between experimental and calculated quantities lays the ground for more sophisticated quantitative structure-to-activity relationships (QSARs), which can serve as a foundation for drug and materials design. (*Bader et al.,Alber and Carloni, 2002*)

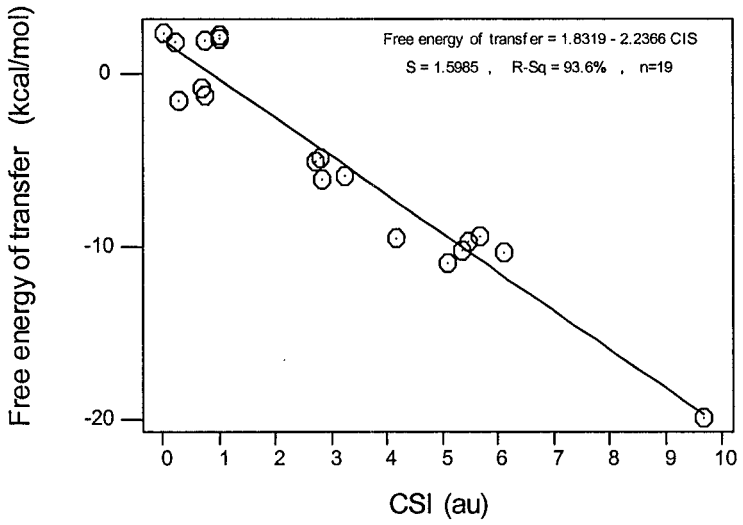


Figure 5.2 A scatter plot of the free energy of transfer from the gas phase to the aqueous phase ($\Delta G_{hydr.}$) versus the side-chain CSI. The linear regression fitted line is superimposed on the data points.

Table 5.9 Experimental and calculated free energy of transfer ($\Delta G_{hydr.}$, kcal/mol) of the amino acid side-chains from the gas to the aqueous phase (CSI values for the side-chains used in the fitting are also listed (in au))*

	$\Delta G_{hydr.}$ (Expt.)	$\Delta G_{hydr.}$ (Calc.)	CSI (R)
$arg^{(+)}$	-19.92	-19.82	9.679
$asp^{(-)}$	-10.95	-9.54	5.083
his	-10.27	-11.87	6.124
$glu^{(-)}$	-10.2	-10.84	5.665
asn	-9.68	-10.44	5.487
$lys^{(+)}$	-9.52	-7.51	4.175
gln	-9.38	-10.13	5.35
tyr	-6.11	-4.49	2.826
trp	-5.88	-5.41	3.238
ser	-5.06	-4.22	2.707
thr	-4.88	-4.44	2.805
met	-1.48	1.22	0.275
cys	-1.24	0.17	0.741
phe	-0.76	0.27	0.698
ala	1.94	1.34	0.221
val	1.99	0.18	0.737
ile	2.15	-0.43	1.01
leu	2.28	-0.45	1.019
gly	2.39	1.81	0.009

* Experimental values taken from (*Wolfenden et al., 1981*).

5.4 Conclusions

Atomic and group properties have been shown to provide a basis for excellent predictive "quantitative structure-to-activity relationship (QSAR)" models. Such models can be used in drug design since biological properties often correlate strongly with physicochemical properties, which in turn were shown to be highly predictable from theoretically-derived properties. More advanced statistical models (e.g. multivariate regression, principal component analysis, etc.) can be utilized in such endeavours in the future to predict the pharmacological activities of members of a series of biologically-active congeners. (*Mager, 1984*)

The geometric, atomic, and bond properties of a functional group were shown to exhibit a high degree of transferability between different conformations of one and the same molecule and between different molecules. These properties depend on their immediate (local) environment and are

quite insensitive to perturbations at remote sites in the molecule. These characteristics are ideal if one is to use the set of atomic electrical multipoles calculated for the free amino acids – for example - to expand the molecular electrostatic potential (MEP) of a protein, since these will be valid as the molecule changes its conformation to within an excellent approximation. The use of the QT-AIM-derived properties to assemble large molecules from small buffered fragments rests on this transferability, as will be shown in the next chapter.

Through natural selection, evolution resulted in a robust genetic code which is directly linked to the physical properties of the amino acids it encodes. This established knowledge has been extended to suggest a direct link between this code and the side-chain polarities measured by their group contributions to the molecular dipole moment as well as by the less conformation-sensitive measure of polarity proposed here: the charge separation index.

APPENDIX: Listing of Individual Bond and Atomic Properties for the Genetically-Encoded Amino Acids Calculated from Densities Obtained at the RHF/6-311++G**//6-31+G* Level of Theory

Table A5.1 Selected bond properties of the genetically-encoded amino acids (values in au, except bonded radii which are given in Å)

AA	Bond	A-B	A	B	r_b (A)	r_b (B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
Gly	0.1	C=O	C	O1	0.394	0.788	0.443	0.171	0.061	0.825	-0.782
Gly	0.2	C-O(H)	C	O2	0.429	0.900	0.308	-0.069	0.162	0.445	-0.462
Gly	0.3	O-H acid	O2	H	0.778	0.168	0.382	-2.896	0.017	0.063	-0.787
Gly	0.4	C-C	Cα	C	0.686	0.830	0.271	-0.810	0.115	0.059	-0.262
Gly	0.5	C-N	Cα	N	0.532	0.906	0.286	-0.928	0.006	0.146	-0.378
Gly	0.7	N-H	N	H1,2	0.736	0.264	0.356	-1.814	0.052	0.059	-0.512
Gly	0.8	C-H (CH2 sat)	Cα	Hα	0.690	0.395	0.290	-1.068	0.038	0.039	-0.306

AA	Bond	A-B	A	B	r_b (A)	r_b (B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
Ala	0.1	C=O	C	O1	0.397	0.793	0.436	0.111	0.059	0.793	-0.765
Ala	0.2	C-O(H)	C	O2	0.430	0.901	0.307	-0.070	0.167	0.442	-0.460
Ala	0.3	O-H acid	O2	H	0.782	0.170	0.375	-2.826	0.017	0.061	-0.768
Ala	0.4	C-C	Cα	C	0.685	0.839	0.269	-0.799	0.099	0.059	-0.258
Ala	0.5	C-N	Cα	N	0.533	0.911	0.284	-0.918	0.009	0.146	-0.375
Ala	0.6	N-H	N	H1	0.736	0.266	0.354	-1.780	0.051	0.059	-0.504
Ala	0.7	N-H	N	H2	0.738	0.264	0.354	-1.792	0.050	0.058	-0.506
Ala	0.8	C-H (CH sat)	Cα	Hα	0.688	0.397	0.293	-1.091	0.032	0.039	-0.312
Ala	1	C-C	Cα	Cβ	0.793	0.740	0.253	-0.677	0.032	0.052	-0.221
Ala	1.1	C-H (CH3)	Cβ	Hβ1	0.683	0.399	0.287	-1.044	0.009	0.042	-0.303
Ala	1.2	C-H (CH3)	Cβ	Hβ2	0.679	0.408	0.283	-1.019	0.011	0.044	-0.299
Ala	1.3	C-H (CH3)	Cβ	Hβ3	0.681	0.402	0.286	-1.042	0.011	0.043	-0.303
Val	0.1	C=O	C	O1	0.397	0.794	0.436	0.109	0.059	0.792	-0.765
Val	0.2	C-O(H)	C	O2	0.430	0.901	0.306	-0.069	0.167	0.442	-0.459
Val	0.3	O-H acid	O2	H	0.782	0.170	0.375	-2.826	0.017	0.062	-0.768
Val	0.4	C-C	Cα	C	0.684	0.840	0.267	-0.789	0.102	0.059	-0.256
Val	0.5	C-N	Cα	N	0.533	0.913	0.282	-0.905	0.016	0.146	-0.372
Val	0.6	N-H	N	H1	0.735	0.266	0.354	-1.771	0.052	0.060	-0.502
Val	0.7	N-H	N	H2	0.738	0.263	0.355	-1.810	0.049	0.058	-0.511
Val	0.8	C-H (CH sat)	Cα	Hα	0.686	0.399	0.292	-1.082	0.032	0.040	-0.310
Val	1	C-C	Cα	Cβ	0.809	0.740	0.248	-0.650	0.022	0.051	-0.213
Val	21	C-C	Cβ	Cγ1	0.774	0.757	0.253	-0.677	0.012	0.052	-0.221
Val	22	C-C	Cβ	Cγ2	0.767	0.766	0.252	-0.668	0.009	0.051	-0.218
Val	1.1	C-H (CH sat)	Cβ	Hβ	0.678	0.407	0.290	-1.062	0.005	0.044	-0.309
Val	21.1	C-H (CH3)	Cγ1	Hγ1-1	0.675	0.411	0.282	-1.015	0.010	0.045	-0.299
Val	21.2	C-H (CH3)	Cγ1	Hγ1-2	0.683	0.401	0.286	-1.041	0.008	0.042	-0.303
Val	21.3	C-H (CH3)	Cγ1	Hγ1-3	0.676	0.410	0.282	-1.016	0.010	0.045	-0.299
Val	22.1	C-H (CH3)	Cγ2	Hγ2-1	0.677	0.410	0.283	-1.018	0.009	0.045	-0.299
Val	22.2	C-H (CH3)	Cγ2	Hγ2-2	0.677	0.408	0.283	-1.018	0.007	0.045	-0.299
Val	22.3	C-H (CH3)	Cγ2	Hγ2-3	0.678	0.407	0.284	-1.022	0.009	0.044	-0.300
Leu	0.1	C=O	C	O1	0.397	0.794	0.434	0.139	0.058	0.793	-0.758
Leu	0.2	C-O(H)	C	O2	0.430	0.900	0.307	-0.100	0.163	0.437	-0.463
Leu	0.3	O-H acid	O2	H	0.782	0.170	0.375	-2.824	0.017	0.062	-0.768
Leu	0.4	C-C	Cα	C	0.686	0.838	0.268	-0.790	0.102	0.059	-0.256
Leu	0.5	C-N	Cα	N	0.536	0.908	0.283	-0.914	0.012	0.142	-0.371
Leu	0.6	N-H	N	H1	0.735	0.266	0.354	-1.771	0.052	0.059	-0.502
Leu	0.7	N-H	N	H2	0.738	0.264	0.354	-1.788	0.050	0.058	-0.505
Leu	0.8	C-H (CH sat)	Cα	Hα	0.686	0.397	0.294	-1.094	0.033	0.040	-0.313
Leu	1	C-C	Cα	Cβ	0.801	0.739	0.250	-0.662	0.030	0.052	-0.217
Leu	2	C-C	Cβ	Cγ	0.788	0.748	0.252	-0.665	0.015	0.051	-0.218
Leu	31	C-C	Cγ	Cδ1	0.760	0.772	0.253	-0.670	0.010	0.051	-0.219
Leu	32	C-C	Cγ	Cδ2	0.760	0.772	0.253	-0.672	0.006	0.051	-0.219
Leu	1.1	C-H (CH2 sat)	Cβ	Hβ1	0.677	0.412	0.284	-1.022	0.009	0.045	-0.301
Leu	1.2	C-H (CH2 sat)	Cβ	Hβ2	0.679	0.407	0.286	-1.038	0.009	0.044	-0.304
Leu	2.1	C-H (CH sat)	Cγ	Hγ	0.681	0.404	0.291	-1.068	0.001	0.043	-0.310
Leu	31.1	C-H (CH3)	Cδ1	Hδ1-1	0.676	0.411	0.282	-1.016	0.007	0.045	-0.299
Leu	31.2	C-H (CH3)	Cδ1	Hδ1-2	0.676	0.412	0.281	-1.009	0.007	0.045	-0.298
Leu	31.3	C-H (CH3)	Cδ2	Hδ1-3	0.676	0.410	0.283	-1.016	0.007	0.045	-0.299
Leu	32.1	C-H (CH3)	Cδ2	Hδ2-1	0.676	0.412	0.281	-1.005	0.008	0.045	-0.297
Leu	32.2	C-H (CH3)	Cδ2	Hδ2-2	0.675	0.410	0.284	-1.024	0.008	0.045	-0.301
Leu	32.3	C-H (CH3)	Cδ1	Hδ1-3	0.676	0.410	0.283	-1.018	0.007	0.045	-0.299

AA	Bond	A-B	A	B	r_b (A)	r_b (B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
Ile	0.1	C=O	C	O1	0.397	0.794	0.435	0.102	0.059	0.790	-0.764
Ile	0.2	C-O(H)	C	O2	0.429	0.900	0.307	-0.068	0.161	0.444	-0.461
Ile	0.3	O-H acid	O2	H	0.782	0.171	0.375	-2.825	0.017	0.062	-0.768
Ile	0.4	C-C	Cα	C	0.684	0.841	0.267	-0.787	0.088	0.059	-0.256
Ile	0.5	C-N	Cα	N	0.537	0.913	0.280	-0.896	0.021	0.141	-0.365
Ile	0.6	N-H	N	H1	0.735	0.267	0.354	-1.759	0.053	0.060	-0.500
Ile	0.7	N-H	N	H2	0.739	0.262	0.355	-1.823	0.048	0.058	-0.513
Ile	0.8	C-H (CH sat)	Cα	Hα	0.686	0.398	0.295	-1.097	0.031	0.040	-0.314
Ile	1	C-C	Cα	Cβ	0.811	0.741	0.246	-0.638	0.031	0.050	-0.210
Ile	21	C-C	Cβ	Cγ1	0.777	0.762	0.252	-0.665	0.016	0.050	-0.217
Ile	22	C-C	Cβ	Cγ2	0.769	0.764	0.251	-0.665	0.005	0.051	-0.217
Ile	31	C-C	Cγ1	Cδ1	0.764	0.767	0.251	-0.667	0.008	0.051	-0.218
Ile	1.1	C-H (CH sat)	Cβ	Hβ	0.682	0.404	0.289	-1.057	0.004	0.043	-0.307
Ile	21.1	C-H (CH2 sat)	Cγ1	Hγ1-1	0.675	0.412	0.285	-1.033	0.007	0.045	-0.304
Ile	21.2	C-H (CH2 sat)	Cγ1	Hγ1-2	0.676	0.411	0.286	-1.035	0.006	0.045	-0.304
Ile	22.1	C-H (CH3)	Cγ2	Hγ2-1	0.677	0.407	0.284	-1.027	0.010	0.045	-0.302
Ile	22.2	C-H (CH3)	Cγ2	Hγ2-2	0.677	0.409	0.283	-1.019	0.009	0.045	-0.299
Ile	22.3	C-H (CH3)	Cγ2	Hγ2-3	0.676	0.408	0.284	-1.026	0.008	0.045	-0.302
Ile	31.1	C-H (CH3)	Cδ1	Hδ1-1	0.676	0.410	0.282	-1.015	0.009	0.045	-0.299
Ile	31.2	C-H (CH3)	Cδ1	Hδ1-2	0.675	0.410	0.284	-1.026	0.009	0.045	-0.302
Ile	31.3	C-H (CH3)	Cδ1	Hδ1-3	0.677	0.409	0.283	-1.019	0.009	0.045	-0.300
Ile	h	O...H-C	O2	Hγ1-2	1.520	1.219	0.008	0.030	0.260	0.006	0.001
Ser	0.1	C=O	C	O1	0.396	0.792	0.438	0.126	0.065	0.802	-0.771
Ser	0.2	C-O(H)	C	O2	0.431	0.905	0.302	-0.063	0.176	0.433	-0.449
Ser	0.3	O-H acid	O2	H	0.784	0.170	0.374	-2.817	0.017	0.061	-0.765
Ser	0.4	C-C	Cα	C	0.682	0.843	0.267	-0.785	0.100	0.060	-0.256
Ser	0.5	C-N	Cα	N	0.537	0.907	0.284	-0.920	0.006	0.142	-0.372
Ser	0.6	N-H	N	H1	0.738	0.263	0.355	-1.807	0.050	0.058	-0.510
Ser	0.7	N-H	N	H2	0.737	0.265	0.354	-1.781	0.050	0.059	-0.504
Ser	0.8	C-H (CH sat)	Cα	Hα	0.692	0.393	0.292	-1.081	0.031	0.038	-0.309
Ser	1	C-C	Cα	Cβ	0.793	0.747	0.257	-0.704	0.074	0.049	-0.225
Ser	2	C-O(H)	Cβ	Oγ	0.449	0.947	0.265	-0.108	0.063	0.349	-0.376
Ser	1.1	C-H (CH2 sat)	Cβ	Hβ1	0.693	0.386	0.301	-1.149	0.038	0.035	-0.322
Ser	1.2	C-H (CH2 sat)	Cβ	Hβ2	0.685	0.405	0.291	-1.076	0.044	0.039	-0.308
Ser	2.1	O-H alcohol	Oγ	Hγ	0.775	0.172	0.384	-2.860	0.023	0.068	-0.783
Ser	h	O...H-O	O2	Hγ	1.330	0.856	0.016	0.067	0.022	0.015	0.002
Thr	0.1	C=O	C	O1	0.396	0.792	0.438	0.122	0.063	0.801	-0.770
Thr	0.2	C-O(H)	C	O2	0.430	0.901	0.306	-0.074	0.164	0.440	-0.459
Thr	0.3	O-H acid	O2	H	0.782	0.170	0.375	-2.827	0.017	0.062	-0.768
Thr	0.4	C-C	Cα	C	0.694	0.833	0.267	-0.782	0.100	0.056	-0.252
Thr	0.5	C-N	Cα	N	0.541	0.900	0.287	-0.939	0.013	0.138	-0.373
Thr	0.6	N-H	N	H1	0.737	0.264	0.355	-1.800	0.049	0.059	-0.508
Thr	0.7	N-H	N	H2	0.741	0.261	0.354	-1.832	0.048	0.057	-0.515
Thr	0.8	C-H (CH sat)	Cα	Hα	0.687	0.400	0.290	-1.072	0.032	0.040	-0.308
Thr	1	C-C	Cα	Cβ	0.777	0.756	0.262	-0.731	0.019	0.049	-0.231
Thr	21	C-O(H)	Cβ	Oγ1	0.451	0.961	0.252	-0.049	0.042	0.339	-0.351
Thr	22	C-C	Cβ	Cγ2	0.782	0.740	0.262	-0.731	0.038	0.052	-0.235
Thr	1.1	C-H (CH sat)	Cβ	Hβ	0.687	0.397	0.299	-1.129	0.038	0.038	-0.320
Thr	21.1	O-H alcohol	Oγ1	Hγ1	0.770	0.177	0.384	-2.819	0.024	0.071	-0.775
Thr	22.1	C-H (CH3)	Cγ2	Hγ2-1	0.686	0.395	0.288	-1.055	0.006	0.041	-0.305
Thr	22.2	C-H (CH3)	Cγ2	Hγ2-2	0.675	0.413	0.281	-1.006	0.007	0.046	-0.298
Thr	22.3	C-H (CH3)	Cγ2	Hγ2-3	0.678	0.408	0.282	-1.016	0.007	0.044	-0.299

AA	Bond	A-B	A	B	r_b (A)	r_b (B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
Met	0.1	C=O	C	O1	0.397	0.793	0.436	0.107	0.065	0.794	-0.767
Met	0.2	C-O(H)	C	O2	0.430	0.902	0.306	-0.069	0.166	0.440	-0.458
Met	0.3	O-H acid	O2	H	0.783	0.170	0.375	-2.827	0.017	0.061	-0.768
Met	0.4	C-C	Cα	C	0.685	0.839	0.269	-0.802	0.028	0.058	-0.258
Met	0.5	C-N	Cα	N	0.541	0.915	0.274	-0.855	0.010	0.139	-0.352
Met	0.6	N-H	N	H1	0.738	0.264	0.354	-1.796	0.054	0.058	-0.507
Met	0.7	N-H	N	H2	0.738	0.265	0.353	-1.782	0.054	0.058	-0.504
Met	0.8	C-H (CH sat)	Cα	Hα	0.687	0.396	0.295	-1.104	0.017	0.039	-0.314
Met	1	C-C	Cα	Cβ	0.794	0.734	0.258	-0.707	0.033	0.053	-0.229
Met	2	C-C	Cβ	Cγ	0.748	0.780	0.254	-0.677	0.015	0.052	-0.221
Met	3	C-S (C)	Cγ	Sδ	0.860	0.957	0.189	-0.371	0.112	0.044	-0.136
Met	4	C-S (C)	Cϵ	Sδ	0.860	0.948	0.191	-0.383	0.102	0.044	-0.140
Met	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.685	0.399	0.289	-1.058	0.010	0.042	-0.306
Met	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.684	0.400	0.288	-1.052	0.007	0.042	-0.305
Met	2.1	C-H (CH2 sat)	Cγ	Hγ-1	0.682	0.403	0.288	-1.046	0.007	0.042	-0.304
Met	2.2	C-H (CH2 sat)	Cγ	Hγ-2	0.687	0.396	0.291	-1.068	0.005	0.041	-0.308
Met	4.1	C-H (CH3)	Cϵ	Hϵ-1	0.685	0.399	0.287	-1.044	0.014	0.041	-0.302
Met	4.2	C-H (CH3)	Cϵ	Hϵ-2	0.687	0.396	0.287	-1.044	0.009	0.041	-0.302
Met	4.3	C-H (CH3)	Cϵ	Hϵ-3	0.685	0.398	0.287	-1.046	0.014	0.041	-0.302
											0.000
Cys	0.1	C=O	C	O1	0.397	0.794	0.436	0.100	0.062	0.790	-0.765
Cys	0.2	C-O(H)	C	O2	0.429	0.899	0.308	-0.071	0.160	0.446	-0.463
Cys	0.3	O-H acid	O2	H	0.783	0.170	0.375	-2.827	0.017	0.061	-0.768
Cys	0.4	C-C	Cα	C	0.688	0.835	0.270	-0.809	0.034	0.057	-0.259
Cys	0.5	C-N	Cα	N	0.539	0.916	0.274	-0.856	0.013	0.140	-0.354
Cys	0.6	N-H	N	H1	0.738	0.263	0.354	-1.805	0.054	0.058	-0.509
Cys	0.7	N-H	N	H2	0.737	0.265	0.353	-1.779	0.054	0.059	-0.503
Cys	0.8	C-H (CH sat)	Cα	Hα	0.691	0.391	0.297	-1.118	0.015	0.037	-0.316
Cys	1	C-C	Cα	Cβ	0.776	0.749	0.261	-0.722	0.040	0.051	-0.232
Cys	2	C-S (H)	Cβ	Sγ	0.857	0.963	0.187	-0.361	0.086	0.043	-0.133
Cys	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.691	0.390	0.292	-1.077	0.014	0.039	-0.309
Cys	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.690	0.391	0.292	-1.072	0.013	0.039	-0.308
Cys	2.1	S-H	Sγ	Hγ	0.651	0.675	0.226	-0.393	0.388	0.197	-0.295
CysS	0.1	C=O	C	O1	0.397	0.793	0.437	0.106	0.062	0.794	-0.768
CysS	0.2	C-O(H)	C	O2	0.428	0.896	0.311	-0.068	0.145	0.453	-0.470
CysS	0.3	O-H acid	O2	H	0.784	0.170	0.374	-2.825	0.017	0.061	-0.767
CysS	0.4	C-C	Cα	C	0.698	0.829	0.269	-0.797	0.056	0.055	-0.254
CysS	0.5	C-N	Cα	N	0.544	0.905	0.281	-0.901	0.019	0.135	-0.360
CysS	0.6	N-H	N	H1	0.737	0.265	0.354	-1.777	0.054	0.059	-0.503
CysS	0.7	N-H	N	H2	0.741	0.262	0.354	-1.821	0.050	0.057	-0.512
CysS	0.8	C-H (CH sat)	Cα	Hα	0.690	0.392	0.297	-1.115	0.022	0.037	-0.316
CysS	1	C-C	Cα	Cβ	0.764	0.771	0.253	-0.676	0.036	0.051	-0.220
CysS	2	C-S (S)	Cβ	Sγ	0.852	0.976	0.186	-0.353	0.108	0.043	-0.131
CysS	3	S-S	Sγi	Sγk	1.026	1.026	0.152	-0.176	0.007	0.043	-0.087
CysS	2.1	C-H (CH2 sat)	Cβ	Hβ-1	0.690	0.389	0.294	-1.090	0.012	0.039	-0.311
CysS	2.2	C-H (CH2 sat)	Cβ	Hβ-2	0.692	0.390	0.293	-1.079	0.008	0.039	-0.308
Asn(I)	0.1	C=O	C	O1	0.396	0.793	0.437	0.118	0.063	0.797	-0.768
Asn(I)	0.2	C-O(H)	C	O2	0.429	0.897	0.310	-0.079	0.149	0.448	-0.468
Asn(I)	0.3	O-H acid	O2	H	0.783	0.170	0.374	-2.825	0.017	0.061	-0.767
Asn(I)	0.4	C-C	Cα	C	0.696	0.828	0.269	-0.797	0.095	0.056	-0.255
Asn(I)	0.5	C-N	Cα	N	0.534	0.916	0.280	-0.894	0.013	0.145	-0.368

AA	Bond	A-B	A	B	r_b (A)	r_b (B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
Asn(I)	0.6	N-H	N	H1	0.738	0.264	0.354	-1.790	0.049	0.058	-0.505
Asn(I)	0.7	N-H	N	H2	0.741	0.262	0.353	-1.816	0.046	0.057	-0.510
Asn(I)	0.8	C-H (CH sat)	Cα	Hα	0.691	0.393	0.294	-1.095	0.030	0.038	-0.312
Asn(I)	1	C-C	Cα	Cβ	0.791	0.749	0.250	-0.660	0.035	0.051	-0.216
Asn(I)	2	C-C	Cβ	Cγ	0.737	0.789	0.265	-0.754	0.053	0.050	-0.239
Asn(I)	31	C=O	Cγ	Oδ1	0.401	0.802	0.426	-0.031	0.048	0.735	-0.742
Asn(I)	32	C-N	Cγ	Nδ2	0.456	0.890	0.333	-0.843	0.016	0.309	-0.519
Asn(I)	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.686	0.402	0.283	-1.016	0.010	0.043	-0.297
Asn(I)	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.691	0.390	0.293	-1.080	0.017	0.039	-0.309
Asn(I)	32.1	N-H	Nδ2	Hδ2-1	0.760	0.237	0.352	-2.085	0.054	0.048	-0.569
Asn(I)	32.2	N-H	Nδ2	Hδ2-2	0.746	0.249	0.353	-1.954	0.056	0.050	-0.539
Asn(I)	h	N...H-N	N	Hδ2-1	1.395	0.790	0.019	0.072	0.053	0.016	0.002
Asn(II)	0.1	C=O	C	O1	0.398	0.796	0.433	0.078	0.055	0.778	-0.758
Asn(II)	0.2	C-O(H)	C	O2	0.427	0.892	0.316	-0.083	0.139	0.460	-0.480
Asn(II)	0.3	O-H acid	O2	H	0.783	0.170	0.375	-2.833	0.017	0.061	-0.769
Asn(II)	0.4	C-C	Cα	C	0.695	0.827	0.271	-0.817	0.036	0.054	-0.259
Asn(II)	0.5	C-N	Cα	N	0.542	0.912	0.277	-0.873	0.015	0.137	-0.355
Asn(II)	0.6	N-H	N	H1	0.737	0.263	0.354	-1.796	0.056	0.058	-0.507
Asn(II)	0.7	N-H	N	H2	0.737	0.265	0.353	-1.781	0.054	0.059	-0.504
Asn(II)	0.8	C-H (CH sat)	Cα	Hα	0.694	0.387	0.299	-1.126	0.012	0.036	-0.317
Asn(II)	1	C-C	Cα	Cβ	0.771	0.765	0.253	-0.672	0.028	0.051	-0.219
Asn(II)	2	C-C	Cβ	Cγ	0.723	0.797	0.267	-0.771	0.050	0.052	-0.245
Asn(II)	31	C=O	Cγ	Oδ1	0.401	0.801	0.426	-0.012	0.046	0.739	-0.742
Asn(II)	32	C-N	Cγ	Nδ2	0.459	0.894	0.329	-0.854	0.007	0.299	-0.512
Asn(II)	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.685	0.398	0.288	-1.050	0.008	0.042	-0.304
Asn(II)	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.682	0.401	0.289	-1.052	0.007	0.043	-0.306
Asn(II)	32.1	N-H	Nδ2	Hδ2-1	0.749	0.247	0.354	-1.988	0.056	0.050	-0.548
Asn(II)	32.2	N-H	Nδ2	Hδ2-2	0.747	0.250	0.353	-1.954	0.054	0.050	-0.539
Asn(II)	h	O...H-N	O1	Hδ2-1	1.541	1.117	0.007	0.027	0.330	0.006	0.001
Gln	0.1	C=O	C	O1	0.397	0.794	0.435	0.095	0.065	0.788	-0.764
Gln	0.2	C-O(H)	C	O2	0.430	0.901	0.306	-0.060	0.169	0.444	-0.459
Gln	0.3	O-H acid	O2	H	0.783	0.170	0.375	-2.827	0.017	0.061	-0.768
Gln	0.4	C-C	Cα	C	0.682	0.841	0.269	-0.800	0.074	0.059	-0.259
Gln	0.5	C-N	Cα	N	0.535	0.914	0.279	-0.887	0.010	0.144	-0.366
Gln	0.6	N-H	N	H1	0.735	0.266	0.354	-1.770	0.055	0.059	-0.502
Gln	0.7	N-H	N	H2	0.738	0.263	0.354	-1.807	0.052	0.058	-0.510
Gln	0.8	C-H (CH sat)	Cα	Hα	0.688	0.395	0.296	-1.109	0.030	0.038	-0.315
Gln	1	C-C	Cα	Cβ	0.797	0.739	0.252	-0.673	0.039	0.052	-0.220
Gln	2	C-C	Cβ	Cγ	0.747	0.781	0.254	-0.679	0.020	0.052	-0.222
Gln	3	C-C	Cγ	Cδ	0.718	0.801	0.268	-0.773	0.061	0.054	-0.247
Gln	41	C=O	Cδ	Oϵ1	0.401	0.800	0.427	-0.007	0.045	0.743	-0.745
Gln	42	C-N	Cδ	Nϵ2	0.457	0.898	0.325	-0.791	0.012	0.305	-0.503
Gln	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.686	0.397	0.290	-1.063	0.011	0.041	-0.307
Gln	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.686	0.398	0.290	-1.060	0.010	0.041	-0.306
Gln	2.1	C-H (CH2 sat)	Cγ	Hγ-1	0.680	0.407	0.285	-1.027	0.014	0.044	-0.301
Gln	2.2	C-H (CH2 sat)	Cγ	Hγ-2	0.684	0.403	0.284	-1.021	0.006	0.044	-0.299
Gln	42.1	N-H	Nϵ2	Hϵ2-1	0.742	0.252	0.355	-1.939	0.063	0.053	-0.538
Gln	42.2	N-H	Nϵ2	Hϵ2-2	0.748	0.248	0.353	-1.964	0.058	0.050	-0.541
$Asp^{(-)}$	0.1	C=O	C	O1	0.398	0.796	0.432	0.066	0.049	0.772	-0.756
$Asp^{(-)}$	0.2	C-O(H)	C	O2	0.431	0.907	0.300	-0.061	0.195	0.431	-0.446
$Asp^{(-)}$	0.3	O-H acid	O2	H	0.778	0.173	0.380	-2.836	0.019	0.064	-0.773

AA	Bond	A-B	A	B	r_b (A)	r_b (B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
$Asp^{(-)}$	0.4	C-C	Cα	C	0.655	0.855	0.272	-0.836	0.088	0.067	-0.276
$Asp^{(-)}$	0.5	C-N	Cα	N	0.550	0.902	0.281	-0.910	0.054	0.129	-0.356
$Asp^{(-)}$	0.6	N-H	N	H1	0.751	0.252	0.353	-1.925	0.041	0.053	-0.535
$Asp^{(-)}$	0.7	N-H	N	H2	0.734	0.270	0.353	-1.728	0.048	0.061	-0.493
$Asp^{(-)}$	0.8	C-H (CH sat)	Cα	Hα	0.689	0.392	0.297	-1.115	0.024	0.038	-0.317
$Asp^{(-)}$	1	C-C	Cα	Cβ	0.803	0.736	0.251	-0.665	0.024	0.052	-0.219
$Asp^{(-)}$	2	C-C	Cβ	Cγ	0.810	0.746	0.250	-0.681	0.066	0.047	-0.217
$Asp^{(-)}$	31	C=O (-ve)	Cγ	Oδ1	0.411	0.828	0.391	-0.190	0.012	0.611	-0.659
$Asp^{(-)}$	32	C=O (-ve)	Cγ	Oδ2	0.409	0.822	0.399	-0.180	0.021	0.632	-0.677
$Asp^{(-)}$	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.677	0.411	0.281	-1.004	0.026	0.046	-0.297
$Asp^{(-)}$	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.677	0.410	0.282	-1.010	0.027	0.046	-0.298
$Asp^{(-)}$	h	O...H-N	Oδ1	H1	1.369	0.864	0.017	0.062	0.115	0.014	0.002
$Glu^{(-)}$	0.1	C=O	C	O1	0.398	0.796	0.432	0.075	0.047	0.774	-0.755
$Glu^{(-)}$	0.2	C-O(H)	C	O2	0.430	0.902	0.305	-0.058	0.181	0.442	-0.457
$Glu^{(-)}$	0.3	O-H acid	O2	H	0.780	0.172	0.378	-2.832	0.018	0.063	-0.771
$Glu^{(-)}$	0.4	C-C	Cα	C	0.668	0.853	0.268	-0.801	0.096	0.063	-0.263
$Glu^{(-)}$	0.5	C-N	Cα	N	0.526	0.928	0.275	-0.846	0.023	0.154	-0.366
$Glu^{(-)}$	0.6	N-H	N	H1	0.731	0.271	0.354	-1.719	0.053	0.062	-0.491
$Glu^{(-)}$	0.7	N-H	N	H2	0.737	0.265	0.354	-1.788	0.051	0.059	-0.506
$Glu^{(-)}$	0.8	C-H (CH sat)	Cα	Hα	0.684	0.400	0.294	-1.093	0.039	0.040	-0.313
$Glu^{(-)}$	1	C-C	Cα	Cβ	0.834	0.707	0.247	-0.656	0.029	0.056	-0.220
$Glu^{(-)}$	2	C-C	Cβ	Cγ	0.793	0.734	0.254	-0.685	0.010	0.053	-0.224
$Glu^{(-)}$	3	C-C	Cγ	Cδ	0.809	0.745	0.251	-0.687	0.067	0.047	-0.218
$Glu^{(-)}$	41	C=O (-ve)	Cδ	Oϵ1	0.410	0.825	0.394	-0.172	0.015	0.622	-0.665
$Glu^{(-)}$	42	C=O (-ve)	Cδ	Oϵ2	0.411	0.826	0.395	-0.204	0.016	0.617	-0.667
$Glu^{(-)}$	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.686	0.396	0.290	-1.066	0.015	0.041	-0.308
$Glu^{(-)}$	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.679	0.407	0.286	-1.035	0.017	0.044	-0.303
$Glu^{(-)}$	2.1	C-H (CH2 sat)	Cγ	Hγ-1	0.674	0.415	0.280	-0.999	0.027	0.047	-0.297
$Glu^{(-)}$	2.2	C-H (CH2 sat)	Cγ	Hγ-2	0.672	0.415	0.282	-1.010	0.028	0.048	-0.300
$Lys^{(+)}$	0.1	C=O	C	O1	0.389	0.802	0.436	0.393	0.088	0.850	-0.752
$Lys^{(+)}$	0.2	C-O(H)	C	O2	0.421	0.903	0.316	-0.050	0.129	0.486	-0.499
$Lys^{(+)}$	0.3	O-H acid	O2	H	0.780	0.174	0.376	-2.515	0.018	0.067	-0.696
$Lys^{(+)}$	0.4	C-C	Cα	C	0.698	0.829	0.270	-0.797	0.049	0.054	-0.253
$Lys^{(+)}$	0.5	C-N	Cα	N	0.560	0.892	0.282	-0.966	0.029	0.123	-0.365
$Lys^{(+)}$	0.6	N-H	N	H1	0.757	0.244	0.360	-1.981	0.051	0.056	-0.551
$Lys^{(+)}$	0.7	N-H	N	H2	0.752	0.251	0.358	-1.953	0.054	0.059	-0.547
$Lys^{(+)}$	0.8	C-H (CH sat)	Cα	Hα	0.672	0.414	0.297	-1.141	0.016	0.042	-0.327
$Lys^{(+)}$	1	C-C	Cα	Cβ	0.777	0.751	0.260	-0.704	0.037	0.050	-0.226
$Lys^{(+)}$	2	C-C	Cβ	Cγ	0.747	0.784	0.254	-0.669	0.017	0.051	-0.218
$Lys^{(+)}$	3	C-C	Cγ	Cδ	0.725	0.808	0.251	-0.658	0.014	0.053	-0.218
$Lys^{(+)}$	4	C-C	Cδ	Cϵ	0.695	0.827	0.259	-0.718	0.023	0.058	-0.238
$Lys^{(+)}$	5	C-N	Cϵ	Nζ	0.481	1.038	0.213	-0.193	0.002	0.251	-0.299
$Lys^{(+)}$	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.667	0.417	0.292	-1.099	0.011	0.045	-0.320
$Lys^{(+)}$	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.665	0.422	0.290	-1.080	0.007	0.046	-0.316
$Lys^{(+)}$	2.1	C-H (CH2 sat)	Cγ	Hγ-1	0.668	0.415	0.294	-1.117	0.008	0.045	-0.324
$Lys^{(+)}$	2.2	C-H (CH2 sat)	Cγ	Hγ-2	0.662	0.426	0.288	-1.062	0.011	0.047	-0.313
$Lys^{(+)}$	3.1	C-H (CH2 sat)	Cδ	Hδ-1	0.666	0.422	0.289	-1.076	0.017	0.046	-0.315
$Lys^{(+)}$	3.2	C-H (CH2 sat)	Cδ	Hδ-2	0.665	0.423	0.289	-1.072	0.017	0.046	-0.314
$Lys^{(+)}$	4.1	C-H (CH2 sat)	Cϵ	Hϵ-1	0.683	0.398	0.303	-1.215	0.045	0.036	-0.340
$Lys^{(+)}$	4.2	C-H (CH2 sat)	Cϵ	Hϵ-2	0.681	0.399	0.303	-1.210	0.046	0.037	-0.339
$Lys^{(+)}$	5.1	N-H	Nζ	Hζ-1	0.785	0.225	0.354	-2.043	0.001	0.039	-0.550
$Lys^{(+)}$	5.1	N-H	Nζ	Hζ-2	0.786	0.225	0.354	-2.042	0.001	0.039	-0.550

AA	Bond	A-B	A	B	r_b (A)	r_b (B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
$Lys^{(+)}$	5.3	N-H	Nζ	Hζ-3	0.786	0.225	0.354	-2.042	0.001	0.039	-0.550
$Lys^{(+)}$	h	O...H-C	O1	Hγ-1	1.464	1.053	0.011	0.041	0.365	0.009	0.001
$Arg^{(+)}$	0.1	C=O	C	O1	0.396	0.792	0.438	0.116	0.065	0.800	-0.771
$Arg^{(+)}$	0.2	C-O(H)	C	O2	0.429	0.899	0.308	-0.067	0.150	0.447	-0.463
$Arg^{(+)}$	0.3	O-H acid	O2	H	0.785	0.169	0.373	-2.817	0.017	0.060	-0.764
$Arg^{(+)}$	0.4	C-C	Cα	C	0.693	0.833	0.269	-0.796	0.060	0.056	-0.255
$Arg^{(+)}$	0.5	C-N	Cα	N	0.547	0.899	0.284	-0.922	0.024	0.132	-0.363
$Arg^{(+)}$	0.6	N-H	N	H1	0.739	0.263	0.354	-1.802	0.053	0.058	-0.508
$Arg^{(+)}$	0.7	N-H	N	H2	0.742	0.261	0.353	-1.827	0.050	0.056	-0.513
$Arg^{(+)}$	0.8	C-H (CH sat)	Cα	Hα	0.688	0.394	0.296	-1.110	0.023	0.038	-0.315
$Arg^{(+)}$	1	C-C	Cα	Cβ	0.774	0.762	0.253	-0.676	0.040	0.051	-0.220
$Arg^{(+)}$	2	C-C	Cβ	Cγ	0.738	0.794	0.252	-0.672	0.016	0.052	-0.220
$Arg^{(+)}$	3	C-C	Cγ	Cδ	0.707	0.818	0.256	-0.703	0.027	0.056	-0.232
$Arg^{(+)}$	4	C-N	Cδ	Nϵ	0.474	0.998	0.237	-0.249	0.011	0.262	-0.324
$Arg^{(+)}$	5	C-N	Cζ	Nϵ	0.463	0.854	0.365	-1.166	0.130	0.290	-0.581
$Arg^{(+)}$	61	C-N	Cζ	Nη1	0.460	0.865	0.355	-1.079	0.085	0.294	-0.564
$Arg^{(+)}$	62	C-N	Cζ	Nη2	0.463	0.866	0.353	-1.099	0.084	0.283	-0.558
$Arg^{(+)}$	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.685	0.399	0.289	-1.056	0.011	0.042	-0.305
$Arg^{(+)}$	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.684	0.403	0.287	-1.043	0.009	0.042	-0.303
$Arg^{(+)}$	2.1	C-H (CH2 sat)	Cγ	Hγ-1	0.682	0.405	0.286	-1.039	0.011	0.043	-0.303
$Arg^{(+)}$	2.2	C-H (CH2 sat)	Cγ	Hγ-2	0.681	0.405	0.287	-1.044	0.010	0.043	-0.304
$Arg^{(+)}$	3.1	C-H (CH2 sat)	Cδ	Hδ-1	0.690	0.394	0.294	-1.097	0.043	0.038	-0.312
$Arg^{(+)}$	3.2	C-H (CH2 sat)	Cδ	Hδ-2	0.690	0.394	0.294	-1.097	0.044	0.038	-0.312
$Arg^{(+)}$	4.1	N-H	Nϵ	Hϵ	0.752	0.246	0.354	-2.003	0.057	0.049	-0.550
$Arg^{(+)}$	61.1	N-H	Nη1	Hη1-1	0.758	0.238	0.352	-2.085	0.053	0.046	-0.567
$Arg^{(+)}$	61.2	N-H	Nη1	Hη1-2	0.756	0.240	0.351	-2.051	0.053	0.046	-0.559
$Arg^{(+)}$	62.1	N-H	Nη2	Hη2-1	0.756	0.241	0.351	-2.043	0.055	0.047	-0.557
$Arg^{(+)}$	62.2	N-H	Nη2	Hη2-2	0.756	0.241	0.351	-2.042	0.056	0.047	-0.557
His(I)	0.1	C=O	C	O1	0.398	0.796	0.432	0.079	0.052	0.776	-0.756
His(I)	0.2	C-O(H)	C	O2	0.429	0.898	0.310	-0.070	0.165	0.449	-0.467
His(I)	0.3	O-H acid	O2	H	0.783	0.170	0.375	-2.827	0.017	0.061	-0.768
His(I)	0.4	C-C	Cα	C	0.680	0.844	0.269	-0.803	0.042	0.059	-0.260
His(I)	0.5	C-N	Cα	N	0.541	0.914	0.275	-0.862	0.017	0.139	-0.354
His(I)	0.6	N-H	N	H1	0.739	0.262	0.354	-1.814	0.053	0.058	-0.511
His(I)	0.7	N-H	N	H2	0.737	0.266	0.353	-1.770	0.055	0.059	-0.501
His(I)	0.8	C-H (CH sat)	Cα	Hα	0.691	0.392	0.296	-1.108	0.016	0.037	-0.314
His(I)	1	C-C	Cα	Cβ	0.793	0.739	0.256	-0.693	0.041	0.051	-0.225
His(I)	2	C-C	Cβ	Cγ	0.717	0.782	0.265	-0.750	0.040	0.056	-0.244
His(I)	31	C-N (arom)	Cγ	Nδ1	0.450	0.924	0.300	-0.340	0.130	0.366	-0.451
His(I)	32	C-C (arom)	Cγ	Cδ2	0.716	0.641	0.341	-1.084	0.446	0.127	-0.398
His(I)	41	C-N (arom)	Cϵ1	Nδ1	0.444	0.904	0.320	-0.431	0.129	0.387	-0.494
His(I)	42	C-N (arom)	Cδ2	Nϵ2	0.470	0.901	0.323	-0.823	0.113	0.288	-0.493
His(I)	5	C-N (arom)	Cϵ1	Nϵ2	0.445	0.846	0.384	-0.940	0.273	0.404	-0.639
His(I)	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.687	0.395	0.290	-1.065	0.012	0.041	-0.307
His(I)	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.684	0.402	0.286	-1.037	0.012	0.042	-0.302
His(I)	31.1	N-H (arom)	Nδ1	Hδ1	0.759	0.237	0.354	-2.112	0.042	0.044	-0.572
His(I)	32.1	C-H (arom)	Cδ2	Hδ2	0.682	0.388	0.296	-1.119	0.047	0.035	-0.315
His(I)	41.1	C-H (arom)	Cϵ1	Hϵ1	0.691	0.380	0.300	-1.157	0.047	0.031	-0.320
His(I)	h	O...H-N	O1	Hδ1	1.381	0.901	0.013	0.054	0.127	0.012	0.002
His(II)	0.1	C=O	C	O1	0.397	0.794	0.436	0.101	0.059	0.791	-0.766
His(II)	0.2	C-O(H)	C	O2	0.429	0.897	0.310	-0.071	0.152	0.449	-0.467

AA	Bond	A-B	A	B	r_b (A)	r_b (B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
His(II)	0.3	O-H acid	O2	H	0.782	0.170	0.376	-2.831	0.017	0.061	-0.769
His(II)	0.4	C-C	Cα	C	0.694	0.831	0.269	-0.800	0.068	0.056	-0.256
His(II)	0.5	C-N	Cα	N	0.537	0.914	0.279	-0.891	0.019	0.142	-0.364
His(II)	0.6	N-H	N	H1	0.735	0.267	0.354	-1.759	0.052	0.060	-0.499
His(II)	0.7	N-H	N	H2	0.740	0.263	0.354	-1.808	0.049	0.057	-0.509
His(II)	0.8	C-H (CH sat)	Cα	Hα	0.687	0.395	0.296	-1.109	0.028	0.038	-0.316
His(II)	1	C-C	Cα	Cβ	0.792	0.749	0.250	-0.657	0.042	0.051	-0.215
His(II)	2	C-C	Cβ	Cγ	0.720	0.779	0.268	-0.769	0.042	0.055	-0.247
His(II)	31	C-N (arom)	Cγ	Nδ1	0.472	0.903	0.321	-0.828	0.108	0.281	-0.488
His(II)	32	C-C (arom)	Cγ	Cδ2	0.646	0.709	0.342	-1.086	0.447	0.127	-0.399
His(II)	41	C-N (arom)	Cϵ1	Nδ1	0.445	0.845	0.385	-0.949	0.263	0.402	-0.640
His(II)	42	C-N (arom)	Cδ2	Nϵ2	0.448	0.925	0.297	-0.282	0.129	0.374	-0.445
His(II)	5	C-N (arom)	Cϵ1	Nϵ2	0.444	0.903	0.320	-0.423	0.126	0.389	-0.495
His(II)	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.684	0.399	0.289	-1.057	0.012	0.042	-0.306
His(II)	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.687	0.399	0.287	-1.046	0.008	0.041	-0.303
His(II)	31.1	N-H (arom)	Nϵ2	Hϵ2	0.749	0.245	0.356	-2.035	0.046	0.047	-0.556
His(II)	32.1	C-H (arom)	Cδ2	Hδ2	0.686	0.383	0.295	-1.111	0.052	0.035	-0.313
His(II)	41.1	C-H (arom)	Cϵ1	Hϵ1	0.692	0.379	0.301	-1.160	0.045	0.030	-0.321
His(II)	other	O...N	O1	Hδ1	1.609	1.639	0.006	0.023	0.668	0.005	0.001
$His^{(+)}$	0.1	C=O	C	O1	0.399	0.800	0.426	0.047	0.035	0.752	-0.740
$His^{(+)}$	0.2	C-O(H)	C	O2	0.424	0.886	0.324	-0.075	0.133	0.479	-0.498
$His^{(+)}$	0.3	O-H acid	O2	H	0.788	0.167	0.370	-2.818	0.015	0.057	-0.762
$His^{(+)}$	0.4	C-C	Cα	C	0.691	0.838	0.267	-0.787	0.080	0.056	-0.253
$His^{(+)}$	0.5	C-N	Cα	N	0.557	0.887	0.286	-0.934	0.032	0.125	-0.359
$His^{(+)}$	0.6	N-H	N	H1	0.743	0.258	0.354	-1.866	0.051	0.055	-0.522
$His^{(+)}$	0.7	N-H	N	H2	0.739	0.263	0.353	-1.795	0.054	0.058	-0.506
$His^{(+)}$	0.8	C-H (CH sat)	Cα	Hα	0.694	0.393	0.292	-1.083	0.015	0.038	-0.308
$His^{(+)}$	1	C-C	Cα	Cβ	0.772	0.770	0.251	-0.665	0.045	0.049	-0.215
$His^{(+)}$	2	C-C	Cβ	Cγ	0.667	0.835	0.261	-0.746	0.024	0.065	-0.252
$His^{(+)}$	31	C-N (arom)	Cγ	Nδ1	0.451	0.934	0.294	-0.310	0.052	0.360	-0.438
$His^{(+)}$	32	C-C (arom)	Cγ	Cδ2	0.670	0.677	0.350	-1.130	0.476	0.130	-0.412
$His^{(+)}$	41	C-N (arom)	Cϵ1	Nδ1	0.440	0.870	0.357	-0.654	0.170	0.413	-0.577
$His^{(+)}$	42	C-N (arom)	Cδ2	Nϵ2	0.448	0.935	0.290	-0.216	0.091	0.376	-0.430
$His^{(+)}$	5	C-N (arom)	Cϵ1	Nϵ2	0.440	0.875	0.351	-0.598	0.160	0.414	-0.564
$His^{(+)}$	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.695	0.386	0.293	-1.085	0.016	0.038	-0.309
$His^{(+)}$	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.691	0.394	0.288	-1.049	0.016	0.040	-0.302
$His^{(+)}$	31.1	N-H (arom)	Nδ1	Hδ1	0.792	0.214	0.342	-2.286	0.028	0.034	-0.606
$His^{(+)}$	32.1	C-H (arom)	Cδ2	Hδ2	0.703	0.366	0.300	-1.156	0.043	0.029	-0.318
$His^{(+)}$	41.1	C-H (arom)	Cϵ1	Hϵ1	0.716	0.354	0.305	-1.215	0.027	0.023	-0.327
$His^{(+)}$	42.1	N-H (arom)	Nϵ2	Hϵ2	0.767	0.232	0.351	-2.172	0.035	0.039	-0.582
$His^{(+)}$	h	O...H-N	O1	Hδ1	1.246	0.705	0.025	0.110	0.035	0.024	0.003
Phe	0.1	C=O	C	O1	0.397	0.793	0.436	0.112	0.059	0.793	-0.765
Phe	0.2	C-O(H)	C	O2	0.430	0.901	0.307	-0.071	0.164	0.443	-0.461
Phe	0.3	O-H acid	O2	H	0.783	0.170	0.375	-2.825	0.017	0.061	-0.768
Phe	0.4	C-C	Cα	C	0.686	0.838	0.269	-0.796	0.093	0.058	-0.257
Phe	0.5	C-N	Cα	N	0.534	0.910	0.284	-0.920	0.018	0.144	-0.374
Phe	0.6	N-H	N	H1	0.736	0.266	0.354	-1.773	0.053	0.059	-0.503
Phe	0.7	N-H	N	H2	0.739	0.263	0.354	-1.803	0.050	0.058	-0.509
Phe	0.8	C-H (CH sat)	Cα	Hα	0.689	0.394	0.294	-1.097	0.028	0.038	-0.313
Phe	1	C-C	Cα	Cβ	0.795	0.751	0.247	-0.644	0.031	0.050	-0.211
Phe	2	C-C	Cβ	Cγ	0.747	0.766	0.262	-0.724	0.027	0.052	-0.233
Phe	31	C-C (arom)	Cγ	Cδ1	0.702	0.690	0.322	-1.010	0.236	0.097	-0.349

AA	Bond	A-B	A	B	r_b (A)	r_b (B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
Phe	32	C-C (arom)	Cγ	Cδ2	0.697	0.694	0.321	-1.004	0.244	0.097	-0.348
Phe	41	C-C (arom)	Cδ1	Cϵ1	0.692	0.695	0.323	-1.023	0.233	0.098	-0.354
Phe	42	C-C (arom)	Cδ2	Cϵ2	0.695	0.693	0.323	-1.019	0.233	0.097	-0.352
Phe	51	C-C (arom)	Cϵ1	Cζ	0.694	0.693	0.323	-1.024	0.229	0.097	-0.353
Phe	52	C-C (arom)	Cϵ2	Cζ	0.696	0.691	0.324	-1.026	0.233	0.098	-0.355
Phe	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.682	0.405	0.286	-1.038	0.007	0.043	-0.303
Phe	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.681	0.403	0.289	-1.054	0.007	0.043	-0.307
Phe	31.1	C-H (arom)	Cδ1	Hδ1	0.683	0.392	0.294	-1.105	0.019	0.037	-0.313
Phe	32.1	C-H (arom)	Cδ2	Hδ2	0.680	0.397	0.291	-1.082	0.022	0.039	-0.309
Phe	41.1	C-H (arom)	Cϵ1	Hϵ1	0.682	0.394	0.293	-1.095	0.019	0.038	-0.311
Phe	42.1	C-H (arom)	Cϵ2	Hϵ2	0.681	0.394	0.293	-1.094	0.019	0.038	-0.311
Phe	51.1	C-H (arom)	Cζ	Hζ	0.681	0.394	0.293	-1.093	0.020	0.038	-0.311
Tyr	0.1	C=O	C	O1	0.397	0.793	0.436	0.111	0.059	0.793	-0.765
Tyr	0.2	C-O(H)	C	O2	0.430	0.901	0.307	-0.070	0.165	0.443	-0.460
Tyr	0.3	O-H acid	O2	H	0.783	0.170	0.375	-2.825	0.017	0.061	-0.768
Tyr	0.4	C-C	Cα	C	0.685	0.838	0.269	-0.796	0.094	0.059	-0.258
Tyr	0.5	C-N	Cα	N	0.533	0.911	0.284	-0.917	0.018	0.145	-0.374
Tyr	0.6	N-H	N	H1	0.736	0.266	0.354	-1.773	0.053	0.059	-0.503
Tyr	0.7	N-H	N	H2	0.739	0.263	0.354	-1.802	0.050	0.058	-0.508
Tyr	0.8	C-H (CH sat)	Cα	Hα	0.689	0.394	0.294	-1.097	0.028	0.038	-0.312
Tyr	1	C-C	Cα	Cβ	0.797	0.749	0.248	-0.646	0.030	0.050	-0.212
Tyr	2	C-C	Cβ	Cγ	0.749	0.764	0.261	-0.718	0.034	0.052	-0.232
Tyr	31	C-C (arom)	Cγ	Cδ1	0.691	0.704	0.319	-0.997	0.233	0.095	-0.345
Tyr	32	C-C (arom)	Cγ	Cδ2	0.680	0.707	0.323	-1.011	0.263	0.101	-0.354
Tyr	41	C-C (arom)	Cδ1	Cϵ1	0.691	0.690	0.325	-1.025	0.252	0.101	-0.357
Tyr	42	C-C (arom)	Cδ2	Cϵ2	0.693	0.697	0.320	-0.998	0.242	0.097	-0.346
Tyr	51	C-C (arom)	Cϵ1	Cζ	0.644	0.745	0.326	-1.059	0.266	0.101	-0.366
Tyr	52	C-C (arom)	Cϵ2	Cζ	0.648	0.736	0.327	-1.051	0.297	0.105	-0.368
Tyr	6	C-O(H)	Cζ	Oη	0.435	0.920	0.287	-0.010	0.100	0.415	-0.417
Tyr	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.682	0.405	0.286	-1.036	0.007	0.043	-0.302
Tyr	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.680	0.404	0.289	-1.054	0.009	0.043	-0.307
Tyr	31.1	C-H (arom)	Cδ1	Hδ1	0.685	0.389	0.296	-1.116	0.013	0.036	-0.315
Tyr	32.1	C-H (arom)	Cδ2	Hδ2	0.681	0.396	0.292	-1.090	0.018	0.038	-0.310
Tyr	41.1	C-H (arom)	Cϵ1	Hϵ1	0.685	0.390	0.292	-1.089	0.027	0.037	-0.310
Tyr	42.2	C-H (arom)	Cϵ2	Hϵ2	0.677	0.400	0.288	-1.058	0.032	0.041	-0.306
Tyr	6.1	O-H phenolic	Oη	Hη	0.773	0.174	0.384	-2.855	0.022	0.067	-0.781
Trp	0.1	C=O	C	O1	0.397	0.793	0.437	0.102	0.061	0.793	-0.768
Trp	0.2	C-O(H)	C	O2	0.430	0.901	0.306	-0.064	0.162	0.443	-0.459
Trp	0.3	O-H acid	O2	H	0.783	0.170	0.375	-2.825	0.017	0.061	-0.767
Trp	0.4	C-C	Cα	C	0.688	0.838	0.268	-0.797	0.048	0.057	-0.256
Trp	0.5	C-N	Cα	N	0.540	0.913	0.278	-0.879	0.015	0.139	-0.359
Trp	0.6	N-H	N	H1	0.736	0.266	0.354	-1.766	0.054	0.059	-0.501
Trp	0.7	N-H	N	H2	0.740	0.263	0.354	-1.805	0.051	0.058	-0.509
Trp	0.8	C-H (CH sat)	Cα	Hα	0.686	0.396	0.297	-1.112	0.024	0.039	-0.317
Trp	1	C-C	Cα	Cβ	0.792	0.741	0.254	-0.684	0.039	0.052	-0.223
Trp	2	C-C	Cβ	Cγ	0.734	0.772	0.261	-0.723	0.052	0.054	-0.235
Trp	31	C-C (arom)	Cγ	Cδ1	0.629	0.724	0.340	-1.068	0.445	0.133	-0.400
Trp	32	C-C (arom)	Cγ	Cδ2	0.716	0.734	0.288	-0.826	0.170	0.078	-0.284
Trp	41	C-N (arom)	Cδ1	Nϵ1	0.454	0.922	0.302	-0.458	0.074	0.341	-0.455
Trp	42	C-C (arom)	Cδ2	Cϵ2	0.663	0.737	0.323	-1.023	0.242	0.099	-0.354
Trp	43	C-C (arom)	Cδ2	Cϵ3	0.708	0.693	0.314	-0.976	0.206	0.090	-0.335
Trp	51	C-N (arom)	Cϵ2	Nϵ1	0.454	0.915	0.307	-0.502	0.041	0.340	-0.465

AA	Bond	A-B	A	B	r_b (A)	r_b (B)	ρ_b	$\nabla^2\rho_b$	ϵ	G_b	$H_b=G_b+V_b$
Trp	52	C-C (arom)	Cϵ2	Cζ2	0.739	0.657	0.317	-0.998	0.233	0.096	-0.345
Trp	53	C-C (arom)	Cϵ3	Cζ3	0.700	0.677	0.327	-1.033	0.279	0.105	-0.363
Trp	62	C-C (arom)	Cζ2	Cη	0.697	0.680	0.327	-1.030	0.283	0.105	-0.362
Trp	63	C-C (arom)	Cζ3	Cη	0.694	0.712	0.312	-0.965	0.210	0.090	-0.331
Trp	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.687	0.398	0.288	-1.047	0.013	0.041	-0.303
Trp	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.684	0.402	0.287	-1.042	0.011	0.043	-0.303
Trp	31.1	C-H (arom)	Cδ1	Hδ1	0.687	0.383	0.298	-1.131	0.040	0.034	-0.317
Trp	41.1	N-H (arom)	Nϵ1	Hϵ1	0.745	0.247	0.356	-1.996	0.057	0.050	-0.549
Trp	43.1	C-H (arom)	Cϵ3	Hϵ3	0.682	0.394	0.292	-1.093	0.018	0.037	-0.311
Trp	52.1	C-H (arom)	Cζ2	Hζ2	0.680	0.396	0.290	-1.075	0.028	0.039	-0.308
Trp	53.1	C-H (arom)	Cζ3	Hζ3	0.680	0.395	0.292	-1.086	0.028	0.039	-0.310
Trp	62.1	C-H (arom)	Cη	Hη	0.681	0.394	0.293	-1.094	0.021	0.038	-0.311
Trp	h	O...H-C	O2	Hδ1	1.588	1.221	0.006	0.023	0.572	0.005	0.001
Pro	0.1	C=O	C	O1	0.397	0.794	0.435	0.106	0.058	0.791	-0.764
Pro	0.2	C-O(H)	C	O2	0.430	0.901	0.307	-0.073	0.166	0.442	-0.461
Pro	0.3	O-H acid	O2	H	0.782	0.170	0.376	-2.828	0.017	0.062	-0.769
Pro	0.4	C-C	Cα	C	0.683	0.838	0.269	-0.795	0.104	0.059	-0.258
Pro	0.5	C-N	Cα	N	0.545	0.902	0.285	-0.917	0.025	0.137	-0.366
Pro	0.6	N-H	N	H1	0.738	0.260	0.358	-1.854	0.058	0.058	-0.522
Pro	0.7	C-H (CH sat)	Cα	Hα	0.689	0.397	0.292	-1.081	0.035	0.039	-0.309
Pro	1	C-C	Cα	Cβ	0.800	0.739	0.250	-0.656	0.021	0.053	-0.217
Pro	2	C-C	Cβ	Cγ	0.773	0.758	0.253	-0.668	0.003	0.054	-0.221
Pro	3	C-C	Cγ	Cδ	0.763	0.771	0.255	-0.682	0.021	0.052	-0.222
Pro	4	C-N	Cδ	N	0.534	0.936	0.265	-0.792	0.033	0.145	-0.343
Pro	1.1	C-H (CH2 sat)	Cβ	Hβ-1	0.680	0.402	0.288	-1.055	0.006	0.042	-0.306
Pro	1.2	C-H (CH2 sat)	Cβ	Hβ-2	0.680	0.405	0.287	-1.044	0.009	0.043	-0.304
Pro	2.1	C-H (CH2 sat)	Cγ	Hγ-1	0.677	0.409	0.286	-1.035	0.006	0.044	-0.303
Pro	2.2	C-H (CH2 sat)	Cγ	Hγ-2	0.677	0.407	0.286	-1.038	0.001	0.043	-0.303
Pro	3.1	C-H (CH2 sat)	Cδ	Hδ-1	0.681	0.405	0.291	-1.080	0.034	0.040	-0.310
Pro	3.2	C-H (CH2 sat)	Cδ	Hδ-2	0.681	0.404	0.290	-1.071	0.026	0.041	-0.308

Table A5.2. Selected atomic properties of the genetically-encoded amino acids (all values in au)

Explanation of the table: The first three columns list the atomic electron populations $N(\Omega)$ (Eq. 2.62), the atomic energies $E(\Omega)$ (Eq.2.67), and the atomic volume integrated to the 0.001 au isodensity envelope (v(Ω)). The three columns labeled $\mu_i(\Omega)$ list the components of the atomic dipolar polarization (Eq. 2.68) and the five labeled $Q_{ij}(\Omega)$ list the unique components of the symmetric traceless quadrupole tensor (Eq. 2.70) – both given with respect to the molecular coordinate system defined by the nuclear coordinates x,y, and z given in the last three columns. The magnitudes of the dipolar and quadrupolar polarizations are labeled $|\boldsymbol{\mu}|$ and $|\mathbf{Q}|$, respectively, and are defined in Eqs. 2.69 and 2.74. Two components of the diagonalized quadrupolar polarization traceless tensor (labeled $\mathcal{Q}_{ii}(\Omega)$) are also given with respect to a local (rotated) atomic coordinate system. Only the Q_{xx} and Q_{yy} components of the original and of the rotated quadrupolar polarization tensors are given since this tensor being traceless two components are sufficient to define the third (Eq. 2.73). (Also see section 2.5.2).

		$N(\Omega)$	$E(\Omega)$	v(Ω)	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\mathcal{Q}_{xx}(\Omega)$	$\mathcal{Q}_{yy}(\Omega)$	$\lvert\mathbf{Q}(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
Gly	C	4.210	-36.69443	32.1	1.790	0.801	0.582	-0.001	0.990	1.531	-0.612	1.541	-0.107	-0.909	-0.001	1.026	0.000	1.0257	0.2128	0.0009
Gly	O1	9.354	-75.69582	136.0	-1.354	0.007	0.583	-0.000	0.583	0.050	0.495	0.603	0.050	0.008	-0.000	0.495	0.000	1.1149	2.4452	0.0006
Gly	O2	9.305	-75.62877	121.4	-1.305	0.057	-0.301	-0.000	0.306	0.017	0.837	0.976	0.796	0.179	-0.001	0.058	-0.000	3.0801	-1.2312	-0.0014
Gly	H	0.349	-0.32763	18.4	0.651	-0.114	-0.084	0.000	0.142	0.038	-0.067	0.068	-0.027	-0.051	0.000	-0.002	0.000	4.5297	-0.1845	-0.0025
Gly	N	8.153	-54.89487	115.5	-1.153	0.148	0.167	-0.000	0.223	1.121	1.233	2.356	-1.122	-1.663	0.004	-0.111	0.003	-3.6931	0.0374	-0.0013
Gly	H1	0.631	-0.48157	31.0	0.369	0.006	-0.114	-0.143	0.183	-0.037	-0.064	0.102	-0.034	-0.009	-0.021	-0.025	0.069	-3.8166	1.1583	1.5142
Gly	H2	0.631	-0.48162	30.9	0.369	0.006	-0.113	0.143	0.183	-0.037	-0.064	0.102	-0.034	-0.009	0.021	-0.026	-0.068	-3.8164	1.1515	-1.5219
Gly	Cα	5.383	-37.46920	49.3	0.617	-0.321	0.348	-0.000	0.473	-0.220	1.101	1.165	0.972	-0.391	0.001	-0.092	-0.001	-1.3654	-1.3642	0.0020
Gly	Hα	0.991	-0.62601	47.3	0.009	-0.007	0.071	-0.101	0.124	-0.159	-0.255	0.418	-0.159	-0.014	0.015	-0.015	-0.321	-1.2841	-2.5923	1.6422
Gly	Hα	0.991	-0.62596	47.2	0.009	-0.007	0.071	0.101	0.124	-0.159	-0.255	0.418	-0.158	-0.014	-0.016	-0.013	0.321	-1.2828	-2.5980	-1.6339
Gly	SUM	39.997	-282.92589	629.1	0.003															
Gly	SCF	40	-282.92550		0	0.402	-0.317	-0.001	0.512											
Ala	C	4.236	-36.71416	31.7	1.764	0.938	-0.226	0.288	1.007	-0.679	-0.855	1.537	-0.572	0.429	-0.205	0.991	-0.897	1.4767	-0.3137	-0.0161
Ala	O1	9.350	-75.68277	136.0	-1.350	0.166	-0.476	0.286	0.579	0.066	0.496	0.615	0.084	-0.054	0.156	0.223	-0.425	2.1886	-2.1354	1.0931
Ala	O2	9.303	-75.62921	120.3	-1.303	-0.030	0.268	-0.150	0.309	0.830	0.012	0.966	0.801	0.112	0.187	-0.158	-0.334	3.0419	1.5162	-0.7398
Ala	H	0.351	-0.32673	18.6	0.649	-0.131	0.034	-0.042	0.142	0.037	0.027	0.065	-0.048	0.025	-0.028	0.028	0.004	4.7108	1.1012	-0.2071
Ala	N	8.155	-54.91603	113.2	-1.155	0.196	-0.086	0.063	0.223	1.069	1.278	2.350	-2.043	0.566	-0.794	1.051	0.247	-2.8783	-1.9507	-0.5454
Ala	H1	0.638	-0.48447	31.4	0.362	-0.065	0.138	0.105	0.185	-0.044	-0.064	0.108	-0.041	-0.014	-0.020	0.021	0.083	-2.3003	-3.3493	-1.6823
Ala	H2	0.630	-0.47982	31.0	0.370	-0.007	0.069	-0.169	0.183	-0.029	-0.060	0.092	-0.031	0.009	-0.002	-0.045	-0.043	-2.8517	-2.6445	1.2167
Ala	Cα	5.407	-37.47807	41.8	0.593	-0.152	-0.473	0.046	0.499	1.013	-0.154	1.092	0.520	0.634	0.128	0.192	-0.261	-1.2529	0.2295	-0.7511
Ala	Hα	1.009	-0.64179	46.9	-0.009	-0.007	-0.029	0.120	0.124	-0.165	-0.228	0.394	-0.173	0.021	0.005	-0.161	-0.179	-1.2297	0.8344	-2.7092
Ala	Cβ	5.850	-37.72000	63.7	0.150	0.017	-0.057	-0.022	0.063	-0.033	-0.166	0.213	-0.081	-0.001	-0.081	0.097	0.140	-2.2837	2.3999	0.8643
Ala	Hβ1	1.010	-0.62704	49.0	-0.010	0.120	-0.041	0.030	0.130	-0.193	-0.256	0.451	0.360	-0.120	0.201	-0.187	-0.011	-4.1972	2.8165	0.2731

		$N(\Omega)$	$E(\Omega)$	$\nu(\Omega)$	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\lvert\mathbf{Q}(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
Ala	Hβ2	1.040	-0.63803	50.9	-0.040	0.001	0.031	-0.126	0.130	-0.261	-0.200	0.462	-0.256	-0.011	-0.023	-0.158	-0.170	-2.3283	1.8696	2.8473
Ala	Hβ3	1.021	-0.63294	48.6	-0.021	-0.059	-0.107	0.007	0.123	-0.185	-0.279	0.467	0.046	0.313	-0.074	0.214	-0.051	-1.1403	4.0862	0.6703
Ala	SUM	48.001	-321.97106	495.2	-0.001															
Ala	SCF	48	-321.97134		0	0.252	0.480	-0.193	0.576											
Val	C	4.239	-36.71436	31.0	1.761	-0.914	-0.276	0.333	1.011	-0.689	-0.845	1.537	-0.654	-0.257	-0.248	0.715	1.083	-2.6593	-0.1692	-0.1803
Val	O1	9.352	-75.68739	135.1	-1.352	-0.332	0.298	0.367	0.578	0.064	0.469	0.582	0.105	-0.006	-0.336	0.034	0.355	-4.0122	0.9324	1.2391
Val	O2	9.303	-75.63251	119.6	-1.303	0.136	-0.200	-0.196	0.311	0.830	0.016	0.968	0.467	0.517	-0.342	0.051	0.273	-3.3007	-2.3795	-1.1937
Val	H	0.351	-0.32694	18.6	0.649	0.128	0.037	-0.049	0.142	0.038	0.027	0.066	-0.047	-0.020	0.031	0.030	0.013	-4.9250	-2.8516	-0.5780
Val	N	8.162	-54.92279	109.9	-1.162	-0.193	-0.029	0.051	0.201	1.234	0.987	2.226	-1.735	-1.073	0.533	0.737	0.069	0.4031	3.3461	-0.4169
Val	H2	0.625	-0.47883	29.8	0.375	0.044	-0.036	-0.169	0.178	-0.060	-0.032	0.093	-0.034	0.010	-0.019	-0.054	0.007	-0.0184	3.6888	1.3960
Val	H1	0.640	-0.48548	31.4	0.360	0.117	-0.109	0.093	0.185	-0.045	-0.065	0.111	-0.031	-0.038	0.050	0.019	-0.067	-0.6987	4.4934	-1.4419
Val	Cα	5.431	-37.49858	40.8	0.569	-0.104	0.494	0.118	0.519	1.033	-0.097	1.140	-0.119	0.590	-0.260	0.685	0.397	-0.0651	0.7312	-1.0505
Val	Hα	1.014	-0.64463	46.5	-0.014	0.012	0.002	0.124	0.124	-0.171	-0.205	0.376	-0.173	0.007	-0.024	-0.196	0.059	-0.0650	0.5638	-3.0942
Val	Cβ	5.854	-37.71594	46.6	0.146	-0.056	0.019	-0.029	0.065	-0.075	-0.177	0.258	0.046	-0.129	0.165	0.001	-0.065	2.0341	-1.0403	-0.0406
Val	Hβ	1.047	-0.65973	46.8	-0.047	0.024	0.114	0.027	0.120	-0.159	-0.253	0.416	-0.083	0.186	0.060	0.297	0.143	1.4697	-2.9540	-0.5203
Val	Cγ1	5.852	-37.71888	62.4	0.148	-0.010	-0.019	-0.008	0.023	0.056	0.095	0.153	0.042	-0.073	-0.039	-0.107	-0.045	4.5438	-0.4913	-1.3681
Val	Hγ1-1	1.058	-0.64601	51.3	-0.058	0.005	0.014	0.130	0.130	-0.195	-0.276	0.473	-0.187	0.002	0.073	-0.266	0.080	4.3744	-0.7167	-3.4019
Val	Hγ1-2	1.018	-0.63154	48.3	-0.018	-0.063	-0.112	-0.020	0.130	-0.181	-0.267	0.450	-0.159	0.134	0.005	0.389	0.139	5.1750	1.4193	-0.9922
Val	Hγ1-3	1.054	-0.64439	51.2	-0.054	-0.097	0.080	-0.036	0.131	-0.277	-0.192	0.471	0.105	-0.318	0.136	0.033	-0.177	5.9921	-1.7919	-0.7169
Val	Cγ2	5.869	-37.72406	62.7	0.131	0.013	-0.003	0.020	0.024	-0.068	-0.024	0.095	0.013	0.002	-0.055	-0.063	-0.022	2.3299	-0.8850	2.8368
Val	Hγ2-1	1.049	-0.64346	50.3	-0.049	0.103	0.025	-0.064	0.124	-0.288	-0.192	0.484	0.345	0.092	-0.253	-0.262	-0.073	0.5695	-1.2602	3.8241
Val	Hγ2-2	1.044	-0.64047	50.7	-0.044	-0.085	0.086	-0.049	0.131	-0.270	-0.194	0.466	0.073	-0.339	0.117	0.070	-0.136	3.7006	-2.2705	3.4770
Val	Hγ2-3	1.041	-0.64108	49.3	-0.041	-0.043	-0.109	-0.051	0.128	-0.263	-0.189	0.454	-0.192	0.185	0.069	0.366	0.123	3.0060	0.9618	3.4191
Val	SUM	64.002	-400.05706	638.4	-0.002															
Val	SCF	64	-400.05767		0	-0.013	-0.473	-0.319	0.571											
Leu	C	4.232	-36.71607	31.2	1.768	0.882	-0.392	-0.273	1.003	-0.672	-0.847	1.522	-0.474	0.508	-0.413	0.620	-0.997	3.7175	-0.4715	-0.0275
Leu	O1	9.355	-75.67333	135.9	-1.355	0.403	0.253	-0.339	0.584	0.080	0.479	0.604	0.149	0.010	-0.391	0.059	-0.295	5.3071	0.4432	-1.3292
Leu	O2	9.300	-75.62376	120.0	-1.300	-0.174	-0.175	0.182	0.307	0.834	0.009	0.968	0.341	-0.547	-0.311	0.220	-0.259	3.9722	-2.7535	0.9973
Leu	H	0.351	-0.32700	18.6	0.649	-0.125	0.054	0.041	0.142	0.038	0.027	0.065	-0.044	0.031	0.024	0.023	-0.015	5.5513	-3.4494	0.4849
Leu	N	8.150	-54.92043	110.7	-1.150	0.188	-0.064	-0.058	0.207	1.232	0.989	2.226	-1.297	1.376	0.655	0.363	-0.231	1.1840	3.4501	0.1773
Leu	H1	0.640	-0.48528	31.4	0.360	-0.134	-0.082	-0.097	0.185	-0.044	-0.064	0.109	-0.007	0.040	0.063	-0.003	0.057	2.4831	4.3360	1.2311
Leu	H2	0.630	-0.47967	30.9	0.370	-0.046	-0.035	0.173	0.182	-0.027	-0.062	0.091	-0.042	-0.013	-0.027	-0.037	-0.023	1.6199	3.8167	-1.6294
Leu	Cα	5.429	-37.50539	40.5	0.571	0.172	0.466	-0.081	0.504	1.011	-0.147	1.093	-0.257	-0.390	-0.268	0.868	-0.275	1.2040	0.7613	0.6437
Leu	Hα	1.012	-0.64543	45.1	-0.012	0.001	0.014	-0.118	0.119	-0.152	-0.240	0.396	-0.138	-0.015	-0.138	-0.205	-0.094	0.9211	0.4496	2.6472
Leu	Cβ	5.874	-37.75052	52.4	0.126	0.098	-0.005	-0.023	0.101	-0.019	-0.312	0.372	0.147	0.190	0.216	-0.022	-0.045	-0.9160	-0.5713	-0.8375

		$N(\Omega)$	$E(\Omega)$	v(Ω)	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\lvert\mathbf{Q}(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
Leu	Hβ1	1.040	-0.64735	48.1	-0.040	-0.005	0.123	-0.014	0.124	-0.119	-0.294	0.425	-0.114	-0.058	0.033	0.377	-0.138	-0.7983	-2.5901	-0.4834
Leu	Hβ2	1.038	-0.65398	46.9	-0.038	0.022	-0.120	0.031	0.126	-0.164	-0.243	0.409	-0.167	0.040	0.015	0.390	-0.096	-3.6797	2.3659	-0.6201
Leu	Cγ	5.846	-37.70897	47.9	0.154	0.015	-0.014	-0.022	0.030	-0.021	-0.224	0.272	0.118	-0.016	-0.208	-0.020	0.009	-3.6039	0.3520	-0.2454
Leu	Hγ	1.052	-0.64342	51.1	-0.052	0.125	-0.038	-0.013	0.131	-0.273	-0.193	0.468	0.358	-0.229	-0.103	-0.173	0.072	-7.3942	-0.2855	-1.6863
Leu	Cδ1	5.858	-37.72053	62.1	0.142	-0.026	-0.021	-0.022	0.040	-0.057	-0.013	0.075	0.017	-0.007	-0.044	-0.051	-0.015	-5.4858	-0.9699	-2.0051
Leu	Hδ1-1	1.059	-0.64501	51.8	-0.059	0.006	0.133	-0.014	0.134	-0.267	-0.190	0.459	-0.225	-0.004	0.040	0.437	-0.113	-5.4936	-3.0021	-1.6964
Leu	Hδ1-2	1.050	-0.64293	50.9	-0.050	0.123	-0.034	-0.030	0.131	-0.277	-0.187	0.467	0.400	-0.194	-0.069	-0.221	0.030	-6.2860	0.4866	2.8435
Leu	Hδ1-3	1.056	-0.64519	51.2	-0.056	-0.023	-0.016	0.128	0.131	-0.276	-0.191	0.470	-0.156	0.052	-0.157	-0.249	-0.118	-5.0275	-0.6394	-3.9790
Leu	Cδ2	5.861	-37.72181	62.2	0.139	-0.014	-0.007	0.036	0.039	-0.005	-0.058	0.069	0.014	0.016	0.037	0.009	0.039	-4.3374	-0.0717	2.5220
Leu	Hδ2-1	1.052	-0.64675	49.7	-0.052	-0.063	-0.061	-0.089	0.125	-0.288	-0.187	0.478	0.079	0.249	0.222	-0.072	0.234	-3.1741	1.0277	3.8023
Leu	Hδ2-2	1.059	-0.65319	50.2	-0.059	-0.033	-0.012	0.126	0.131	-0.280	-0.133	0.421	-0.128	0.041	-0.072	-0.261	-0.091	-0.5508	-0.3037	-2.8453
Leu	Hδ2-3	1.060	-0.64489	52.0	-0.060	-0.010	0.127	-0.041	0.134	-0.267	-0.191	0.460	-0.251	-0.071	-0.009	0.416	-0.149	-4.1640	-2.0532	3.0435
Leu	SUM	72.003	-439.10090	1240.8	-0.003															
Leu	SCF	72	-439.10149		0	-0.119	-0.412	0.264	0.503											
Ile	C	4.254	-36.72072	31.9	1.746	0.614	0.609	0.542	1.021	-0.290	-1.549	1.978	-0.367	-0.958	1.086	0.847	-0.543	2.7528	1.1075	0.2046
Ile	O1	9.353	-75.68995	134.4	-1.353	0.363	-0.185	0.406	0.575	0.052	0.451	0.552	-0.085	-0.065	0.454	0.090	-0.102	4.1979	0.4777	1.8104
Ile	O2	9.302	-75.63649	118.9	-1.302	-0.182	0.157	-0.204	0.315	0.797	0.018	0.931	-0.230	0.284	0.555	0.519	0.244	2.3597	3.5251	-0.3575
Ile	H	0.352	-0.32750	18.7	0.648	-0.080	-0.080	-0.087	0.143	0.039	0.027	0.066	-0.002	-0.032	-0.031	0.009	-0.035	3.3665	4.5433	0.7345
Ile	N	8.156	-54.91466	112.1	-1.156	0.141	0.125	0.093	0.210	1.046	1.330	2.381	0.075	-1.385	-0.953	-0.678	-1.006	2.4620	-3.1074	-1.7449
Ile	H1	0.644	-0.48752	31.7	0.356	-0.156	-0.026	0.100	0.187	-0.045	-0.072	0.117	0.011	-0.008	-0.092	-0.047	-0.003	3.9720	-2.9196	-2.8714
Ile	H2	0.621	-0.47750	29.1	0.379	-0.067	0.063	-0.151	0.177	-0.034	-0.064	0.100	-0.056	0.003	0.026	-0.030	-0.031	3.1039	-3.7710	-0.0934
Ile	Cα	5.440	-37.49883	41.1	0.560	0.384	-0.344	-0.025	0.516	0.985	-0.073	1.098	-0.392	-0.324	0.348	0.844	0.378	1.1959	-0.7018	-1.4101
Ile	Hα	1.015	-0.64775	45.7	-0.015	-0.001	-0.049	0.113	0.123	-0.206	-0.167	0.374	-0.187	-0.049	0.063	-0.066	-0.215	0.9990	0.1612	-3.2559
Ile	Cβ	5.867	-37.72626	45.7	0.133	0.062	-0.040	-0.054	0.092	-0.061	-0.201	0.274	0.152	-0.034	-0.185	-0.102	-0.054	-1.5090	-1.1040	-0.3498
Ile	Hβ	1.036	-0.65269	47.4	-0.036	0.063	0.083	0.080	0.131	-0.123	-0.241	0.370	-0.051	0.118	0.129	0.043	0.264	-2.3252	-2.4677	-1.6489
Ile	Cγ1	5.850	-37.71918	52.2	0.150	0.009	0.032	-0.012	0.036	-0.046	-0.031	0.078	0.053	-0.047	-0.014	-0.022	0.001	-3.1203	1.3140	-0.4661
Ile	Hγ1-1	1.061	-0.65733	49.3	-0.061	-0.015	-0.058	0.112	0.128	-0.144	-0.273	0.423	-0.126	0.026	-0.118	-0.128	-0.267	-2.7935	2.2498	-2.2652
Ile	Hγ1-2	1.058	-0.65631	48.9	-0.058	-0.027	-0.080	-0.093	0.125	-0.166	-0.256	0.426	-0.081	0.150	0.145	0.023	0.295	-2.5163	2.6398	0.9804
Ile	Cδ1	5.849	-37.70470	63.7	0.151	-0.024	-0.005	0.002	0.025	0.004	0.036	0.044	0.016	0.006	-0.030	0.007	-0.008	-5.9591	0.8352	-0.1897
Ile	Hδ1-1	1.054	-0.64427	51.4	-0.054	0.050	0.086	0.088	0.132	-0.193	-0.275	0.470	-0.134	0.124	0.143	0.044	0.340	-6.6294	-0.4938	-1.6045
Ile	Hδ1-2	1.054	-0.64757	50.1	-0.054	0.041	0.043	-0.111	0.126	-0.186	-0.292	0.483	-0.185	0.009	-0.108	-0.150	-0.280	-6.4452	0.0863	1.6553
Ile	Hδ1-3	1.049	-0.64237	51.1	-0.049	0.072	-0.108	0.014	0.131	-0.216	-0.263	0.480	-0.046	-0.297	0.036	0.296	-0.084	-7.0054	2.5834	-0.4344
Ile	Cγ2	5.864	-37.72376	62.0	0.136	-0.010	-0.014	0.013	0.021	-0.035	-0.001	0.041	-0.016	-0.024	0.005	-0.003	-0.019	-1.5049	-2.2940	2.2922
Ile	Hγ2-1	1.042	-0.64313	48.6	-0.042	-0.061	0.107	-0.015	0.124	0.470	-0.177	0.475	0.001	-0.339	-0.072	0.221	0.000	-0.4121	-4.0261	2.3378
Ile	Hγ2-2	1.045	-0.64216	49.9	-0.045	-0.037	-0.078	-0.093	0.127	-0.256	-0.201	0.458	-0.104	0.210	0.179	0.039	0.270	-0.7309	-1.0200	3.7028
Ile	Hγ2-3	1.046	-0.64459	49.2	-0.046	0.110	0.037	-0.045	0.124	-0.183	-0.297	0.485	0.432	0.094	-0.146	-0.215	-0.078	-3.4104	-2.7774	2.8671

		$N(\Omega)$	$E(\Omega)$	v(Ω)	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\|\mu(\Omega)\|$	$\mathcal{Q}_{xx}(\Omega)$	$\mathcal{Q}_{yy}(\Omega)$	$\|\mathbf{Q}(\Omega)\|$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
Ile	SUM	72.014	-439.10526	1233.1	-0.014															
Ile	SCF	72	-439.09703		0	-0.207	0.537	-0.269	0.635											
Ser	C	4.251	-36.71495	32.6	1.749	-0.715	-0.693	0.231	1.022	-0.681	-0.873	1.558	0.469	-1.068	-0.407	0.278	0.295	-1.7477	-0.8779	-0.0650
Ser	O1	9.344	-75.68674	134.9	-1.344	-0.555	0.103	0.171	0.589	0.086	0.473	0.602	0.373	-0.068	-0.284	0.099	0.019	-3.8702	-0.5595	0.5916
Ser	O2	9.310	-75.63040	116.9	-1.310	0.275	-0.102	-0.085	0.306	0.100	0.811	1.000	0.267	0.320	-0.264	0.585	-0.085	-0.7689	-3.1913	-0.3202
Ser	H	0.345	-0.32302	18.2	0.655	0.099	0.094	-0.030	0.139	0.037	0.027	0.064	-0.016	-0.047	0.014	-0.006	0.015	-2.0399	-4.4078	0.0676
Ser	N	8.151	-54.91100	112.4	-1.151	-0.165	-0.133	0.054	0.219	1.198	1.087	2.286	-0.243	-1.647	0.450	-0.750	0.449	-0.8799	3.7026	-0.2537
Ser	H1	0.626	-0.47874	30.6	0.374	0.148	-0.027	0.099	0.180	-0.037	-0.063	0.100	0.024	-0.018	0.075	-0.031	-0.027	-2.3922	4.0264	-1.3429
Ser	H2	0.634	-0.48163	31.3	0.366	0.061	-0.013	-0.173	0.184	-0.034	-0.061	0.096	-0.049	0.002	-0.036	-0.032	0.024	-1.4678	3.8920	1.5370
Ser	Cα	5.435	-37.51261	42.1	0.565	-0.341	0.366	0.087	0.508	1.021	-0.090	1.130	-0.210	-0.106	-0.321	0.991	0.159	0.1341	1.2108	-0.7004
Ser	Hα	0.986	-0.62909	46.2	0.014	-0.006	0.016	0.124	0.125	-0.208	-0.174	0.382	-0.160	-0.026	-0.139	-0.183	0.036	0.5705	1.0256	-2.6958
Ser	Cβ	5.257	-37.35266	45.5	0.743	0.317	-0.490	-0.228	0.627	-0.747	-0.387	1.153	0.041	-0.857	-0.197	0.256	0.382	2.6408	0.9054	0.7437
Ser	Hβ1	0.982	-0.63279	45.6	0.018	-0.069	-0.104	0.005	0.125	-0.140	-0.246	0.391	-0.005	0.225	-0.045	0.233	-0.095	3.7034	2.6208	0.4558
Ser	Hβ2	1.052	-0.65635	50.3	-0.052	0.034	0.003	-0.125	0.130	-0.183	-0.242	0.427	-0.179	-0.010	-0.091	-0.224	-0.077	2.2780	0.7180	2.7618
Ser	Oγ	9.263	-75.49259	120.1	-1.263	0.182	0.083	-0.072	0.212	-0.063	0.882	0.984	-0.122	-0.029	-0.215	0.843	0.238	4.1457	-1.0726	-0.1357
Ser	Hγ	0.369	-0.34379	16.5	0.631	0.064	0.124	-0.018	0.141	-0.046	-0.002	0.054	0.044	0.004	0.013	-0.042	0.014	3.2912	-2.6364	0.0372
Ser	SUM	56.004	-396.84636	843.1	-0.004															
Ser	SCF	56	-396.84708		0	-0.441	-0.528	0.130	0.700											
Thr	C	4.215	-36.69314	30.7	1.785	-0.908	-0.265	0.270	0.984	-0.650	-0.850	1.505	-0.553	-0.414	-0.297	0.899	0.907	-2.6367	-0.1374	-0.2203
Thr	O1	9.346	-75.68624	135.0	-1.346	-0.378	0.323	0.304	0.583	0.062	0.465	0.576	0.167	-0.078	-0.309	0.116	0.294	-4.1662	1.0609	0.9044
Thr	O2	9.304	-75.62962	119.2	-1.304	0.164	-0.212	-0.151	0.308	0.831	0.014	0.967	0.516	0.457	-0.299	0.133	0.227	-3.0958	-2.4711	-1.0408
Thr	H	0.350	-0.32650	18.5	0.650	0.131	0.037	-0.040	0.142	0.038	0.027	0.065	-0.051	-0.021	0.025	0.031	0.010	-4.7564	-2.9437	-0.5325
Thr	N	8.157	-54.92149	110.0	-1.157	-0.181	-0.074	0.085	0.214	1.267	0.937	2.212	-1.175	-1.212	1.007	0.626	0.389	0.4903	3.3394	-0.1170
Thr	H1	0.629	-0.48055	30.5	0.371	0.133	-0.112	0.048	0.181	-0.029	-0.068	0.099	0.032	-0.053	0.034	0.015	-0.041	-0.8772	4.5297	-0.6571
Thr	H2	0.612	-0.47193	29.1	0.388	0.007	-0.010	-0.174	0.175	-0.067	-0.037	0.105	-0.058	-0.002	0.036	-0.031	-0.032	0.5210	3.3690	1.7762
Thr	Cα	5.418	-37.51223	40.5	0.582	-0.046	0.444	0.188	0.485	-0.166	-0.750	0.976	-0.125	0.496	-0.117	0.589	0.407	-0.0258	0.8154	-0.9970
Thr	Hα	1.012	-0.64071	47.3	-0.012	0.008	0.003	0.122	0.123	-0.215	-0.176	0.391	-0.183	0.016	-0.019	-0.207	-0.006	-0.0221	0.8179	-3.0507
Thr	Cβ	5.335	-37.39484	39.8	0.665	-0.090	-0.005	0.631	0.637	-0.555	-0.491	1.047	-0.513	-0.031	-0.048	-0.532	0.003	2.0318	-1.0318	-0.1331
Thr	Hβ	1.027	-0.65900	45.3	-0.027	0.006	0.109	0.051	0.120	-0.143	-0.256	0.405	-0.076	0.169	0.060	0.256	0.196	1.5623	-2.8871	-0.8652
Thr	Oγ1	9.254	-75.48423	117.5	-1.254	-0.141	0.138	0.038	0.201	-0.850	-0.120	1.057	-0.241	-0.589	0.329	-0.268	-0.434	1.8860	-1.0728	2.5300
Thr	Hγ1	0.390	-0.35427	20.3	0.610	-0.105	0.104	-0.056	0.158	0.029	0.034	0.063	-0.012	0.044	-0.019	-0.013	0.015	3.0630	-2.2352	3.2130
Thr	Cγ2	5.856	-37.73627	63.2	0.144	-0.027	-0.036	0.000	0.045	0.156	0.027	0.198	0.116	-0.071	-0.077	-0.102	-0.080	4.6547	-0.2830	-1.0453
Thr	Hγ2-1	0.996	-0.62257	46.9	0.004	-0.058	-0.106	-0.042	0.128	-0.174	-0.270	0.447	-0.160	0.088	0.026	0.347	0.225	5.1750	1.5645	-0.3451
Thr	Hγ2-2	1.059	-0.64531	51.6	-0.059	-0.088	0.086	-0.042	0.130	-0.203	-0.274	0.479	0.097	-0.319	0.130	0.074	-0.178	6.0406	-1.6553	-0.3980
Thr	Hγ2-3	1.041	-0.63754	50.8	-0.041	-0.006	0.002	0.128	0.128	-0.203	-0.268	0.472	-0.203	-0.006	-0.012	-0.267	-0.021	4.7260	-0.2510	-3.0959

		$N(\Omega)$	$E(\Omega)$	v(Ω)	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\lvert\mathbf{Q}(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
Thr	SUM	64.002	-435.89644	996.1	-0.002															
Thr	SCF	64	-435.89702		0	0.515	-0.904	-0.479	1.146											
Met	C	4.236	-36.70715	30.9	1.764	-0.387	-0.838	0.409	1.009	-0.653	-0.873	1.530	-0.086	0.001	1.126	-0.653	-0.016	-3.7539	-1.2964	0.2469
Met	O1	9.348	-75.67070	134.2	-1.348	0.160	-0.241	0.508	0.584	0.062	0.447	0.555	-0.127	0.170	0.270	-0.084	-0.309	-3.1976	-2.2835	2.1885
Met	O2	9.302	-75.61042	120.6	-1.302	-0.119	0.077	-0.272	0.306	0.842	0.019	0.984	-0.165	0.707	0.257	0.337	-0.273	-5.5171	-2.2426	-1.2798
Met	H	0.351	-0.32669	18.6	0.649	0.053	0.117	-0.060	0.142	0.027	0.038	0.065	0.020	-0.026	0.020	-0.034	0.035	-6.1978	-3.7330	-0.5330
Met	N	8.147	-54.88401	114.4	-1.147	-0.064	-0.143	0.062	0.168	1.159	1.358	2.519	0.828	-1.087	0.543	-1.628	1.138	-4.4200	3.2581	-0.1083
Met	H1	0.628	-0.47894	30.9	0.372	0.167	0.029	0.064	0.181	-0.032	-0.075	0.110	0.051	0.002	0.084	-0.033	0.010	-6.1294	3.0252	-0.8856
Met	H2	0.633	-0.47987	31.6	0.367	0.029	-0.006	-0.183	0.186	-0.067	-0.036	0.105	-0.055	-0.013	0.008	-0.036	0.042	-4.6338	3.4389	1.7649
Met	Cα	5.431	-37.49333	41.3	0.569	-0.429	0.244	0.128	0.510	0.791	-0.090	0.866	-0.148	-0.314	-0.309	0.665	0.057	-2.7091	1.1995	-0.7426
Met	Hα	1.011	-0.64527	46.6	-0.011	-0.008	0.014	0.122	0.123	-0.168	-0.209	0.377	-0.175	0.016	-0.013	-0.199	0.030	-2.6457	1.0778	-2.7854
Met	Cβ	5.838	-37.72740	52.2	0.162	-0.031	-0.040	-0.033	0.061	-0.116	-0.236	0.359	0.294	-0.147	0.078	-0.091	0.026	-0.0723	1.7734	0.2806
Met	Hβ-1	1.014	-0.63537	48.0	-0.014	-0.040	-0.118	0.036	0.130	-0.126	-0.274	0.409	-0.100	0.104	-0.042	0.299	-0.207	0.4187	3.6503	-0.3764
Met	Hβ-2	1.017	-0.63595	48.0	-0.017	0.003	-0.022	-0.127	0.129	-0.228	-0.173	0.402	-0.171	0.002	-0.029	-0.227	-0.020	-0.1299	1.8302	2.3287
Met	Cγ	5.998	-37.74285	57.4	0.002	-0.086	-0.037	-0.052	0.107	-0.007	-0.093	0.111	-0.065	-0.034	-0.025	-0.024	-0.033	1.9285	-0.1254	-0.5771
Met	Hγ-1	1.035	-0.64472	49.4	-0.035	0.017	0.027	0.133	0.136	-0.266	-0.142	0.414	-0.149	-0.032	-0.017	-0.257	0.024	1.9392	-0.2745	-2.6232
Met	Hγ-2	1.005	-0.63501	46.3	-0.005	0.027	0.123	-0.023	0.128	-0.120	-0.282	0.413	-0.092	0.102	-0.083	0.240	-0.258	1.5468	-1.9769	0.2048
Met	Sδ	16.001	-397.67300	197.4	-0.001	-0.239	-0.753	-0.432	0.900	5.238	1.227	6.867	4.634	-1.771	0.835	-0.023	3.187	5.0378	0.8912	0.4689
Met	Cϵ	5.988	-37.71882	70.0	0.012	0.068	-0.085	-0.032	0.114	-0.087	0.027	0.089	-0.066	0.049	0.006	0.024	0.015	6.9201	-1.7376	-0.6356
Met	Hϵ-1	1.013	-0.62917	49.5	-0.013	-0.001	0.032	0.132	0.135	-0.196	-0.245	0.441	-0.213	-0.021	0.052	-0.221	0.039	6.8043	-1.9286	-2.6702
Met	Hϵ-2	0.995	-0.62017	48.4	0.005	-0.125	-0.014	-0.027	0.129	-0.206	-0.244	0.451	0.380	0.120	0.164	-0.189	0.049	8.8639	-1.3589	-0.1233
Met	Hϵ-3	1.009	-0.62758	49.2	-0.009	0.026	0.126	-0.042	0.135	-0.195	-0.244	0.439	-0.145	0.158	-0.111	0.238	-0.254	6.3364	-3.4871	0.2515
Met	SUM	80.000	-797.58644	1284.7	-0.000															
Met	SCF	80	-797.58672		0	-0.661	-0.647	-0.695	1.157											
Cys	C	4.224	-36.69783	30.7	1.776	-0.454	-0.830	0.309	0.996	-0.653	-0.865	1.522	-0.070	-0.166	1.102	-0.607	-0.284	-1.9900	-1.1162	0.1494
Cys	O1	9.351	-75.66814	133.9	-1.351	0.142	-0.310	0.471	0.581	0.074	0.445	0.561	-0.047	0.127	0.275	-0.049	-0.371	-1.5105	-2.3610	1.9609
Cys	O2	9.302	-75.61226	120.4	-1.302	-0.116	0.117	-0.262	0.309	0.828	0.017	0.966	-0.033	0.692	0.308	0.271	-0.245	-3.8415	-1.7079	-1.4396
Cys	H	0.349	-0.32538	18.4	0.651	0.062	0.118	-0.045	0.141	0.027	0.037	0.064	0.015	-0.032	0.019	-0.036	0.027	-4.6442	-3.2171	-0.8734
Cys	N	8.152	-54.88638	114.5	-1.152	-0.087	-0.143	0.041	0.172	1.163	1.344	2.510	0.457	-1.468	0.419	-1.592	0.614	-2.2620	3.5038	0.2207
Cys	H1	0.622	-0.47599	30.6	0.378	0.161	0.001	0.080	0.179	-0.033	-0.072	0.107	0.035	-0.016	0.085	-0.032	-0.005	-3.9065	3.5667	-0.7125
Cys	H2	0.634	-0.48039	31.6	0.366	0.044	0.013	-0.180	0.186	-0.066	-0.038	0.106	-0.059	-0.013	-0.007	-0.041	0.024	-2.6296	3.4758	2.0785
Cys	Cα	5.411	-37.49360	40.1	0.589	-0.365	0.260	0.171	0.480	-0.115	0.763	0.823	-0.071	-0.286	-0.238	0.620	0.153	-0.7181	1.3666	-0.5601
Cys	Hα	0.991	-0.63593	45.7	0.009	-0.009	0.000	0.124	0.124	-0.207	-0.166	0.374	-0.170	0.016	-0.033	-0.201	-0.023	-0.5582	1.4306	-2.5970
Cys	Cβ	5.974	-37.74284	57.0	0.026	-0.110	0.070	0.008	0.130	-0.196	-0.097	0.299	0.262	-0.090	0.065	-0.173	-0.036	1.9019	1.6116	0.6133
Cys	Hβ-1	0.978	-0.62075	45.7	0.022	-0.049	-0.116	0.026	0.129	-0.141	-0.253	0.399	-0.077	0.167	-0.049	0.288	-0.150	2.6371	3.4511	0.1142

		$N(\Omega)$	$E(\Omega)$	$v(\Omega)$	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\lvert Q(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
Cys	Hβ-2	0.986	-0.62396	46.1	0.014	0.013	-0.016	-0.129	0.131	-0.173	-0.220	0.394	-0.172	0.011	-0.031	-0.212	-0.068	1.7997	1.5072	2.6516
Cys	Sγ	15.670	-397.46389	200.8	0.330	-0.751	-0.457	0.842	1.218	5.518	1.500	7.391	-1.649	-1.674	-3.780	4.474	-2.919	4.2081	-0.6704	-0.5267
Cys	Hγ	1.348	-0.77761	56.8	-0.348	-0.077	-0.161	0.117	0.214	-0.142	-0.272	0.421	-0.056	0.154	-0.178	0.122	-0.257	3.4593	-2.6349	0.8348
Cys	SUM	63.993	-719.50497	972.1	0.007															
Cys	SCF	64	-719.50331		0	-1.047	0.118	-0.033	1.055											
CysS	Ci	4.208	-36.68498	30.5	1.792	0.491	-0.802	-0.249	0.973	-0.819	-0.645	1.467	-0.756	-0.017	-0.240	-0.568	-0.483	7.6145	-0.9166	-0.0138
CysS	O1i	9.349	-75.66457	135.4	-1.349	0.190	-0.134	-0.530	0.579	-0.490	0.011	0.560	-0.326	-0.228	-0.242	-0.034	0.076	8.3753	-1.4831	-2.0520
CysS	O2i	9.301	-75.61617	120.5	-1.301	-0.068	0.012	0.311	0.319	0.802	0.017	0.936	-0.467	-0.635	-0.098	0.470	-0.154	7.9450	-2.3470	2.0132
CysS	Hi	0.348	-0.32428	18.5	0.652	-0.065	0.119	0.039	0.141	0.027	0.037	0.065	0.007	0.036	0.013	-0.034	-0.027	8.7701	-3.8745	1.5317
CysS	Ni	8.146	-54.88890	113.8	-1.146	0.113	-0.157	-0.019	0.194	1.334	1.082	2.421	0.132	1.726	0.177	-1.230	-0.122	7.4222	3.7066	-0.5788
CysS	H1i	0.635	-0.48209	31.3	0.365	-0.165	-0.025	-0.079	0.185	-0.071	-0.047	0.118	0.032	0.060	0.071	-0.028	0.030	9.0550	4.1023	0.2938
CysS	H2i	0.616	-0.47257	29.9	0.384	-0.042	0.029	0.170	0.177	-0.069	-0.022	0.095	-0.067	-0.008	-0.017	-0.014	0.029	7.8136	3.3899	-2.4051
CysS	Cαi	5.417	-37.48202	41.2	0.583	0.278	0.296	-0.247	0.475	-0.146	-0.671	0.872	-0.288	0.141	-0.322	0.635	-0.378	6.1935	1.5314	0.5415
CysS	Hαi	0.995	-0.63782	45.8	0.005	0.003	-0.019	-0.119	0.121	-0.214	-0.165	0.380	-0.211	-0.013	-0.015	-0.161	0.059	6.1244	1.8005	2.5673
CysS	Cβi	5.988	-37.75746	57.5	0.012	0.080	0.074	-0.024	0.112	0.262	-0.108	0.263	0.152	0.171	-0.003	-0.006	-0.023	3.4942	1.3940	-0.5112
CysS	Hβ-1i	0.981	-0.62508	46.6	0.019	0.059	-0.115	-0.006	0.129	-0.128	-0.260	0.395	-0.031	-0.197	-0.050	0.272	0.095	2.6267	3.2162	-0.2111
CysS	Hβ-2i	0.980	-0.62166	46.6	0.020	-0.004	0.004	0.131	0.132	-0.149	-0.232	0.384	-0.154	0.015	-0.030	-0.191	0.140	3.5467	1.0042	-2.5170
CysS	Sγi	16.039	-397.66287	191.6	-0.039	-0.088	0.427	-0.537	0.691	4.875	0.904	6.217	3.370	3.061	1.805	-1.515	-2.797	1.6311	-1.0621	1.0480
CysS	Ck	4.208	-36.68509	30.3	1.792	-0.490	-0.801	0.251	0.972	-0.827	-0.651	1.482	-0.763	0.021	-0.231	-0.571	0.500	-7.6145	-0.9167	0.0143
CysS	O1k	9.349	-75.66453	135.4	-1.349	-0.190	-0.134	0.530	0.579	-0.490	0.011	0.560	-0.326	0.228	-0.243	-0.034	-0.075	-8.3758	-1.4825	2.0525
CysS	O2k	9.301	-75.61626	120.5	-1.301	0.068	0.012	-0.311	0.319	0.802	0.016	0.935	-0.467	0.635	-0.099	0.471	0.154	-7.9436	-2.3481	-2.0122
CysS	Hk	0.348	-0.32428	18.4	0.652	0.065	0.119	-0.039	0.141	0.027	0.037	0.065	0.007	-0.036	0.013	-0.034	0.027	-8.7682	-3.8757	-1.5304
CysS	Nk	8.146	-54.88900	113.8	-1.146	-0.113	-0.157	0.018	0.194	1.334	1.082	2.420	0.130	-1.728	0.175	-1.227	0.118	-7.4224	3.7064	0.5792
CysS	H1k	0.635	-0.48209	31.3	0.365	0.165	-0.025	0.079	0.185	-0.071	-0.046	0.118	0.032	-0.060	0.071	-0.028	-0.030	-9.0540	4.1039	-0.2948
CysS	H2k	0.616	-0.47253	29.9	0.384	0.043	0.029	-0.169	0.177	-0.069	-0.022	0.095	-0.067	0.008	-0.018	-0.014	-0.029	-7.8158	3.3888	2.4049
CysS	Cαk	5.420	-37.48168	41.6	0.580	-0.279	0.300	0.255	0.483	-0.131	-0.656	0.843	-0.263	-0.132	-0.303	0.643	0.333	-6.1939	1.5314	-0.5415
CysS	Hαk	0.995	-0.63782	45.8	0.005	-0.003	-0.019	0.119	0.121	-0.214	-0.165	0.380	-0.211	0.013	-0.015	-0.161	-0.059	-6.1256	1.8004	-2.5674
CysS	Cβk	5.988	-37.75749	57.5	0.012	-0.080	0.074	0.024	0.112	0.262	-0.109	0.263	0.152	-0.172	-0.002	-0.006	0.021	-3.4942	1.3942	0.5101
CysS	Hβ-1k	0.981	-0.62509	46.7	0.019	-0.059	-0.115	0.006	0.129	-0.128	-0.260	0.395	-0.031	0.197	-0.050	0.272	-0.095	-2.6269	3.2163	0.2091
CysS	Hβ-2k	0.980	-0.62166	46.6	0.020	0.004	0.004	-0.132	0.132	-0.149	-0.232	0.384	-0.154	-0.015	-0.030	-0.191	-0.140	-3.5461	1.0053	2.5162
CysS	Sγk	16.039	-397.66287	191.7	-0.039	0.088	0.427	0.537	0.691	4.874	0.905	6.217	3.369	-3.060	1.806	-1.513	2.797	-1.6313	-1.0623	-1.0486
CysS	SUM	126.009	-1437.84087	1818.7	-0.009															
CysS	SCF	126	-1437.83669		0	-0.000	0.364	-0.001	0.364											
Asn(I)	C	4.209	-36.69673	31.0	1.791	-0.895	-0.302	0.215	0.969	-0.628	-0.882	1.517	-0.495	-0.421	-0.360	0.681	1.023	-3.5510	-0.3601	-0.0779
Asn(I)	O1	9.345	-75.68637	135.0	-1.345	-0.393	0.288	0.324	0.585	0.078	0.484	0.609	0.183	-0.031	-0.350	0.052	0.330	-5.0974	0.6980	1.1624

		$N(\Omega)$	$E(\Omega)$	v(Ω)	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\|\mu(\Omega)\|$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\|Q(\Omega)\|$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
Asn (I)	O2	9.302	-75.63747	119.6	-1.302	0.168	-0.202	-0.184	0.320	0.805	0.007	0.933	0.438	0.499	-0.216	0.125	0.308	-3.9532	-2.5924	-1.1438
Asn(I)	H	0.346	-0.32362	18.3	0.654	0.128	0.043	-0.034	0.140	0.036	0.027	0.063	-0.050	-0.024	0.022	0.027	0.010	-5.6064	-3.1660	-0.7184
Asn(I)	N	8.175	-54.93875	106.0	-1.175	-0.183	-0.044	0.021	0.189	1.043	0.852	1.899	-1.322	-1.148	0.172	0.444	0.002	-0.6909	3.3281	-0.0955
Asn(I)	H1	0.629	-0.47975	30.9	0.371	0.111	-0.093	0.109	0.181	-0.045	-0.061	0.106	-0.028	-0.033	0.050	0.005	-0.065	-1.7650	4.3354	-1.2867
Asn(I)	H2	0.616	-0.47241	29.8	0.384	0.061	-0.038	-0.162	0.177	-0.022	-0.063	0.089	-0.021	0.013	-0.046	-0.042	0.023	-1.3280	3.7155	1.6471
Asn(I)	Cα	5.440	-37.49648	41.8	0.560	-0.107	0.474	0.113	0.499	-0.219	-0.808	1.083	-0.199	0.403	-0.144	0.851	0.318	-0.9074	0.6484	-0.6126
Asn(I)	Hα	0.993	-0.63479	46.2	0.007	-0.016	0.012	0.120	0.122	-0.162	-0.221	0.385	-0.162	0.009	-0.111	-0.181	0.099	-0.5366	0.3366	-2.6033
Asn(I)	Cβ	5.898	-37.78251	56.1	0.102	0.021	0.033	-0.020	0.044	-0.184	-0.316	0.506	0.326	-0.265	0.137	-0.071	-0.122	1.0102	-0.8847	0.9494
Asn(I)	Hβ-1	1.007	-0.62636	48.7	-0.007	0.014	-0.035	-0.122	0.127	-0.244	-0.177	0.423	-0.199	-0.043	-0.032	-0.175	0.163	0.8589	-0.3373	2.9253
Asn(I)	Hβ-2	0.969	-0.61920	43.9	0.031	0.038	0.112	-0.002	0.118	-0.106	-0.290	0.410	-0.081	0.115	0.035	0.363	0.055	0.5881	-2.8794	0.8394
Asn(I)	Cγ	4.219	-36.70250	32.0	1.781	0.794	-0.069	-0.162	0.813	-0.338	-1.089	1.491	-0.359	-0.270	-0.124	1.238	-0.603	3.7791	-0.6311	0.1943
Asn(I)	Oδ1	9.350	-75.63986	139.0	-1.350	0.317	-0.406	0.042	0.517	0.640	-0.048	0.713	0.165	-0.351	-0.067	0.366	-0.179	5.1755	-2.4144	0.3834
Asn(I)	Nδ2	8.433	-55.20805	122.5	-1.433	0.147	0.197	-0.065	0.254	1.929	0.889	2.881	1.588	-0.494	-0.988	0.834	-0.680	4.5515	1.6466	-0.6367
Asn(I)	Hδ2-1	0.514	-0.41908	21.0	0.486	0.077	-0.122	-0.002	0.144	-0.063	-0.074	0.138	0.094	-0.082	-0.011	-0.020	0.005	3.3974	3.1351	-0.6068
Asn(I)	Hδ2-2	0.562	-0.44179	27.7	0.438	-0.160	-0.025	0.032	0.165	-0.039	-0.035	0.074	0.068	0.006	-0.023	-0.037	-0.004	6.3826	1.8731	-1.0068
Asn(I)	SUM	70.007	-489.80573	1049.5	-0.007															
Asn(I)	SCF	70	-489.80640		0	-0.983	0.988	-0.458	1.467											
Asn(II)	C	4.193	-36.68506	30.1	1.807	0.362	-0.884	-0.136	0.965	-0.620	-0.836	1.461	-0.391	0.177	0.888	-0.571	0.282	2.3147	-1.2095	0.0917
Asn(II)	O1	9.359	-75.68688	133.2	-1.359	-0.099	-0.327	-0.451	0.566	0.059	0.436	0.540	-0.173	-0.126	0.285	0.051	0.313	1.9509	-2.5564	-1.6805
Asn(II)	O2	9.293	-75.64204	119.5	-1.293	0.102	0.134	0.289	0.334	0.796	0.017	0.930	-0.346	-0.581	0.367	0.465	0.019	3.6989	-1.8952	2.0488
Asn(II)	H	0.349	-0.32558	18.5	0.651	-0.052	0.129	0.025	0.141	0.026	0.037	0.064	0.018	0.028	0.011	-0.048	-0.018	4.3705	-3.5389	1.7513
Asn(II)	N	8.146	-54.89766	114.5	-1.146	-0.148	0.049	-0.059	0.167	1.281	1.182	2.463	-1.801	0.901	-1.079	0.983	0.370	3.2796	3.1181	-0.6206
Asn(II)	H1	0.626	-0.47776	30.9	0.374	0.035	-0.173	-0.044	0.182	-0.042	-0.069	0.112	-0.049	0.011	-0.007	0.073	0.072	3.0232	4.9002	-0.0427
Asn(II)	H2	0.634	-0.48063	31.6	0.366	0.003	-0.007	0.185	0.186	-0.031	-0.071	0.104	-0.034	0.013	-0.032	-0.057	0.023	3.3612	3.1129	-2.5124
Asn(II)	Cα	5.418	-37.49276	41.5	0.582	0.417	0.169	-0.214	0.498	-0.078	0.665	0.727	0.269	0.331	-0.210	0.235	-0.227	1.2928	1.4697	0.3188
Asn(II)	Hα	0.972	-0.62807	44.7	0.028	0.016	-0.028	-0.114	0.119	-0.209	-0.162	0.372	-0.155	-0.024	-0.069	-0.188	0.100	1.0322	1.8761	2.3038
Asn(II)	Cβ	5.896	-37.79627	54.3	0.104	-0.044	-0.021	0.035	0.060	0.321	0.009	0.376	0.106	0.175	-0.239	-0.116	0.078	-1.2354	1.7513	-1.0792
Asn(II)	Hβ-1	0.999	-0.62768	47.6	0.001	0.027	-0.124	0.011	0.128	-0.159	-0.250	0.413	-0.116	-0.155	-0.026	0.361	0.039	-1.7862	3.7208	-0.9940
Asn(II)	Hβ-2	1.019	-0.63834	46.7	-0.019	-0.006	0.014	0.123	0.124	-0.196	-0.216	0.412	-0.187	-0.021	-0.078	-0.143	0.196	-0.9979	1.2268	-3.0438
Asn(II)	Cγ	4.253	-36.72908	32.1	1.747	-0.643	-0.322	0.446	0.846	-0.390	-1.055	1.496	-0.685	0.184	-0.288	0.263	1.097	-3.3508	0.2544	0.1612
Asn(II)	Oδ1	9.349	-75.64027	138.6	-1.349	-0.226	0.167	0.439	0.522	-0.057	-0.556	0.677	-0.145	0.086	-0.405	-0.138	0.335	-4.3162	0.9694	2.0896
Asn(II)	Nδ2	8.381	-55.17116	120.1	-1.381	-0.061	-0.245	-0.071	0.262	0.752	1.938	2.776	-0.053	0.993	-2.028	0.217	0.802	-4.0755	-1.8824	-1.0405
Asn(II)	Hδ2-1	0.551	-0.43894	25.8	0.449	-0.090	0.070	0.108	0.157	-0.055	-0.037	0.092	0.045	-0.031	-0.053	-0.041	0.026	-2.9452	-2.6760	-2.3187
Asn(II)	Hδ2-2	0.564	-0.44307	27.9	0.436	0.102	0.104	-0.078	0.165	-0.043	-0.033	0.077	0.018	0.046	-0.037	-0.009	-0.026	-5.2827	-3.0098	-0.1348
Asn(II)	SUM	70.004	-489.80125	1057.7	-0.004															
Asn(II)	SCF	70	-489.80188		0	0.534	-0.169	-0.793	0.971											

		$N(\Omega)$	$E(\Omega)$	$v(\Omega)$	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\lvert\mathbf{Q}(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
Gln	C	4.245	-36.72047	31.1	1.755	-0.352	-0.846	0.426	1.010	-0.716	-0.816	1.534	-0.660	-0.266	-0.416	0.086	1.064	-3.5878	-1.3245	0.1169
Gln	O1	9.350	-75.68353	136.5	-1.350	-0.003	-0.533	-0.230	0.580	-0.460	-0.045	0.559	-0.342	0.096	-0.161	0.382	0.258	-3.6295	-3.4082	-0.7328
Gln	O2	9.306	-75.63454	119.3	-1.306	-0.023	0.258	0.172	0.311	-0.861	0.016	0.985	-0.573	0.329	-0.534	0.322	-0.295	-4.4692	-0.7985	2.4118
Gln	H	0.349	-0.32568	18.4	0.651	0.047	0.120	-0.058	0.141	0.027	0.038	0.065	0.018	-0.027	0.009	-0.039	0.035	-5.0763	-2.3153	3.1671
Gln	N	8.156	-54.91227	113.3	-1.156	-0.096	-0.120	0.103	0.186	1.093	1.349	2.447	0.707	-1.071	0.803	-1.052	1.354	-4.0468	3.2264	-1.0294
Gln	H1	0.641	-0.48493	31.6	0.359	0.173	0.035	0.061	0.187	-0.048	-0.071	0.119	0.059	-0.007	0.087	-0.050	0.004	-5.7692	2.9710	-1.7709
Gln	H2	0.624	-0.47737	30.2	0.376	0.027	-0.049	-0.170	0.179	-0.068	-0.026	0.097	-0.068	-0.001	-0.003	-0.020	0.026	-4.2666	3.7635	0.7719
Gln	Cα	5.419	-37.49100	41.2	0.581	-0.403	0.292	0.088	0.505	0.982	-0.159	1.055	-0.350	-0.427	-0.296	0.842	-0.168	-2.5023	0.9760	-1.2306
Gln	Hα	1.006	-0.64336	46.4	-0.006	-0.003	0.030	0.120	0.123	-0.207	-0.167	0.375	-0.204	-0.010	-0.038	-0.131	0.134	-2.3994	0.4519	-3.2058
Gln	Cβ	5.844	-37.73507	52.3	0.156	-0.014	-0.034	-0.027	0.046	-0.056	-0.349	0.440	0.332	-0.137	0.173	-0.081	0.080	0.1696	1.5572	-0.2542
Gln	Hβ-1	1.002	-0.63157	47.2	-0.002	-0.046	-0.101	0.063	0.128	-0.132	-0.262	0.401	-0.086	0.134	-0.065	0.197	-0.263	0.7970	3.2455	-1.2255
Gln	Hβ-2	1.006	-0.63345	46.5	-0.006	0.022	-0.028	-0.119	0.124	-0.127	-0.268	0.403	-0.127	0.004	0.005	-0.237	0.138	0.0982	2.0197	1.7393
Gln	Cγ	5.902	-37.77296	56.5	0.098	0.093	0.000	0.019	0.095	-0.247	-0.037	0.308	0.231	-0.024	0.137	-0.153	0.102	2.0570	-0.5826	-0.6915
Gln	Hγ-1	1.034	-0.64203	49.2	-0.034	-0.010	0.034	0.122	0.127	-0.138	-0.283	0.429	-0.153	-0.050	-0.017	-0.233	0.148	2.1540	-1.0597	-2.6862
Gln	Hγ-2	1.011	-0.63129	47.2	-0.011	0.023	0.103	-0.060	0.121	-0.153	-0.282	0.441	-0.069	0.188	-0.088	0.201	-0.268	1.4799	-2.2996	0.2777
Gln	Cδ	4.247	-36.72152	32.2	1.753	0.739	0.267	0.353	0.861	-0.410	-1.033	1.487	-0.197	-0.498	-0.330	0.361	1.097	4.6845	0.0693	0.2619
Gln	Oϵ1	9.348	-75.64799	136.3	-1.348	0.083	0.391	0.343	0.527	0.070	0.547	0.675	0.051	-0.051	0.115	0.063	0.563	5.0807	1.7294	1.7568
Gln	Nϵ2	8.388	-55.16814	121.8	-1.388	0.199	-0.151	-0.077	0.261	0.838	2.047	2.967	1.172	0.015	0.910	-0.189	2.142	6.5692	-1.3859	-0.6811
Gln	Hϵ2-1	0.574	-0.45016	28.2	0.426	0.029	0.133	0.098	0.167	-0.035	-0.043	0.079	-0.032	0.015	0.017	0.024	0.057	6.2380	-2.8529	-1.8056
Gln	Hϵ2-2	0.557	-0.43902	27.4	0.443	-0.153	-0.015	-0.058	0.164	-0.038	-0.034	0.072	0.052	0.019	0.037	-0.034	0.007	8.3053	-1.1623	0.0100
Gln	SUM	78.009	-528.84634	1213.0	-0.009															
Gln	SCF	78	-528.84750		0	-0.255	-0.894	-0.339	0.989											
$Asp^{(-)}$	C	4.274	-36.73491	31.3	1.726	-0.964	-0.349	0.293	1.066	-0.687	-0.825	1.515	-0.718	-0.178	-0.124	0.055	1.093	-3.4701	-0.2581	0.0002
$Asp^{(-)}$	O1	9.370	-75.67308	137.1	-1.370	-0.286	0.183	0.430	0.548	0.069	0.500	0.621	0.032	0.121	-0.342	-0.153	0.372	-4.7430	0.3741	1.7530
$Asp^{(-)}$	O2	9.296	-75.61083	121.3	-1.296	0.112	-0.123	-0.243	0.295	0.887	0.016	1.034	0.508	0.649	-0.268	-0.238	0.340	-4.2827	-2.0354	-1.6050
$Asp^{(-)}$	H	0.370	-0.33886	19.4	0.630	0.135	0.049	-0.050	0.152	0.027	0.044	0.072	-0.050	-0.028	0.029	0.024	0.019	-5.8798	-2.6120	-1.0151
$Asp^{(-)}$	N	8.177	-54.90797	116.7	-1.177	0.127	-0.183	0.142	0.264	1.021	1.459	2.493	-0.311	1.286	-1.261	-0.202	1.102	-0.6623	3.3277	0.1032
$Asp^{(-)}$	H1	0.575	-0.45327	24.6	0.425	-0.147	-0.060	0.019	0.160	-0.056	-0.105	0.164	0.125	-0.034	-0.084	-0.056	-0.003	1.1186	3.8520	-0.2856
$Asp^{(-)}$	H2	0.658	-0.49244	32.4	0.342	0.008	-0.017	-0.194	0.195	-0.033	-0.079	0.115	-0.036	0.005	-0.078	-0.035	0.012	-0.9271	3.5093	1.9712
$Asp^{(-)}$	Cα	5.415	-37.49805	40.3	0.585	-0.201	0.460	0.135	0.520	0.878	0.025	1.029	0.025	0.458	-0.269	0.592	0.381	-0.8466	0.6855	-0.6117
$Asp^{(-)}$	Hα	0.998	-0.63982	45.2	0.002	0.005	0.010	0.128	0.128	-0.220	-0.134	0.358	-0.136	0.029	-0.090	-0.200	0.037	-0.5932	0.5694	-2.6353
$Asp^{(-)}$	Cβ	5.932	-37.79200	56.6	0.068	-0.157	-0.029	0.025	0.162	-0.515	-0.158	0.704	-0.129	-0.444	0.334	0.151	-0.207	1.0853	-1.0814	0.6556
$Asp^{(-)}$	Hβ-1	1.054	-0.64539	51.1	-0.054	0.034	-0.002	-0.136	0.140	-0.278	-0.140	0.426	-0.155	-0.050	-0.044	-0.257	0.034	0.8671	-0.9999	2.6985
$Asp^{(-)}$	Hβ-2	1.050	-0.64542	50.2	-0.050	0.044	0.124	0.028	0.134	-0.129	-0.303	0.443	-0.121	0.081	0.068	0.351	0.203	0.7396	-3.0311	0.1069
$Asp^{(-)}$	Cγ	3.962	-36.49601	28.6	2.038	0.632	0.113	-0.129	0.655	-0.221	-1.209	1.540	-0.115	-0.352	0.214	0.370	-1.225	3.8896	-0.4771	0.0156

		$N(\Omega)$	$E(\Omega)$	v(Ω)	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\lvert\mathbf{Q}(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
$Asp^{(-)}$	Oδ1	9.438	-75.58061	139.6	-1.438	0.036	0.289	-0.221	0.366	0.206	0.679	0.927	0.165	-0.102	-0.186	0.171	-0.717	4.2661	1.3246	-1.4310
$Asp^{(-)}$	Oδ2	9.436	-75.57740	146.9	-1.436	0.251	-0.226	0.164	0.375	-0.131	-0.759	0.961	0.400	-0.431	0.277	-0.052	-0.538	5.4297	-1.8878	1.0429
$Asp^{(-)}$	SUM	70.003	-509.08606	1041.5	-1.003															
$Asp^{(-)}$	SCF	70	-509.08686		-1															
$Glu^{(-)}$	C	4.263	-36.72997	30.9	1.737	-0.447	-0.783	0.544	1.053	-0.711	-0.827	1.539	-0.540	-0.338	-0.508	-0.047	1.042	-3.6107	-1.0887	0.3075
$Glu^{(-)}$	O1	9.369	-75.67363	137.4	-1.369	-0.237	0.044	0.496	0.551	0.027	-0.542	0.611	-0.322	0.052	-0.418	0.019	0.069	-4.6396	-0.9976	2.3140
$Glu^{(-)}$	O2	9.297	-75.62521	119.3	-1.297	0.121	-0.087	-0.268	0.306	0.835	0.004	0.966	-0.553	0.367	-0.414	0.650	-0.157	-3.4632	-3.2398	-0.9898
$Glu^{(-)}$	H	0.363	-0.33449	19.1	0.637	0.055	0.109	-0.084	0.148	0.027	0.042	0.070	0.013	-0.030	0.019	-0.016	0.047	-4.1250	-4.5626	0.0337
$Glu^{(-)}$	N	8.153	-54.90162	113.7	-1.153	-0.109	-0.135	0.143	0.225	1.453	0.973	2.442	1.022	-0.746	0.762	-1.107	1.482	-3.8335	3.4519	-0.6786
$Glu^{(-)}$	H1	0.667	-0.49827	33.0	0.333	0.190	0.031	0.034	0.196	-0.041	-0.073	0.116	0.084	0.007	0.068	-0.041	0.000	-5.6608	3.2061	-1.1123
$Glu^{(-)}$	H2	0.635	-0.48207	31.3	0.365	0.003	-0.050	-0.180	0.186	-0.060	-0.029	0.091	-0.051	0.003	0.034	-0.021	0.032	-3.7680	3.9772	1.1396
$Glu^{(-)}$	Cα	5.434	-37.49331	41.1	0.566	-0.489	0.281	0.092	0.571	1.080	0.002	1.248	-0.621	-0.396	-0.507	0.987	-0.087	-2.3943	1.1413	-1.0388
$Glu^{(-)}$	Hα	1.027	-0.65111	47.2	-0.027	0.015	0.026	0.124	0.127	-0.224	-0.154	0.380	-0.223	-0.004	-0.030	-0.125	0.120	-2.3675	0.7057	-3.0401
$Glu^{(-)}$	Cβ	5.848	-37.73269	51.7	0.152	-0.137	0.062	0.001	0.150	-0.406	-0.240	0.652	0.536	0.113	0.271	-0.389	0.056	0.3379	1.5346	-0.1121
$Glu^{(-)}$	Hβ-1	1.004	-0.63286	46.4	-0.004	-0.025	-0.111	0.070	0.133	-0.099	-0.278	0.392	-0.013	0.170	-0.077	0.148	-0.245	1.0516	3.1945	-1.0686
$Glu^{(-)}$	Hβ-2	1.045	-0.64738	49.4	-0.045	0.028	-0.034	-0.128	0.136	-0.131	-0.261	0.399	-0.133	-0.011	0.016	-0.230	0.134	0.2935	1.9989	1.8871
$Glu^{(-)}$	Cγ	5.904	-37.76875	55.2	0.096	-0.164	-0.034	-0.043	0.173	-0.512	-0.148	0.693	-0.231	-0.459	-0.163	0.395	0.068	2.1276	-0.6827	-0.5604
$Glu^{(-)}$	Hγ-1	1.068	-0.65238	51.6	-0.068	0.026	0.035	0.134	0.141	-0.140	-0.277	0.424	-0.153	-0.040	0.007	-0.218	0.170	2.0900	-1.2457	-2.5386
$Glu^{(-)}$	Hγ-2	1.069	-0.65705	49.9	-0.069	0.058	0.098	-0.066	0.132	-0.123	-0.300	0.436	-0.061	0.155	-0.081	0.171	-0.299	1.5492	-2.3425	0.5020
$Glu^{(-)}$	Cδ	3.965	-36.49827	28.8	2.035	0.630	0.102	0.149	0.655	-0.216	-1.214	1.542	-0.184	-0.361	0.091	1.191	0.640	4.9290	-0.1462	0.1405
$Glu^{(-)}$	Oϵ1	9.437	-75.57467	144.5	-1.437	0.065	0.341	0.113	0.365	0.121	0.760	0.955	0.125	0.091	0.225	0.581	0.446	5.4325	2.0124	0.8744
$Glu^{(-)}$	Oϵ2	9.444	-75.56764	147.7	-1.444	0.217	-0.276	-0.040	0.353	0.919	-0.141	0.989	0.319	-0.553	0.041	0.385	0.232	6.3497	-1.9806	-0.1327
$Glu^{(-)}$	SUM	77.993	-548.12138	1198.3	-0.993															
$Glu^{(-)}$	SCF	78	-548.11975		-1															
$Lys^{(+)}$	C	4.133	-36.62343	29.5	1.867	0.282	-0.923	-0.299	1.010	-0.684	-0.792	1.477	0.676	0.341	1.029	-0.615	0.208	4.1333	-1.4048	-0.1412
$Lys^{(+)}$	O1	9.411	-75.62611	133.8	-1.411	-0.336	-0.414	-0.411	0.673	0.179	0.286	0.470	0.168	-0.084	0.232	0.047	0.257	3.0232	-2.8153	-1.5009
$Lys^{(+)}$	O2	9.346	-75.56416	120.6	-1.346	0.247	0.150	0.226	0.367	0.796	0.017	0.930	0.069	-0.503	0.292	0.443	0.279	6.3323	-1.9801	0.9048
$Lys^{(+)}$	H	0.327	-0.30675	16.3	0.673	-0.040	0.120	0.038	0.132	0.037	0.032	0.069	0.028	0.027	0.011	-0.051	-0.028	6.8549	-3.6186	0.3674
$Lys^{(+)}$	N	8.190	-54.88362	115.4	-1.190	0.003	-0.175	-0.022	0.176	1.355	1.128	2.487	0.836	1.200	0.253	-1.753	-0.909	5.0098	3.1757	-0.3223
$Lys^{(+)}$	H1	0.578	-0.45110	27.9	0.422	-0.163	0.017	-0.067	0.177	-0.014	-0.060	0.078	0.035	-0.008	0.060	-0.012	-0.007	6.7274	2.9978	0.4501
$Lys^{(+)}$	H2	0.607	-0.46457	30.1	0.393	-0.024	0.014	0.187	0.189	-0.054	-0.032	0.087	-0.044	0.011	0.002	-0.034	-0.032	5.2063	3.1316	-2.2060
$Lys^{(+)}$	Cα	5.406	-37.53116	41.1	0.594	0.374	0.220	-0.137	0.455	-0.149	-0.563	0.751	-0.087	0.309	-0.243	0.525	-0.174	3.2755	1.2536	0.5831
$Lys^{(+)}$	Hα	1.033	-0.65934	47.5	-0.033	0.006	0.006	-0.106	0.107	-0.167	-0.202	0.370	-0.168	-0.006	0.012	-0.199	0.038	3.2977	1.3471	2.6320
$Lys^{(+)}$	Cβ	5.797	-37.74971	51.5	0.203	-0.002	0.003	-0.017	0.018	-0.166	-0.091	0.261	0.155	0.160	-0.078	-0.040	-0.000	0.6153	1.8222	-0.3871
$Lys^{(+)}$	Hβ-1	1.045	-0.65564	48.5	-0.045	0.040	-0.106	-0.012	0.114	-0.121	-0.276	0.406	-0.115	-0.058	-0.025	0.366	0.121	0.2750	3.8139	-0.0444

		$N(\Omega)$	$E(\Omega)$	v(Ω)	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\mathcal{Q}_{xx}(\Omega)$	$\mathcal{Q}_{yy}(\Omega)$	$\lvert\mathbf{Q}(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
$Lys^{(+)}$	Hβ-2	1.058	-0.66034	48.6	-0.058	0.011	-0.004	0.112	0.113	-0.159	-0.230	0.391	-0.159	-0.000	0.006	-0.203	0.126	0.5750	1.5351	-2.4188
$Lys^{(+)}$	Cγ	5.805	-37.73418	53.0	0.195	-0.117	-0.022	0.019	0.121	-0.312	-0.009	0.365	0.167	-0.133	0.209	-0.184	0.090	-1.4652	0.2421	0.8552
$Lys^{(+)}$	Hγ-1	1.035	-0.65529	45.1	-0.035	0.007	0.103	-0.013	0.104	-0.145	-0.296	0.450	-0.080	-0.184	-0.063	0.350	0.115	-1.0419	-1.7511	0.6577
$Lys^{(+)}$	Hγ-2	1.076	-0.66643	51.0	-0.076	0.014	-0.013	-0.112	0.114	-0.154	-0.266	0.425	-0.158	0.019	-0.018	-0.225	0.151	-1.5429	0.6621	2.8664
$Lys^{(+)}$	Cδ	5.808	-37.74580	53.5	0.192	-0.193	-0.033	-0.018	0.197	-0.559	-0.041	0.670	-0.078	0.363	-0.423	0.152	-0.096	-4.0499	0.7784	-0.3385
$Lys^{(+)}$	Hδ-1	1.055	-0.65768	49.5	-0.055	0.023	0.018	0.110	0.114	-0.145	-0.266	0.417	-0.147	0.016	-0.063	-0.229	0.136	-3.9588	0.3735	-2.3511
$Lys^{(+)}$	Hδ-2	1.060	-0.65894	50.1	-0.060	0.049	-0.102	-0.003	0.113	-0.158	-0.261	0.424	-0.140	-0.101	-0.019	0.391	0.079	-4.5199	2.7687	-0.1313
$Lys^{(+)}$	Cϵ	5.577	-37.59755	53.2	0.423	-0.541	-0.012	-0.138	0.558	-0.900	-0.452	1.376	0.764	-0.468	0.852	-0.575	-0.009	-6.1014	-0.8334	0.8699
$Lys^{(+)}$	Hϵ-1	0.980	-0.63239	45.1	0.020	-0.003	0.102	0.011	0.103	-0.174	-0.229	0.404	-0.147	-0.119	-0.015	0.369	0.064	-5.7439	-2.8321	0.6567
$Lys^{(+)}$	Hϵ-2	0.985	-0.63442	45.5	0.015	0.027	-0.019	-0.098	0.103	-0.178	-0.229	0.408	-0.175	-0.001	-0.059	-0.196	0.134	-6.3715	-0.4082	2.8485
$Lys^{(+)}$	Nζ	8.244	-55.06954	95.6	-1.244	0.009	-0.001	0.004	0.009	0.196	0.132	0.331	-0.177	0.104	-0.211	0.139	0.026	-8.6276	-0.3412	-0.4000
$Lys^{(+)}$	Hζ-1	0.482	-0.39470	23.6	0.518	0.042	-0.144	-0.009	0.150	-0.018	-0.029	0.048	-0.017	-0.013	0.001	0.043	0.012	-9.1328	1.4953	-0.2554
$Lys^{(+)}$	Hζ-2	0.481	-0.39394	23.5	0.519	-0.003	0.031	0.146	0.150	-0.018	-0.029	0.048	-0.017	-0.003	-0.008	-0.023	0.020	-8.5578	-0.7584	-2.2630
$Lys^{(+)}$	Hζ-3	0.480	-0.39400	23.0	0.520	0.111	0.079	-0.057	0.148	-0.023	-0.029	0.052	0.010	0.030	-0.022	-0.001	-0.023	-10.0325	-1.3838	0.3673
$Lys^{(+)}$	SUM	79.998	-494.41080	1322.4	1.002															
$Lys^{(+)}$	SCF	80	-494.41030		+1															
$Arg^{(+)}$	C	4.223	-36.71091	30.3	1.777	0.238	0.762	0.575	0.984	-0.835	-0.669	1.506	-0.742	-0.229	0.384	-0.026	-0.965	5.5361	1.3779	0.4907
$Arg^{(+)}$	O1	9.342	-75.69355	134.6	-1.342	0.175	-0.022	0.565	0.592	0.012	-0.494	0.564	-0.413	0.018	0.269	0.012	0.005	6.1475	1.3492	2.6515
$Arg^{(+)}$	O2	9.305	-75.63754	120.5	-1.305	-0.074	0.080	-0.293	0.312	0.793	0.015	0.925	-0.699	0.231	0.260	0.717	0.135	5.3244	3.4974	-0.8364
$Arg^{(+)}$	H	0.344	-0.32178	18.2	0.656	-0.037	-0.107	-0.080	0.139	0.027	0.036	0.063	0.021	-0.018	-0.013	-0.019	-0.046	5.8177	4.8947	0.1914
$Arg^{(+)}$	N	8.152	-54.91269	114.2	-1.152	0.064	0.153	0.069	0.180	1.354	1.084	2.443	0.920	-1.103	-0.491	-1.482	-1.157	6.6641	-3.0033	-0.6584
$Arg^{(+)}$	H1	0.623	-0.47641	30.7	0.377	-0.158	-0.046	0.073	0.180	-0.047	-0.069	0.116	0.054	-0.008	-0.083	-0.056	-0.007	8.3063	-2.6459	-1.5288
$Arg^{(+)}$	H2	0.611	-0.47017	29.6	0.389	-0.038	0.022	-0.170	0.175	-0.067	-0.023	0.094	-0.063	-0.011	0.016	-0.024	-0.010	7.0359	-3.2710	1.1798
$Arg^{(+)}$	Cα	5.411	-37.49316	41.0	0.589	0.362	-0.271	0.144	0.475	0.820	-0.161	0.869	-0.177	-0.493	0.219	0.576	0.117	4.8618	-0.9829	-1.0208
$Arg^{(+)}$	Hα	1.003	-0.64207	46.0	-0.003	0.004	-0.027	0.116	0.120	-0.215	-0.170	0.385	-0.214	-0.006	0.005	-0.133	-0.140	4.8347	-0.4917	-3.0065
$Arg^{(+)}$	Cβ	5.859	-37.74281	53.1	0.141	-0.023	0.030	-0.007	0.038	-0.238	-0.193	0.432	0.260	-0.288	0.016	-0.062	-0.029	2.2370	-1.9404	-0.2337
$Arg^{(+)}$	Hβ-1	1.011	-0.63469	47.7	-0.011	0.046	0.105	0.061	0.130	-0.131	-0.276	0.415	-0.122	0.055	0.044	0.198	0.308	1.8987	-3.6466	-1.3167
$Arg^{(+)}$	Hβ-2	1.024	-0.63913	48.6	-0.024	0.012	0.038	-0.120	0.127	-0.155	-0.256	0.416	-0.153	-0.016	0.038	-0.217	-0.157	2.3117	-2.4819	1.7453
$Arg^{(+)}$	Cγ	5.865	-37.73394	54.6	0.135	-0.164	-0.023	0.033	0.169	0.380	-0.045	0.415	0.255	0.244	-0.118	-0.177	0.067	0.0879	-0.0489	-0.6559
$Arg^{(+)}$	Hγ-1	1.031	-0.64170	49.3	-0.031	0.017	-0.044	0.117	0.126	-0.162	-0.263	0.429	-0.163	-0.008	-0.014	-0.211	-0.180	0.0499	0.5379	-2.6243
$Arg^{(+)}$	Hγ-2	1.032	-0.64433	47.8	-0.032	-0.000	-0.103	-0.063	0.121	-0.145	-0.285	0.437	-0.115	0.119	0.060	0.186	0.315	0.3728	1.6361	0.4817
$Arg^{(+)}$	Cδ	5.519	-37.51277	49.0	0.481	-0.506	0.328	-0.055	0.606	-0.922	-0.378	1.337	0.253	-1.061	0.302	0.144	-0.053	-2.4543	-1.2180	0.0373
$Arg^{(+)}$	Hδ-1	1.002	-0.63701	46.9	-0.002	0.015	0.043	-0.114	0.123	-0.146	-0.253	0.404	-0.146	-0.006	0.009	-0.209	-0.163	-2.4542	-1.7732	2.0090
$Arg^{(+)}$	Hδ-2	1.003	-0.63729	47.1	-0.003	0.034	0.104	0.057	0.123	-0.145	-0.257	0.407	-0.132	0.072	0.043	0.167	0.307	-2.8115	-2.8682	-1.1225
$Arg^{(+)}$	Nϵ	8.462	-55.30531	98.4	-1.462	0.214	0.089	-0.060	0.239	0.747	1.320	2.093	0.822	0.096	-0.402	0.805	-1.062	-4.4881	0.6339	-0.3767
$Arg^{(+)}$	Hϵ	0.542	-0.43522	26.1	0.458	-0.034	-0.142	0.057	0.157	-0.014	-0.048	0.066	0.002	0.028	-0.016	0.032	-0.033	-3.9850	2.3170	-1.0603

		$N(\Omega)$	$E(\Omega)$	v(Ω)	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\mathcal{Q}_{xx}(\Omega)$	$\mathcal{Q}_{yy}(\Omega)$	$\lvert\mathbf{Q}(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
$Arg^{(+)}$	Cζ	3.815	-36.50623	23.6	2.185	-0.009	-0.006	0.004	0.011	0.679	0.713	1.392	0.673	-0.068	-0.237	0.456	-0.651	-6.9075	0.2551	0.0691
$Arg^{(+)}$	Nη1	8.391	-55.19863	117.3	-1.391	-0.082	-0.220	0.090	0.252	1.527	0.924	2.475	1.425	-0.306	-0.406	0.595	-1.051	-7.7129	-1.9419	0.9612
$Arg^{(+)}$	Hη1-1	0.509	-0.41561	23.7	0.491	-0.091	0.110	-0.027	0.145	-0.022	-0.043	0.067	0.027	-0.043	0.007	0.012	-0.014	-6.5281	-3.3600	1.3126
$Arg^{(+)}$	Hη1-2	0.516	-0.41860	24.9	0.484	0.144	0.021	-0.027	0.148	-0.034	-0.035	0.068	0.058	0.022	-0.021	-0.028	-0.005	-9.5360	-2.2484	1.3187
$Arg^{(+)}$	Nη2	8.382	-55.18554	118.2	-1.382	-0.162	0.179	-0.043	0.245	0.989	1.488	2.494	1.214	0.088	-0.489	0.789	-1.168	-8.5762	2.0792	-0.3734
$Arg^{(+)}$	Hη2-1	0.517	-0.41925	25.0	0.483	-0.045	-0.131	0.052	0.148	-0.035	-0.033	0.068	-0.031	0.019	-0.007	0.052	-0.032	-8.0463	3.7617	-1.0315
$Arg^{(+)}$	Hη2-2	0.517	-0.41892	25.0	0.483	0.144	0.022	-0.026	0.148	-0.035	-0.033	0.068	0.066	0.006	-0.014	-0.034	-0.000	-10.4159	1.8337	-0.0565
$Arg^{(+)}$	SUM	94.012	-603.48523	1491.4	0.988															
$Arg^{(+)}$	SCF	94	-603.48685		+1															
His(I)	C	4.238	-36.71270	31.4	1.762	-0.265	-0.975	0.162	1.024	-0.686	-0.843	1.531	-0.006	-0.235	1.102	-0.658	-0.229	-3.2086	-1.3728	-0.1159
His(I)	O1	9.361	-75.69418	130.6	-1.361	0.208	-0.333	0.414	0.570	0.079	0.438	0.557	0.018	0.060	0.382	0.024	-0.286	-2.4582	-2.7515	1.5037
His(I)	O2	9.303	-75.63858	119.7	-1.303	-0.145	0.129	-0.252	0.318	0.823	0.015	0.959	-0.324	0.478	0.436	0.600	-0.033	-4.8736	-2.1245	-1.8330
His(I)	H	0.347	-0.32437	18.3	0.653	0.034	0.134	-0.025	0.140	0.026	0.037	0.064	0.026	-0.019	0.010	-0.055	0.019	-5.3116	-3.8435	-1.5238
His(I)	N	8.150	-54.90169	113.9	-1.150	-0.047	-0.156	0.023	0.165	1.163	1.334	2.498	1.018	-0.797	0.177	-2.252	0.446	-4.6564	3.0097	0.2233
His(I)	H1	0.620	-0.47589	30.2	0.380	0.144	0.036	0.097	0.178	-0.028	-0.076	0.107	0.017	0.002	0.086	-0.030	0.021	-6.1524	2.6912	-0.8894
His(I)	H2	0.637	-0.48188	31.8	0.363	0.058	0.037	-0.174	0.187	-0.070	-0.039	0.110	-0.044	-0.018	-0.028	-0.057	0.016	-5.1922	2.7763	2.0251
His(I)	Cα	5.425	-37.49889	41.4	0.575	-0.477	0.130	0.109	0.506	-0.104	0.819	0.891	0.109	-0.401	-0.249	0.530	0.151	-2.5225	1.4067	-0.4337
His(I)	Hα	0.992	-0.63543	46.5	0.008	-0.025	-0.010	0.121	0.124	-0.207	-0.169	0.377	-0.163	0.035	-0.098	-0.179	-0.089	-2.1152	1.7096	-2.4164
His(I)	Cβ	5.823	-37.71931	53.5	0.177	0.025	-0.051	-0.050	0.076	0.462	-0.268	0.464	0.426	-0.158	0.001	-0.225	0.021	-0.2192	2.1755	1.1438
His(I)	Hβ-1	0.994	-0.62769	47.0	0.006	-0.020	-0.125	0.007	0.127	-0.131	-0.266	0.405	-0.128	0.042	0.002	0.385	-0.076	-0.0367	4.2031	0.9448
His(I)	Hβ-2	1.019	-0.63594	48.1	-0.019	0.029	0.012	-0.125	0.129	-0.222	-0.180	0.403	-0.164	0.033	-0.097	-0.188	-0.135	-0.6003	1.7983	3.1256
His(I)	Cγ	5.605	-37.59709	65.5	0.395	0.120	-0.778	0.233	0.820	2.245	1.145	3.449	0.618	-0.904	-1.492	1.677	-1.284	2.2385	0.9930	0.3812
His(I)	Nδ1	8.533	-55.39707	103.3	-1.533	-0.146	-0.071	0.088	0.184	0.156	1.904	2.294	0.519	0.469	-1.187	0.722	-1.073	2.9045	-1.4592	0.9127
His(I)	Hδ1	0.513	-0.41916	22.6	0.487	0.074	0.102	-0.069	0.144	-0.051	-0.064	0.115	0.073	0.063	-0.034	-0.025	-0.024	1.8046	-2.7267	1.7645
His(I)	Cϵ1	4.635	-36.99658	56.4	1.365	-0.387	0.690	-0.067	0.794	0.671	2.023	2.805	0.996	0.007	-1.654	0.555	-1.145	5.2488	-1.8238	-0.0165
His(I)	Hϵ1	0.960	-0.60944	46.9	0.040	-0.045	0.109	-0.016	0.118	0.479	-0.177	0.484	-0.031	-0.298	-0.021	0.300	-0.068	6.2119	-3.5963	0.1443
His(I)	Cδ2	5.384	-37.46597	74.0	0.616	0.412	-0.534	0.009	0.674	2.194	1.570	3.781	1.239	-0.803	-1.979	1.232	-1.259	4.2698	1.9608	-0.8469
His(I)	Hδ2	0.990	-0.61914	49.1	0.010	-0.001	-0.115	0.039	0.122	-0.174	-0.324	0.506	-0.186	0.073	-0.089	0.387	-0.258	4.5089	3.8402	-1.5586
His(I)	Nϵ2	8.476	-55.23611	130.6	-1.476	0.022	0.136	-0.052	0.147	1.231	-0.100	1.367	-0.191	-0.402	-0.287	1.013	-0.537	6.1371	0.1809	-1.0865
His(I)	SUM	82.006	-545.68712	1260.9	-0.006															
His(I)	SCF	82	-545.68791		0	-1.860	-0.637	0.303	1.989											
His(II)	C	4.211	-36.69723	29.9	1.789	-0.201	0.935	-0.214	0.980	-0.674	-0.811	1.487	-0.199	0.190	0.984	-0.598	0.371	-3.1727	1.4713	0.2238
His(II)	O1	9.352	-75.68517	135.5	-1.352	-0.288	0.189	-0.455	0.570	0.012	0.472	0.552	-0.199	-0.074	0.417	0.071	-0.137	-4.3365	2.2658	-1.5295
His(II)	O2	9.299	-75.63692	119.4	-1.299	0.165	-0.032	0.262	0.312	0.778	-0.012	0.891	-0.565	-0.004	0.333	0.775	0.055	-2.2175	2.9569	2.0014
His(II)	H	0.353	-0.32780	18.6	0.647	0.020	-0.136	0.039	0.143	0.027	0.038	0.066	0.027	0.017	-0.000	-0.054	0.028	-2.4618	4.6717	1.5107

		$N(\Omega)$	$E(\Omega)$	$\nu(\Omega)$	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\lvert Q(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
HisII)	N	8.153	-54.90878	113.6	-1.153	-0.065	0.199	-0.024	0.211	1.349	1.047	2.402	1.120	0.556	-0.185	-2.293	0.200	-4.8980	-2.8544	0.1737
His(II)	H1	0.645	-0.48724	31.9	0.355	0.140	-0.052	-0.114	0.188	-0.072	-0.047	0.119	0.027	0.011	-0.088	-0.061	0.007	-6.2976	-2.4678	1.3893
His(II)	H2	0.622	-0.47578	30.3	0.378	0.068	-0.036	0.162	0.179	-0.024	-0.069	0.096	-0.055	0.005	0.037	-0.019	-0.033	-5.5587	-2.4832	-1.5631
His(II)	Ca	5.421	-37.49385	41.1	0.579	-0.463	-0.148	-0.101	0.497	0.910	-0.139	0.980	0.105	0.555	0.399	0.439	0.058	-2.6831	-1.3282	0.6963
His(II)	Ha	1.011	-0.64492	46.9	-0.011	-0.023	0.018	-0.120	0.124	-0.163	-0.218	0.382	-0.176	-0.003	0.149	-0.160	-0.049	-2.1826	-1.5655	2.6651
His(II)	Cb	5.833	-37.72897	53.9	0.167	0.025	0.037	0.037	0.057	-0.183	-0.266	0.452	0.376	0.198	-0.060	-0.122	0.006	-0.4891	-2.2914	-0.9586
His(II)	Hb-1	1.014	-0.63575	48.2	-0.014	-0.024	0.127	-0.016	0.130	-0.135	-0.270	0.412	-0.136	-0.027	-0.011	0.380	-0.123	-0.3581	-4.3054	-0.6160
His(II)	Hb-2	1.007	-0.63087	48.2	-0.007	0.035	-0.002	0.124	0.128	-0.239	-0.161	0.403	-0.127	-0.022	0.134	-0.222	-0.100	-0.9969	-2.0375	-2.9298
His(II)	Cg	5.424	-37.51476	59.7	0.576	0.250	0.594	-0.180	0.669	1.241	1.917	3.183	0.712	0.598	1.303	1.037	-1.794	2.0243	-1.0738	-0.4865
His(II)	Nd1	8.479	-55.25524	125.4	-1.479	-0.157	-0.049	-0.020	0.165	1.202	-0.132	1.319	0.619	0.736	0.203	0.271	-0.310	2.5967	1.2898	-1.4005
His(II)	He2	0.549	-0.43683	26.9	0.451	-0.147	0.007	-0.062	0.160	-0.040	-0.038	0.078	0.059	-0.005	0.043	-0.039	-0.001	7.5882	-0.2523	1.4046
His(II)	Ce1	4.622	-36.99345	55.2	1.378	-0.409	-0.670	0.156	0.801	0.603	1.959	2.679	1.047	0.070	1.468	0.246	-1.345	4.8786	1.7523	-0.6767
His(II)	He1	0.955	-0.60746	46.4	0.045	-0.072	-0.092	0.015	0.118	-0.178	-0.298	0.481	-0.038	0.286	-0.033	0.271	-0.161	5.9204	3.4410	-1.0742
His(II)	Cd2	5.572	-37.55209	80.5	0.428	0.629	0.536	-0.006	0.826	2.411	1.486	3.933	1.190	1.216	1.538	1.148	-1.913	4.0312	-2.0076	0.7989
His(II)	Hd2	0.965	-0.60510	47.7	0.035	-0.023	0.104	-0.060	0.123	-0.187	-0.306	0.498	-0.192	-0.060	0.079	0.314	-0.318	4.3249	-3.7623	1.7568
His(II)	Ne2	8.511	-55.36771	106.7	-1.511	0.164	-0.061	0.095	0.199	0.073	2.022	2.378	1.473	0.274	1.325	-0.290	-0.763	5.8613	-0.1720	0.6694
His(II)		82.000	-545.68594	1266.0	0.000															
His(II)		82	-545.68511		0	1.566	-0.161	1.145	1.946											
$His^{(+)}$	C	4.206	-36.68805	31.2	1.794	-0.266	-0.955	0.129	1.000	-0.721	-0.802	1.524	0.954	-0.539	0.822	-0.559	-0.294	-3.4404	-1.3882	-0.0987
$His^{(+)}$	O1	9.370	-75.70361	126.8	-1.370	0.317	-0.390	0.230	0.553	0.102	0.571	0.725	0.451	-0.065	0.302	0.079	-0.231	-2.1436	-3.0107	0.8082
$His^{(+)}$	O2	9.304	-75.66832	117.5	-1.304	-0.270	0.177	-0.145	0.354	0.749	0.044	0.892	0.145	0.374	0.285	0.515	-0.119	-5.7385	-1.8143	-0.9132
$His^{(+)}$	H	0.328	-0.31201	17.1	0.672	0.030	0.126	-0.017	0.131	0.033	0.026	0.059	0.030	-0.014	0.005	-0.054	0.013	-6.1794	-3.5488	-0.6844
$His^{(+)}$	N	8.136	-54.91168	111.4	-1.136	-0.054	-0.168	0.020	0.178	1.297	1.052	2.353	0.792	-1.240	0.129	-1.792	0.426	-4.5504	3.2041	-0.0533
$His^{(+)}$	H1	0.596	-0.46434	28.6	0.404	0.129	0.019	0.107	0.169	-0.027	-0.066	0.096	0.014	0.005	0.079	-0.027	0.003	-5.9747	3.0463	-1.2870
$His^{(+)}$	H2	0.623	-0.47551	30.9	0.377	0.071	0.024	-0.164	0.180	-0.059	-0.049	0.109	-0.034	-0.015	-0.051	-0.051	0.025	-5.2824	3.0966	1.6906
$His^{(+)}$	Cα	5.416	-37.48489	41.7	0.584	-0.388	0.210	0.107	0.454	-0.122	0.829	0.896	0.080	-0.388	-0.033	0.618	0.110	-2.6007	1.3485	-0.4941
$His^{(+)}$	Hα	0.985	-0.62903	46.1	0.015	-0.037	0.008	0.116	0.122	-0.205	-0.186	0.391	-0.153	0.010	-0.136	-0.197	-0.060	-2.0763	1.4786	-2.4751
$His^{(+)}$	Cβ	5.818	-37.71839	53.6	0.182	0.168	-0.089	-0.063	0.200	-0.430	-0.355	0.786	0.618	-0.393	-0.140	-0.282	0.046	-0.2934	2.0327	1.1476
$His^{(+)}$	Hβ-1	0.956	-0.61250	44.0	0.044	-0.034	-0.119	0.001	0.124	-0.124	-0.259	0.390	-0.124	0.022	0.012	0.379	-0.040	-0.1088	4.0649	1.0520
$His^{(+)}$	Hβ-2	0.983	-0.62029	46.3	0.017	0.022	0.028	-0.118	0.123	-0.218	-0.185	0.403	-0.168	0.035	-0.101	-0.180	-0.140	-0.6659	1.5562	3.1078
$His^{(+)}$	Cγ	5.528	-37.55873	61.9	0.472	0.323	-0.736	0.014	0.804	2.034	0.795	2.918	0.464	-0.135	-1.036	1.987	-0.453	2.2122	0.9333	0.3933
$His^{(+)}$	Nδ1	8.545	-55.45894	100.9	-1.545	-0.183	-0.030	0.062	0.195	0.167	1.389	1.709	0.743	0.520	-0.661	0.606	-0.334	2.7991	-1.6138	0.5049
$His^{(+)}$	Hδ1	0.417	-0.36270	16.3	0.583	0.069	0.088	-0.030	0.116	-0.058	-0.049	0.108	0.090	0.049	-0.010	-0.039	-0.008	1.5139	-2.9447	0.9434
$His^{(+)}$	Cϵ1	4.577	-36.95980	54.0	1.423	-0.441	0.824	0.045	0.936	0.192	1.688	2.069	1.056	0.519	-0.941	0.496	-0.414	5.1546	-1.9747	-0.1638
$His^{(+)}$	Hϵ1	0.832	-0.55180	39.9	0.168	-0.052	0.095	0.006	0.108	-0.168	-0.226	0.395	-0.052	-0.232	-0.028	0.271	0.017	6.1015	-3.7577	-0.2618
$His^{(+)}$	Cδ2	5.506	-37.53010	76.9	0.494	0.501	-0.593	-0.074	0.780	2.230	1.223	3.501	1.469	-0.599	-1.415	1.524	-0.335	4.3436	2.0934	-0.3715

		$N(\Omega)$	$E(\Omega)$	v(Ω)	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\|\mu(\Omega)\|$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\|\Theta(\Omega)\|$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
$His^{(+)}$	Hδ2	0.879	-0.56984	42.8	0.121	-0.026	-0.109	0.019	0.113	-0.180	-0.249	0.430	-0.164	0.110	-0.037	0.392	-0.096	4.7490	4.0453	-0.6877
$His^{(+)}$	Nϵ2	8.503	-55.40125	103.3	-1.503	0.174	0.139	-0.065	0.232	0.035	1.630	1.903	1.314	0.185	-0.906	0.056	-0.241	6.1502	0.2359	-0.7047
$His^{(+)}$	Hϵ2	0.485	-0.40098	23.7	0.515	-0.132	-0.018	0.041	0.140	-0.033	-0.028	0.062	0.050	0.018	-0.025	-0.030	-0.005	7.9323	0.5107	-1.2635
$His^{(+)}$	SUM	81.995	-546.08278	1215.1	1.005															
$His^{(+)}$	SCF	82	-546.08178		+1															
Phe	C	4.234	-36.71623	31.4	1.766	0.940	0.244	0.255	1.004	-0.671	-0.853	1.528	-0.456	-0.640	0.136	1.233	-0.396	5.2930	0.3890	-0.1020
Phe	O1	9.349	-75.69080	135.2	-1.349	0.419	-0.369	0.160	0.580	0.081	0.493	0.621	0.241	-0.173	0.269	0.240	-0.120	6.9552	-0.9849	0.5350
Phe	O2	9.303	-75.63726	119.6	-1.303	-0.179	0.240	-0.081	0.310	0.002	0.828	0.958	0.487	0.400	0.314	0.258	-0.006	5.5981	2.8776	-0.2789
Phe	H	0.351	-0.32659	18.3	0.649	-0.133	-0.032	-0.038	0.142	0.037	0.027	0.065	-0.053	-0.018	-0.024	0.033	-0.007	7.2826	3.2959	0.1996
Phe	N	8.159	-54.92782	111.9	-1.159	0.189	0.032	0.005	0.192	0.977	1.227	2.208	-1.745	-1.115	-0.187	0.561	0.018	2.4535	-3.0944	-1.3642
Phe	H1	0.639	-0.48460	31.4	0.361	-0.088	0.033	0.160	0.186	-0.069	-0.045	0.115	-0.062	-0.003	-0.032	-0.037	0.036	3.2219	-3.4599	-3.0546
Phe	H2	0.624	-0.47708	30.1	0.376	-0.091	0.100	-0.117	0.179	-0.020	-0.072	0.097	0.001	-0.014	0.066	-0.012	-0.049	3.4091	-4.1261	-0.0952
Phe	Cα	5.416	-37.48751	41.8	0.584	0.127	-0.470	-0.109	0.499	-0.174	-0.805	1.044	-0.083	0.344	0.258	0.770	0.311	2.6146	-0.4409	-0.7555
Phe	Hα	1.000	-0.63835	46.7	-0.000	0.026	-0.062	0.104	0.124	-0.171	-0.208	0.379	-0.124	-0.083	0.146	-0.024	-0.246	2.0238	0.6468	-2.3863
Phe	Cβ	5.857	-37.73548	54.9	0.143	0.011	0.001	-0.020	0.023	0.269	0.003	0.312	0.217	-0.104	-0.116	-0.045	-0.097	0.8529	0.2203	1.4791
Phe	Hβ-1	1.032	-0.64217	49.2	-0.032	-0.030	0.077	-0.099	0.129	-0.183	-0.230	0.414	-0.174	-0.070	0.056	0.065	-0.312	1.2162	-1.1161	2.9943
Phe	Hβ-2	1.027	-0.64483	46.8	-0.027	-0.030	-0.104	-0.050	0.119	-0.138	-0.283	0.430	-0.098	0.146	0.028	0.318	0.191	1.3406	2.0787	2.1905
Phe	Cγ	6.015	-37.85821	69.3	-0.015	-0.005	0.004	-0.034	0.034	1.606	1.495	3.102	1.266	0.194	0.981	1.435	-0.860	-1.9212	0.1908	0.7819
Phe	Cδ1	5.976	-37.82961	81.0	0.024	-0.002	0.061	-0.052	0.081	1.926	1.537	3.470	1.619	0.328	1.131	1.417	-0.850	-3.3387	-2.0220	0.8756
Phe	Hδ1	1.000	-0.62672	48.1	0.000	-0.043	0.105	-0.044	0.121	-0.128	-0.338	0.482	0.011	-0.238	0.097	0.280	-0.162	-2.4495	-3.7588	1.4379
Phe	Cϵ1	5.992	-37.82889	83.8	0.008	0.033	0.054	-0.007	0.064	1.651	1.890	3.543	1.491	0.146	1.258	1.521	-1.003	-5.8774	-2.0155	0.2254
Phe	Hϵ1	1.010	-0.62863	49.5	-0.010	0.065	0.104	-0.005	0.123	0.502	-0.150	0.516	0.016	0.299	0.033	0.316	-0.060	-6.9424	-3.7451	0.3097
Phe	Cδ2	6.001	-37.83677	83.2	-0.001	-0.023	-0.049	0.008	0.054	1.676	1.974	3.654	1.589	0.152	1.317	1.522	-0.999	-3.1124	2.4104	0.0264
Phe	Hδ2	1.020	-0.63287	49.7	-0.020	-0.064	-0.105	0.007	0.123	0.487	-0.115	0.509	0.039	0.285	0.057	0.305	-0.059	-2.0539	4.1477	-0.0418
Phe	Cϵ2	5.992	-37.82814	84.0	0.008	0.036	-0.051	0.025	0.068	1.887	1.666	3.555	1.520	0.338	1.238	1.525	-0.939	-5.6510	2.4244	-0.6325
Phe	Hϵ2	1.011	-0.62881	49.7	-0.011	0.053	-0.105	0.035	0.123	-0.150	-0.353	0.517	-0.038	-0.233	0.128	0.320	-0.195	-6.5359	4.1618	-1.2081
Phe	Cζ	5.991	-37.82616	84.5	0.009	0.070	0.004	0.017	0.072	1.920	1.708	3.631	1.404	0.266	1.211	1.704	-1.041	-7.0433	0.2074	-0.5334
Phe	Hζ	1.011	-0.62847	49.6	-0.011	0.119	-0.000	0.030	0.123	-0.153	-0.353	0.518	0.453	0.009	0.204	-0.160	-0.037	-9.0131	0.2113	-1.0351
	SUM	88.010	-551.56199	1449.8	-0.010															
	SCF	88	-551.56353		0	0.065	0.488	-0.188	0.527											
Tyr	C	4.236	-36.71600	31.4	1.764	-0.944	-0.262	0.217	1.003	-0.689	-0.848	1.539	-0.448	-0.673	-0.159	1.199	0.445	-6.1010	-0.4227	-0.2172
Tyr	O1	9.350	-75.68846	135.5	-1.350	-0.428	0.358	0.160	0.580	0.082	0.493	0.622	0.261	-0.167	-0.254	0.227	0.138	-7.7995	0.9071	0.4180
Tyr	O2	9.303	-75.63467	120.1	-1.303	0.185	-0.234	-0.084	0.310	0.828	0.002	0.957	0.497	0.412	-0.258	0.265	0.054	-6.3711	-2.9065	-0.4970
Tyr	H	0.351	-0.32644	18.6	0.649	0.134	0.035	-0.033	0.142	0.038	0.027	0.065	-0.055	-0.021	0.021	0.033	0.007	-8.0651	-3.3594	-0.0888
Tyr	N	8.160	-54.92703	111.0	-1.160	-0.187	-0.033	-0.002	0.190	-2.158	0.947	2.164	-1.690	-1.117	0.057	0.507	-0.065	-3.2668	3.1380	-1.2511

		$N(\Omega)$	$E(\Omega)$	$\nu(\Omega)$	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\lvert\mu(\Omega)\rvert$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\lvert\mathbf{Q}(\Omega)\rvert$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
Tyr	H1	0.639	-0.48456	31.4	0.361	0.083	-0.039	0.161	0.186	-0.069	-0.045	0.115	-0.064	-0.004	0.028	-0.034	-0.041	-3.9871	3.5622	-2.9490
Tyr	H2	0.623	-0.47690	30.0	0.377	0.096	-0.094	-0.118	0.179	-0.020	-0.072	0.097	0.006	-0.012	-0.068	-0.016	0.046	-4.2763	4.1059	0.0266
Tyr	Cα	5.417	-37.48840	41.8	0.583	-0.143	0.482	-0.098	0.512	-0.112	0.941	1.029	-0.051	0.300	-0.246	0.788	-0.244	-3.4137	0.4597	-0.7532
Tyr	Hα	1.001	-0.63864	46.7	-0.001	-0.030	0.058	0.106	0.125	-0.169	-0.208	0.378	-0.114	-0.083	-0.160	-0.046	0.234	-2.7601	-0.5553	-2.4067
Tyr	Cβ	5.852	-37.72863	54.5	0.148	-0.014	0.005	-0.023	0.028	-0.015	0.248	0.278	0.191	-0.085	0.123	-0.015	0.040	-1.7111	-0.2706	1.5056
Tyr	Hβ-1	1.034	-0.64255	49.3	-0.034	0.035	-0.073	-0.101	0.130	-0.182	-0.232	0.415	-0.170	-0.084	-0.070	0.037	0.305	-2.1403	1.0007	3.0601
Tyr	Hβ-2	1.030	-0.64628	46.8	-0.030	0.030	0.106	-0.045	0.119	-0.133	-0.288	0.431	-0.095	0.147	-0.015	0.335	-0.167	-2.2033	-2.1588	2.1310
Tyr	Cγ	5.995	-37.83888	70.1	0.005	0.039	-0.010	-0.040	0.057	1.776	1.606	3.383	1.424	0.117	-0.944	1.652	0.757	1.0843	-0.1928	0.9062
Tyr	Cδ1	5.974	-37.83238	79.7	0.026	0.031	-0.037	-0.047	0.068	1.826	1.394	3.230	1.610	0.237	-0.930	1.339	0.600	2.4582	2.0563	0.9615
Tyr	Hδ1	0.989	-0.62323	47.5	0.011	0.045	-0.104	-0.039	0.120	-0.127	-0.333	0.475	0.033	-0.253	-0.073	0.272	0.126	1.5139	3.7922	1.4257
Tyr	Cϵ1	5.941	-37.81030	83.5	0.059	0.049	-0.181	-0.044	0.193	1.627	2.187	3.827	1.713	-0.070	-1.235	1.684	0.896	5.0095	2.1110	0.4072
Tyr	Hϵ1	0.985	-0.61622	47.7	0.015	-0.058	-0.108	-0.005	0.123	0.474	-0.138	0.487	0.021	0.280	-0.034	0.300	0.039	6.0551	3.8511	0.4554
Tyr	Cδ2	5.994	-37.83505	81.7	0.006	0.067	0.016	-0.009	0.069	1.535	1.876	3.417	1.606	0.086	-1.109	1.441	0.739	2.3716	-2.3876	0.2782
Tyr	Hδ2	1.015	-0.63137	49.6	-0.015	0.058	0.108	0.005	0.123	0.484	-0.112	0.506	0.025	0.269	-0.055	0.326	0.046	1.3783	-4.1633	0.2311
Tyr	Cϵ2	5.968	-37.81353	84.1	0.032	0.045	0.142	0.017	0.150	2.237	1.781	4.027	1.705	0.411	-1.209	1.916	0.827	4.9357	-2.3660	-0.2916
Tyr	Hϵ2	1.025	-0.63188	50.4	-0.025	-0.067	0.100	0.030	0.124	-0.135	-0.366	0.518	-0.016	-0.240	-0.114	0.332	0.161	5.8855	-4.1026	-0.7692
Tyr	Cζ	5.479	-37.52247	61.5	0.521	0.809	0.035	-0.174	0.829	0.594	2.433	3.207	2.188	0.281	-1.084	0.520	0.516	6.2568	-0.1101	-0.2255
Tyr	Oη	9.273	-75.54572	121.7	-1.273	0.141	0.217	0.003	0.259	-0.017	0.929	1.063	0.442	-0.408	-0.362	0.345	0.288	8.7571	0.0446	-0.7502
Tyr	Hη	0.376	-0.34554	19.6	0.624	-0.054	0.138	0.033	0.152	0.037	0.024	0.061	0.027	0.028	0.004	-0.048	-0.018	9.4326	-1.5633	-1.1519
	SUM	96.008	-626.44512	1514.1	-0.008															
	SCF	96	-626.44633		0	-0.137	-1.026	-0.272	1.071											
Trp	C	4.227	-36.71273	30.1	1.773	0.241	0.378	0.896	1.002	-0.802	-0.696	1.500	-0.790	0.146	-0.084	0.950	-0.939	5.3304	0.5438	1.2639
Trp	O1	9.351	-75.69390	135.3	-1.351	0.034	-0.288	0.500	0.578	-0.480	0.006	0.551	-0.452	0.066	0.100	0.092	-0.205	5.4957	-0.5190	3.2381
Trp	O2	9.304	-75.63675	120.2	-1.304	0.001	0.204	-0.227	0.305	-0.841	0.035	0.951	-0.727	0.264	0.322	0.343	0.329	5.7719	3.0074	1.0256
Trp	H	0.352	-0.32693	18.7	0.648	-0.030	-0.053	-0.129	0.142	0.027	0.038	0.065	0.023	-0.005	-0.018	0.026	-0.033	6.1536	3.6838	2.6509
Trp	N	8.147	-54.90650	113.9	-1.147	0.058	0.083	0.159	0.188	1.086	1.363	2.454	1.083	-0.391	-0.813	0.152	-1.530	6.4783	-2.7813	-1.7622
Trp	H1	0.641	-0.48527	31.7	0.359	-0.173	-0.070	0.014	0.187	-0.047	-0.073	0.120	0.076	0.039	-0.067	-0.034	-0.020	8.2240	-2.1136	-2.0638
Trp	H2	0.623	-0.47610	30.5	0.377	-0.016	0.112	-0.140	0.180	-0.067	-0.026	0.096	-0.063	-0.001	-0.017	0.021	-0.060	6.5716	-3.9959	-0.3115
Trp	Cα	5.419	-37.49701	40.2	0.581	0.403	-0.293	0.020	0.499	-0.729	-0.114	0.915	-0.307	-0.593	-0.020	0.320	0.420	4.6991	-0.7612	-1.2282
Trp	Hα	1.015	-0.64858	45.7	-0.015	-0.011	-0.079	0.089	0.120	-0.217	-0.163	0.382	-0.210	0.005	-0.029	0.106	-0.274	4.8443	0.6105	-2.7372
Trp	Cβ	5.840	-37.72632	54.0	0.160	0.015	0.042	0.015	0.048	-0.138	-0.330	0.480	0.416	-0.149	-0.093	-0.098	0.016	2.0239	-1.8672	-1.1492
Trp	Hβ-1	1.005	-0.63200	47.8	-0.005	0.026	0.076	0.103	0.131	-0.139	-0.256	0.401	-0.137	0.009	0.046	-0.079	0.287	1.7987	-2.9681	-2.8646
Trp	Hβ-2	1.021	-0.63916	48.0	-0.021	0.008	0.091	-0.087	0.126	-0.141	-0.251	0.397	-0.141	0.026	-0.014	-0.010	-0.310	1.9323	-3.1925	0.4159
Trp	Cγ	5.984	-37.82105	72.7	0.016	-0.116	0.147	-0.039	0.191	2.228	1.281	3.551	1.217	-0.510	-0.976	1.955	-0.751	-0.1002	0.0105	-0.9112
Trp	Cδ1	5.540	-37.54130	74.3	0.460	-0.719	0.405	0.105	0.831	2.244	1.227	3.521	1.977	-0.432	-0.947	1.203	-0.692	-0.0903	2.5344	-1.3159
Trp	Hδ1	0.967	-0.61209	45.6	0.033	-0.075	-0.075	0.040	0.113	-0.169	-0.319	0.495	0.272	0.291	-0.104	0.016	-0.101	1.4379	3.7601	-1.8195

		$N(\Omega)$	$E(\Omega)$	v(Ω)	$q(\Omega)$	$\mu_x(\Omega)$	$\mu_y(\Omega)$	$\mu_z(\Omega)$	$\|\mu(\Omega)\|$	$\Theta_{xx}(\Omega)$	$\Theta_{yy}(\Omega)$	$\|\mathbf{Q}(\Omega)\|$	$Q_{xx}(\Omega)$	$Q_{xy}(\Omega)$	$Q_{xz}(\Omega)$	$Q_{yy}(\Omega)$	$Q_{yz}(\Omega)$	x	y	z
Trp	Nϵ1	8.506	-55.32694	106.2	-1.506	-0.040	0.163	-0.014	0.169	1.978	0.262	2.450	0.264	-0.488	-0.465	1.810	-0.471	-2.4718	3.5177	-0.9762
Trp	Hϵ1	0.560	-0.44319	27.5	0.440	0.039	-0.159	0.011	0.164	-0.039	-0.038	0.077	-0.033	-0.027	0.002	0.070	-0.010	-2.9183	5.3348	-1.1140
Trp	Cδ2	5.999	-37.88743	72.0	0.001	-0.087	0.166	-0.007	0.188	1.415	2.155	3.595	1.373	0.140	-1.052	1.837	-0.887	-2.6820	-0.6347	-0.2692
Trp	Cϵ2	5.543	-37.61121	61.1	0.457	0.469	0.588	-0.194	0.777	0.642	2.451	3.264	1.358	0.727	-1.054	1.373	-0.871	-4.0883	1.6061	-0.3274
Trp	Cζ2	5.965	-37.79826	84.7	0.035	0.171	-0.023	-0.033	0.175	1.717	2.096	3.819	1.843	-0.134	-1.173	1.594	-0.819	-6.6690	1.6586	0.2112
Trp	Hζ2	1.013	-0.62632	50.3	-0.013	0.057	-0.111	0.005	0.125	0.503	-0.157	0.515	0.021	-0.303	-0.030	0.313	-0.040	-7.7265	3.3949	0.1618
Trp	Cϵ3	5.991	-37.81745	83.6	0.009	0.022	0.088	-0.015	0.092	1.629	1.834	3.464	1.536	-0.063	-1.039	1.586	-0.779	-3.8978	-2.9041	0.3401
Trp	Hϵ3	1.009	-0.62729	50.1	-0.009	-0.064	0.104	-0.002	0.122	0.495	-0.127	0.514	0.045	-0.288	-0.035	0.310	-0.045	-2.8638	-4.6536	0.3952
Trp	Cζ3	5.975	-37.79864	85.0	0.025	0.050	0.080	-0.022	0.097	2.102	1.725	3.834	1.817	-0.293	-1.150	1.643	-0.794	-6.4453	-2.8709	0.8724
Trp	Hζ3	1.015	-0.63012	50.1	-0.015	0.060	0.106	-0.029	0.125	-0.148	-0.356	0.519	-0.012	0.255	-0.113	0.319	-0.160	-7.4038	-4.5992	1.3454
Trp	Cη	5.979	-37.80663	83.7	0.021	0.069	0.051	-0.022	0.089	1.578	1.969	3.553	1.394	-0.055	-1.039	1.804	-0.824	-7.8187	-0.5985	0.8080
Trp	Hη	1.012	-0.62930	49.9	-0.012	0.121	0.001	-0.026	0.124	-0.147	-0.354	0.515	0.464	0.004	-0.174	-0.151	-0.033	-9.8062	-0.6255	1.2330
	SUM	108.002	-682.35847	1712.9	-0.002															
	SCF	108	-682.35854		0	0.245	1.233	-0.315	1.296											
Pro	C	4.238	-36.71197	31.6	1.762	-0.800	-0.283	-0.536	1.004	-0.695	-0.852	1.551	-0.418	-0.755	-0.196	1.013	0.647	-2.4914	-0.4054	-0.0533
Pro	O1	9.350	-75.68499	135.1	-1.350	-0.052	-0.496	-0.294	0.579	0.091	0.464	0.595	-0.014	-0.088	0.226	0.247	0.385	-2.7786	-2.3093	-1.2166
Pro	O2	9.303	-75.63017	120.0	-1.303	-0.025	0.279	0.137	0.311	0.009	0.829	0.962	0.576	-0.147	0.566	-0.133	0.281	-4.2582	1.3795	0.0533
Pro	H	0.351	-0.32677	18.6	0.649	0.112	0.043	0.077	0.142	0.038	0.027	0.065	-0.025	-0.028	-0.040	0.025	-0.015	-5.6801	0.8497	-0.9150
Pro	N	8.246	-55.01860	99.3	-1.246	-0.097	0.001	-0.160	0.187	1.311	1.018	2.335	0.606	0.108	-1.110	1.294	0.145	1.6904	-1.7285	1.5548
Pro	H1	0.614	-0.47792	28.4	0.386	0.069	0.158	0.018	0.174	-0.028	-0.066	0.097	0.004	0.057	0.036	0.043	0.021	0.9050	-3.4273	1.3182
Pro	H2																			
Pro	Cα	5.448	-37.51921	43.3	0.552	0.190	-0.464	-0.040	0.503	1.003	-0.238	1.048	0.618	-0.638	0.300	-0.163	0.136	-0.1694	0.2728	1.4989
Pro	Hα	1.008	-0.63967	47.3	-0.008	0.055	-0.027	-0.109	0.125	-0.243	-0.154	0.400	-0.106	-0.050	-0.152	-0.214	0.122	-0.8527	0.7095	3.3832
Pro	Cβ	5.867	-37.74628	56.2	0.133	-0.024	-0.065	0.066	0.096	0.306	-0.059	0.324	-0.092	-0.063	0.091	0.284	-0.061	1.2629	2.5322	0.3592
Pro	Hβ-1	1.024	-0.63858	48.0	-0.024	0.062	-0.092	0.046	0.120	-0.193	-0.266	0.461	0.092	-0.275	0.177	0.071	-0.180	0.0222	3.9556	-0.4260
Pro	Hβ-2	1.037	-0.64466	49.7	-0.037	-0.067	-0.062	-0.090	0.128	-0.239	-0.202	0.441	-0.018	0.164	0.252	-0.115	0.200	2.3770	3.4082	1.8404
Pro	Cγ	5.872	-37.74614	56.8	0.128	-0.028	-0.007	0.011	0.031	-0.302	0.010	0.344	0.024	0.206	-0.097	-0.166	-0.112	2.9855	1.2611	-1.5862
Pro	Hγ-1	1.052	-0.65161	50.4	-0.052	0.062	0.015	0.109	0.127	-0.217	-0.227	0.444	-0.038	0.066	0.288	-0.197	0.114	1.9472	0.8975	-3.3188
Pro	Hγ-2	1.045	-0.64455	50.6	-0.045	-0.099	-0.071	0.037	0.127	-0.248	-0.227	0.475	0.220	0.320	-0.120	-0.024	-0.093	4.6174	2.4018	-2.0687
Pro	Cδ	5.454	-37.48838	50.4	0.546	-0.343	-0.113	0.351	0.504	-0.434	-0.383	0.818	0.289	0.102	-0.603	-0.395	-0.054	3.7040	-1.2319	-0.2923
Pro	Hδ-1	1.053	-0.65910	49.9	-0.053	-0.121	-0.010	-0.047	0.131	-0.241	-0.191	0.433	0.264	0.058	0.277	-0.223	0.008	5.4989	-1.0586	0.6852
Pro	Hδ-2	1.043	-0.65125	49.7	-0.043	-0.022	0.096	0.081	0.128	-0.213	-0.230	0.443	-0.211	-0.027	-0.029	0.150	0.331	3.8766	-2.7711	-1.6342
	SUM	62.005	-398.87988	985.4	-0.005															
	SCF	62	-398.88070		0	-0.038	0.658	-0.164	0.679											

PART III:

PROPERTIES OF LARGE SYSTEMS FROM THE ADDITIVE CONTRIBUTIONS OF PROPER OPEN QUANTUM SUB-SYSTEMS

Chapter 6

Assembling Large Molecules from Buffered Fragments Determined as Proper Open Quantum Systems: The Properties of the Oripavine PEO, Enkephalins, and Morphine

6.1 Statement of the Problem

Ab initio methods scale rapidly with the size of the basis set and for this reason cannot be applied to large molecules of biological interest. This chapter describes an attempt to solve this problem by developing a method that produces results of *ab initio* quality but scale more slowly with size of the basis set/molecule. Some opioids are selected as a test case for the method. The method described here exploits the transferability of atomic and functional group properties which is amply demonstrated in the previous chapters of the thesis. The present chapter is based on (*Matta, 2001c*) which has been recently reviewed as to its implications for medicinal chemistry (*Bader et al.,Alber and Carloni, 2002*) and was reported upon in an article in *Chemical and Engineering News*. (*Wilson, 2001*)

6.2 Introduction

A new method of computing the properties of large molecules, not amenable to direct computation, is presented. In this method, the large molecule is assembled from judiciously chosen fragment molecules which are small enough for direct computation but large enough to retain good approximations to the electron densities of the respective moieties in the large molecule. The atoms in the fragment molecules are determined as proper open systems according to the quantum theory of atoms in molecules (QT-AIM). The properties of the atoms in the fragment molecules corresponding to atoms in the large molecule are summed to obtain the properties of the large molecule. The computational effort is, thus, reduced from the n^{th} power of a large number (the size of the basis set describing the large molecule) to the sum of the n^{th} power of small numbers (the sizes of the basis sets describing each fragment molecule). The time required to obtain the atomic properties by numerical integration is also reduced significantly, since these calculations also scale rapidly with the size of the basis set. The method was tested on the oripavine molecule 7-α-[1-(*R*)-hydroxy-1-methyl-3-phenylpropyl]-6,14-*endo*-etheno-tetrahydrooripavine (PEO) and was used to study some opioids. The properties of PEO computed from the fragment molecules were found to be remarkably close to those computed for the intact molecule, demonstrating the feasibility and usefulness of the method.

6.3 Assembling a Large Molecule from Buffered Fragments

In order to assemble a large molecule from fragments, the molecule is broken into separate parts (fragments) each of which is supplanted with hydrogen atoms as required to yield a set of closed-shell "fragment molecules". Separate self-consistent-field (SCF) calculations are performed on each one of the fragment molecules – which will be referred to as "fragments" from now on – followed by numerical integrations of the atomic properties (Eq. 2.61). The atomic properties of those atoms in the fragments which correspond to atoms in the large molecule are summed (Eq. 2.49) to yield approximations to the molecular properties of the large molecule. The accuracy with which the

properties of the large molecule are recovered from such a sum is determined by the transferability from the electronic environment in the intact molecule to the corresponding environment in the fragments.

One can anticipate the transferability of the properties of the atoms which are sufficiently removed from the severed edges of a fragment based on simple chemical intuition. QT-AIM (*Bader, 1990*) has been shown to recover the experimental transferability of group contributions to the heats of formation, (*Wiberg et al., 1987*) magnetic susceptibility, (*Keith and Bader, 1992; Keith and Bader, 1993; Keith and Bader, 1996*) electric polarizabilities, (*Bader and Keith, 1991; Bader et al., 1992*) partial molar volumes at infinite dilution, (*Matta and Bader, 2002a*) and to the heat of transfer from the gas to the solution phase. (*Matta and Bader, 2002a*) On the other hand, since the density uniquely determines the external potential (Eq. 1.42) up to an additive constant, as required by the Hohenberg and Kohn theorem (HK), (*Hohenberg and Kohn, 1964*) *perfect transferability* is an unattainable limit since the external potential of the intact molecule is different than that of any fragment molecule. The HK theorem, however, imposes no restrictions as to how close this limit can be approached. (*Bader and Becker, 1988*)

6.4 Previous Related Work

It is beyond the scope of this section to present a complete review of the different fragment approaches. Only those methods that are closely related to the present one are briefly reviewed here.

Massa, Huang and Karle (MHK) (*Huang et al., 1996; Massa et al., 1995*) express the density of a large molecule as a sum of densities of fragments. Both the MHK method and the buffered fragments method described in this chapter require as an input the geometry of the target molecule (obtained either from experiment or from theory), a geometry which is kept fixed and which serves to define the geometries of the fragments. A fragment in the MHK approach consists of an inner core "kernel" and a number of neighbouring atoms termed the "neighbourhood". The molecule

is divided into a suitable number of kernels, which, when recombined, form the complete molecule. The neighbourhood of each kernel acts as a "buffer" to provide an electronic environment as close as possible to the one experienced by the kernel in the intact molecule, an aspect of the MHK method similar to the buffered fragments approach described here. Hydrogen atoms are added when necessary to obtain closed-shell fragment molecules, and separate SCF calculations are perfomed on the fragments to obtain their respective electron densities. Up to this point the MHK and the buffered fragments methodologies are essentially similar but they differ in their following steps. In the MHK method, the density matrix of a fragment consists of the contributions of the atoms belonging to the kernel taken together with one-half of the contributions involving basis functions centered on atoms of the neighbourhood. The sum of the fragment density matrices yields the density matrix for the whole molecule which is constrained to be idempotent[1] and normalized. The fragment matrices can be obtained from either theoretical calculations or from experimental x-ray structure factors. In the latter case, additional constraints are imposed by requiring that the predicted density matrix reproduces the measured structure factors. (*Clinton et al., 1973; Clinton and Massa, 1972; Frishberg and Massa, 1981*) The density martices can then be used to obtain all the properties of the large inaccessible molecule.

Lecomte's group (*Pichon-Pesme et al., 1995; Pichon-Pesme et al., 1992; Wiest et al., 1994*) is building a data bank of transferable multipolar electron density parameters obtained from a fitting of the x-ray structure factors. This is brought about using the MOLLY program (*Hansen and Coppens, 1978*) which expands the density in terms of a set of overlapping atom-centred basis functions. They have shown that similar atoms in different environments, such as the atoms of the four different peptide groups of Leu-enkephalin, have statistically identical experimental multipolar density parameters. In other words, within this modelling of the density, similar atoms are transferable between different environments. Their strategy is essentially to transfer the multipolar charge-density

---

[1] For a definition of an "idempotent density matrix", see chapter 1: page 18.

parameters obtained from ultra high resolution X-ray diffraction experiments (d < 0.5Å) on amino acids and small peptides to similar atoms in a protein, significantly improving the refinement statistics and its electron density description. (*Jelsch et al., 1998*) The method was shown to dramatically enhance the resolution of two proteins: crambin (*Jelsch et al., 2000*) and scorpion toxin. (*Housset et al., 2000*) The same group used the multipolar densities so obtained to compute the molecular electrostatic potential of an octapeptide, a potential that was shown to be significantly different from one obtained from a density generated by spherical atoms refinement. (*Jelsch et al., 1998*)

Chang and Bader (*Chang and Bader, 1992*) have shown that the density of a polypeptide can be assembled by linking of amino acid residues (fragments) matched at their amidic zero-flux surfaces that bound and define each amino acid residue. In this approach, there is no question of overlapping basis functions from neighbouring fragments. Chang and Bader (*Chang and Bader, 1992*) have demonstrated that the properties of di- and tri-peptides can be closely approximated by linking such amino acid density fragments. In an extension to this method, Martín and Bader (*Bader and Martín, 1998; Martín, 2001*) obtained a complete library of accurate densities of tripeptides of the type Gly|Aa|Gly, where Aa stands for any of the 20 genetically-encoded amino acid residues and the vertical bars refer to the zero-flux amidic surfaces bounding the residue within the tripeptide mold. They showed that the shape of these amidic zero-flux surfaces and the density they bound are insensitive to a change in the nature of the side chain, (*Bader and Martín, 1998*) i.e. the main chain backbone atoms are essentially transferable and so are the side chain atoms. Exploiting this transferability and the almost perfect matching of the amino acid residues' zero-flux surfaces obtained in the Gly|Aa|Gly molds, software was developed to link these amino acid residue densities (stored in the library) sequentially in an arbitrary order and thus obtain essentially an excellent approximation to the accurate density of a peptide of arbitrary length. (*Martín, 2001*)

More recently, Hernández-Trujillo and Bader succeeded in constructing a one-electron density matrix and its derived electron density distribution iteratively, using the from an initial guess of the density of a molecule assembled from the densities of groups determined as proper open systems.

Their method takes advantage of the observation that the zero-flux boundary condition defining a proper open system maximises the transferability of a functional group from one molecule to another. In order to construct a pure-state density matrix, Hernández-Trujillo and Bader subject the density matrix to the iterative procedure of Clinton and co-workers, which imposes the idempotency and normalization constraints on the density matrix. (*Clinton et al., 1969*) The constructed densities and properties are shown to reproduce closely those calculated directly from the SCF procedure. This procedure corrects for the small mismatches of the zero-flux surfaces of neighbouring fragments extracted from different parent molecules and provides a density that in consistent with a properly antisymmetrized wavefunction. (*Hernández-Trujillo and Bader, 2001*)

In another effort, Breneman's group introduced the "Transferable Atom Equivalents" (TAE), (*Breneman et al., 1995; Breneman and Weber,Jeffrey and Piniella, 1991*) which are atomic density fragments bounded by "flexible" zero-flux surfaces to allow for correction of the surface mismatches. In Breneman's approach, the RECON program is used to build densities sequentially by joining TAEs stored in a library at their approximately matching surfaces, translating the nuclei to the positions initially specified in an input z-matrix. At each addition of a new TAE to the previous one, the two nearly matching zero-flux surfaces are rotated about the inter-nuclear axis to optimize the matching of their respective zero-flux surfaces. Next, the bond length is adjusted by balancing the surface electronic "pressure" on each side of the surface, a pressure readily computed from the quantum stress tensor. In the process, the two zero-flux surfaces are fused into a single new surface which is constrained to satisfy the zero-flux condition (Eq. 2.6). This procedure of "melding" the two surfaces of a new TAE to the previous one is repeated until the total molecular density is reconstructed. In building the TAE library, information about the derivatives of each of the atomic properties with respect to small radial variations in each unique position of the zero-flux surface are also stored. This information is used to correct for the changes in the atomic properties induced by the melding procedure. The TAE approach has been shown to be of great value in quantitative-structure-activity relationship studies. (*Breneman and Rhem, 1997*)

6.5 Selection of the Test Molecule for the Fragments Method: PEO

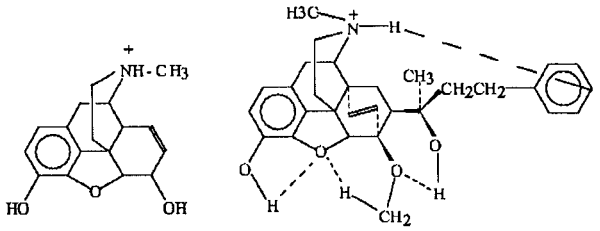
The success of a related approach in polypeptides (*Chang and Bader, 1992; Martín, 2001*) motivated the generalization to molecules which do not consist of simple repetitive units. PEO, an oripavine opioid (scheme 6.1), was selected as a test case for a number of reasons. (See list of abbreviation on page xxiv for the chemical name of PEO).

The molecule is theoretically interesting, since it contains a number of chemical features: aromatic rings, saturated rings, heterocyclic rings, 5- and 6-MRs, fused and non-fused rings, intramolecular hydrogen bonds, aliphatic regions, phenolic and methoxy groups, a quaternized nitrogen, etc. Thus, if the method works for this molecule, it is likely to work for a variety of organic molecules. Another advantage is the size of PEO: the molecule is small enough to be amenable intact to an SCF calculation (to compare with the assembled molecule), yet it is large enough to be broken into fragments representing different parts of the molecule.

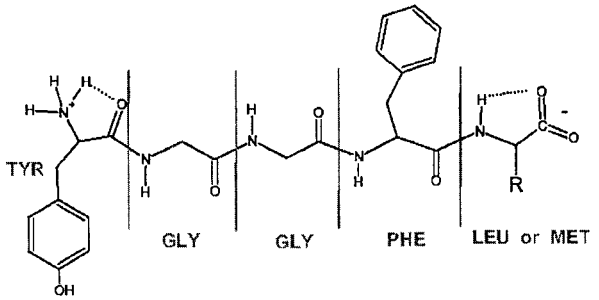
The PEO molecule is also biologically interesting since it has a strong affinity to the three opioid receptor subtypes (μ, κ, and δ). This molecule, a few thousand times more potent than morphine, is believed to be a prototypical substrate of the opioid receptors. (*Bradbury et al., 1976; Feinberg et al., 1976; Schiller,Udenfriend and Meienhofer, 1984; Snyder, 1977*) PEO was selected rather than its methylated derivative at the phenolic hydroxyl group, PET (see list of abbreviation on page xxiv for the chemical name), since blockade of the phenolic group drastically reduces the binding of opioids to the μ receptor. (*Reisine and Pasternak,Hardman et al., 1996*) Furthermore, methoxy opioids are believed to undergo *in vivo* de-methylation to be converted to the biologically-active phenoxy form. (*Reisine and Pasternak,Hardman et al., 1996*)

The ionized (quaternized) state of PEO (with a protonated nitrogen atom), was selected for calculations rather than the free-base state because this is the prevalent state of ionization of opioids at physiological pH (7.2-7.6) and temperature (37°C). (*Kaufman et al., 1975*) The numbering scheme of the PEO molecule is provided in scheme 6.2.

Scheme 6.1

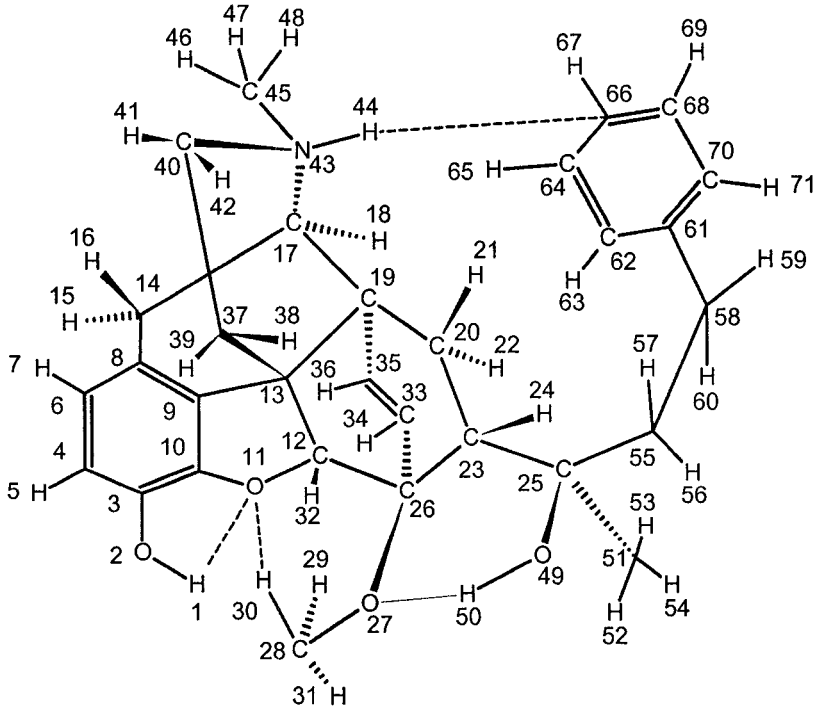


Morphine **PEO**



Enkephalin

Scheme 6.2



6.6 Computational Strategy and Methods

A search of the literature and of the crystrallographic databases revealed that neither the experimental nor a high quality optimised geometry of PEO, or of the closely-related molecule PET, are available. In view of the size of PEO (71 atoms), a fine grid high-level of theory scan of the potential energy surface is computationally prohibitive, even if restricted to the flexible side chain torsion angles. Alternatively, an initial geometry guess for a local full energy optimisation was obtained as follows.

The x-ray crystallographic geometry of a molecule from the oripavine family closely related to PEO, etorphine (in which the phenyl ring of the side chain is replaced by a methyl group), (*Michel et al., 1988*) was used as an initial guess for the geometry of the rigid moiety of PEO. It has been proposed that the flexible side chain of potent opioid receptor agonists such as PEO must adopt a 'scorpion' conformation, since this is the only common conformation accessible to several agonists. (*Feinberg et al., 1976; Schiller, Udenfriend and Meienhofer, 1984*) See figure 6.1.

The scorpion conformation is one in which the groups in the flexible side-chain of PEO adopt a staggered arrangement. This conformation ensures geometric conditions favouring the hydrogen bonding of the –O(49)H(50) of the side-chain with the methoxy oxygen O(27) which was characterised by a bond path. The presence of this hydrogen bond is consistent with the interpretation of the NMR spectrum of etorphine solution. (*Fulmor et al., 1967*) The scorpion conformation also allows another hydrogen bonding interaction between the electrophilic proton H(44) and carbon atom C(66) of the electron rich phenyl ring. This scorpion conformation was subjected to a full energy optimisation at the RHF/3-21G level to obtain geometry (**A**). Geometry **A** was used to probe the potential energy surface (PES) defined by the three torsion angles $\theta1$=C(23)-C(25)-C(55)-C(58), $\theta2$=C(25)-C(55)-C(58)-C(61), and $\theta3$=C(55)-C(58)-C(61)-C(62). See figure 6.1. The torsion angle around the C(23)-C(25) bond was frozen to preserve the O(49)-H(50)-O(27) hydrogen bonding, since it occurs in solution. (*Fulmor et al., 1967*) The PES was sampled at 216 points by a rigid scan in 60° steps around $\theta1$, $\theta2$, and $\theta3$, at the RHF/3-21G level using Gaussian94. (*Frisch et al., 1995*)

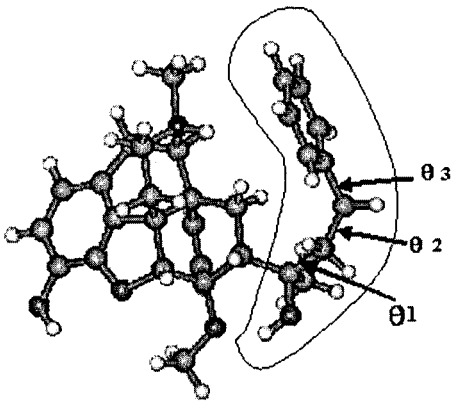


Figure 6.1 PEO in the scorpion (**A**) conformation, the global minimum. The flexible side chain is encircled and the three torsional degrees of freedom labeled: θ1, θ2, and θ3.

The potential energy scan under these conditions revealed that the scorpion geometry (**A**) is the global minimum which was selected for the fragment method.

PEO geometry **A** was used for a single point SCF calculation using a polarised basis set to obtain the electron density. The single point step was performed at the RHF/6-31G** level. The density was analysed and atomic integrations performed using AIMPAC. (*Biegler-König et al., 1982*)

Geometry **A** was then used to specify the geometries of the fragments. PEO was broken into four fragments representing different parts of the molecule, the fragments having some buffer atoms at their boundaries. Hydrogen atoms were added to the atoms participating in the bonding that was severed to obtain closed-shell "fragment molecules". See figure 6.2. Separate single-point SCF calculations were performed at the RHF/6-31G** level on each fragment frozen at a geometry defined by the intact optimised PEO (**A**). The atomic properties of the atoms in the fragments corresponding to those in PEO were summed (Eq. 2.49) to obtain the assembled properties of PEO. Assembling of the charge distributions of Leu- and Met-enkephalin was achieved using atomic charges obtained from (*Martin, 2001*) and calculated at the RHF//6-311++G**/6-31+G* level for the amino acid residues in Gly|Aa|Gly molds. Atomic properties of morphine in both ionization states (quaternized-cationic, and free base) were calculated directly (not from fragments) at the RHF//6-31G**/3-21G level.

6.7 Results and Discussion

6.7.1 Assembled versus Intact Properties of PEO

Table 6.1 compares the atomic electron populations in the four fragments and those of atoms in intact PEO chosen to represent different regions in the molecule. From the comparison one can identify which fragment recovers each of the atomic electron populations best, a comparison based on which fragments are selected as the providers of the atoms used to assemble PEO. The table also lists selected assembled properties: atomic populations, atomic energies, atomic volumes and atomic integrated Laplacian density. Table 6.2 provides a comparison of some atomic properties of a few atoms of intact and assembled PEO taken from different regions of the PEO molecule, showing the remarkable extent to which the assembled properties reproduce the atomic properties of intact PEO. The averages of the absolute deviations (av|dev|) between the two sets of properties are extremely small. As a result of this transferability of atomic properties from the buffered fragments to the intact PEO, the total molecular properties calculated from the buffered fragments recover those obtained for the intact PEO molecule to within a negligible error (see table 6.3).

The success of the buffered fragments method is a result of the high degree of transferability of atoms defined as proper open systems, the same property that underlies their use in molecular recognition. The atomic boundaries uniquely define corresponding fragments in different molecules. Once these fragments are defined, one can then study their chemical reactivity at any desired level of detail.

The comparison of a molecule with another is important for fields such as drug design where the similarities between molecules is quantitated using certain "descriptors" in an attempt to relate the distance between the molecules in the "descriptor" space to the differences in biological/ pharmacological activity of the compared molecules. In the following section it is shown how the partitioning of the real space of a molecule brought about by QT-AIM can contribute to the study of molecular similarity.

6.7.2 The Quantum Theory of Atoms in Molecules and Molecular Similarity

Two objects are similar and they possess similar properties to the extent that they have similar distributions of charge in real space. Thus chemical similarity should be defined and determined using the atoms of QT-AIM whose properties are directly determined by their spatial charge distributions. (*Bader et al., 1992*) Present measures of molecular similarity are couched in terms of Carbó's molecular similarity index (MSI), (*Carbó-Dorca et al., 2000*) which is defined: (*Carbó et al., 1980*)

$$R_{AB} = \frac{\int \rho_A \rho_B \, dv}{\sqrt{\int \rho_A^2 \, dv \int \rho_B^2 \, dv}}, \tag{6.1}$$

where ρ_A and ρ_B stand for the density of molecules A and B respectively, and the integral is over the *entire space*, a procedure that requires the maximisation of the spatial integration of the overlap of the density distributions of two molecules whose similarity is to be determined. The MSI suffers from the dependence on the particular choice of superposition of the molecular geometries (alignment problem), (*O'Brien and Popelier, 2001*) the time and cost of the integration of products of molecular densities, the dominance of the core densities in determining the degree of overlap of ρ_A and ρ_B, and finally, there is the question of how to compare only a region of interest extracted from A and B as opposed to comparing the entire densities. In order to isolate a region – e.g. a pharmacophore – for comparison in a series of related molecules, it has been proposed to modify the above integration to be performed over the region of interest determined as a proper open system. Eq. 6.1 is modified accordingly for an open system: (*Matta, 2001c*)

$$R_{\Omega \cup \Omega'} = \frac{\int_{\Omega \cup \Omega'} \rho_\Omega \rho_{\Omega'} \, dv}{\sqrt{\int_\Omega \rho_\Omega^2 \, dv \int_{\Omega'} \rho_{\Omega'}^2 \, dv}} \tag{6.2}$$

where $\cup$ indicates that the integration is carried out only over the volume necessary to enclose both fragments, each bounded by a zero-flux surface, and no longer over the entire space. Such an index would provide more pharmacologically-relevant comparisons than the whole space integration in Eq. 6.1, yet it still suffers from an alignment problem and remains dominated by the nuclear maxima.

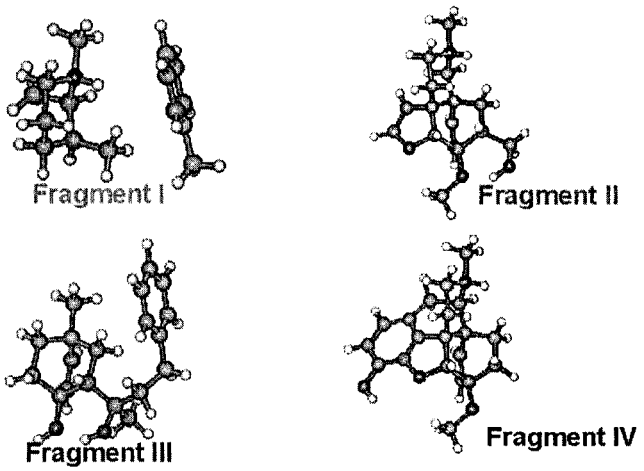


Figure 6.2. Fragment molecule I (Green): Mimics the electron density around the quaternized nitrogen atom and its bonded hydrogen atom. It is a dimer of two separate molecules, one being a toluene molecule. This dimeric fragment molecule is designed to reproduce the electronic environment around the N-(43)-H(44) ...C(60) hydrogen bond. **Fragment molecule II (Red):** Mimics the density around the hydrogen bond between the side-chain -O(49)H(50) group and the methoxy oxygen O(27). **Fragment molecule III (Blue):** Mimics the density of the side-chain and the lower half of the phenyl ring. **Fragment molecule IV (Black):** Mimics the density of the phenolic ring, the ether oxygen O(11), and the rest of the atoms on this side of the molecule. The color-code in this figure corresponds to that in figure 6.3.

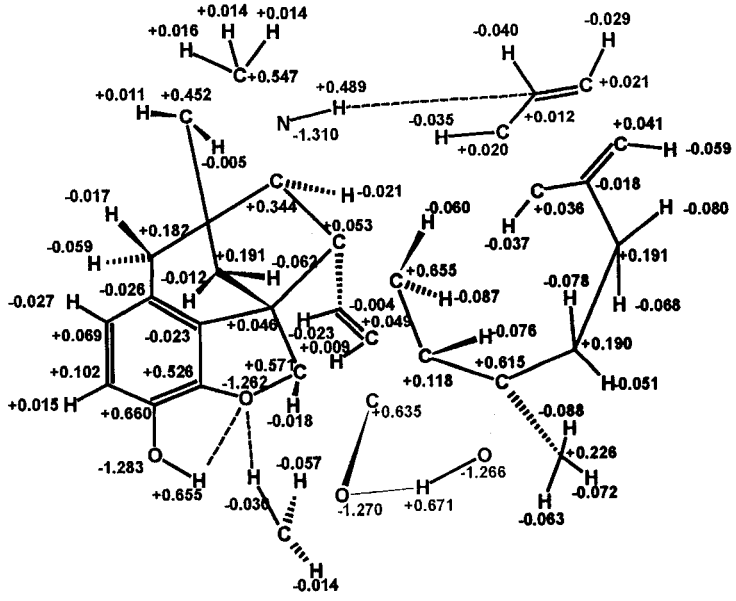


Figure 6.3. Atomic charges of quaternized PEO obtained from the four fragments defined in figure 6.2. The color code identifies the fragment used to generate the atomic charges (and the other atomic properties). The colour code is the same as in figure 6.2: *Green (Fragment I), Red (Fragment II), Blue (Fragment III),* and *Black(Fragment IV)*

Table 6.1 Some atomic properties of PEO from fragments (values in au)*

Ω	$N(\Omega)$						$N(\Omega)$	$E(\Omega)$	$v(\Omega)$	$L(\Omega)$
	intact PEO	frag I	frag II	frag III	frag IV	frag used	assemb PEO	assemb PEO	assemb PEO	assemb PEO
H(1)	0.3466				0.3451	IV	0.3451	-0.31762	17.17	0.00001
O(2)	9.2823				9.2825	IV	9.2825	-75.37406	117.47	-0.00021
C(3)					5.3396	IV	5.3396	-37.44918	57.39	0.00013
C(4)					5.8983	IV	5.8983	-37.79589	78.58	-0.00082
H(5)					0.9854	IV	0.9854	-0.62911	45.97	0.00007
C(6)					5.9312	IV	5.9312	-37.79806	78.59	0.00024
H(7)					1.0268	IV	1.0268	-0.64656	48.49	0.00008
C(8)	6.0245				6.0258	IV	6.0258	-37.88816	66.64	0.00013
C(9)			6.0062		6.0230	IV	6.0230	-38.04155	58.32	0.00015
C(10)			5.3911		5.4739	IV	5.4739	-37.60968	58.13	0.00044
O(11)	9.2642		9.2528		9.2618	IV	9.2618	-75.36266	97.71	0.00013
C(12)			5.4221	5.8120	5.4291	IV	5.4291	-37.50417	39.74	-0.00181
C(13)	5.9526	5.8250	5.9284	5.8299	5.9545	IV	5.9545	-37.88836	39.99	0.00028
C(14)		5.7789			5.8185	IV	5.8185	-37.71727	52.62	0.00192
H(15)		1.0083			1.0170	IV	1.0170	-0.64547	46.41	0.00012
H(16)		1.0564			1.0587	IV	1.0587	-0.66785	45.16	0.00012
C(17)		5.6552	5.5849	5.7775	5.6558	IV	5.6558	-37.60724	43.52	0.00219
H(18)		1.0251	1.0111	1.0848	1.0205	IV	1.0205	-0.65661	45.66	0.00015
C(19)		5.8601	5.9433	5.8963	5.9470	IV	5.9470	-37.84567	40.41	0.00007
C(20)		5.7698	5.8324	5.8143	5.8300	III	5.8143	-37.74292	48.42	0.00228
H(21)		1.0889	1.1293	1.0598	1.1317	III	1.0598	-0.67193	44.51	0.00006
H(22)		1.0526	1.0563	1.0865	1.0521	III	1.0865	-0.68778	43.08	0.00000
C(23)			5.8829	5.8824	5.8203	III	5.8824	-37.77245	43.59	0.00361
H(24)			1.0619	1.0756	1.0774	III	1.0756	-0.68487	44.91	0.00024
C(25)	5.3719		5.2469	5.3855		III	5.3855	-37.37188	35.85	0.00032
C(26)			5.3652	5.4220	5.3328	II	5.3652	-37.42885	35.12	-0.00031
O(27)	9.2738		9.2700	9.3012	9.2473	II	9.2700	-75.39442	89.63	-0.00144
C(28)			5.2700		5.2520	IV	5.2520	-37.35735	53.86	0.00084
H(29)			1.0453		1.0566	IV	1.0566	-0.66662	48.72	0.00011
H(30)			1.0095		1.0303	IV	1.0303	-0.66181	45.34	-0.00023
H(31)			1.0117		1.0143	IV	1.0143	-0.64756	47.42	0.00006
H(32)			1.0129	1.0789	1.0181	IV	1.0181	-0.66064	44.03	0.00014
C(33)			5.9526	5.8061	5.9515	IV	5.9515	-37.88178	73.02	0.00031
H(34)			0.9941	1.0484	0.9915	IV	0.9915	-0.63191	45.66	0.00012
C(35)	5.9966		5.9976	5.9936	6.0042	IV	6.0042	-37.89329	71.43	0.00057
H(36)			1.0212	1.0456	1.0228	IV	1.0228	-0.64453	47.38	0.00011
C(37)		5.8043	5.8118		5.8089	IV	5.8089	-37.75720	50.87	0.00247
H(38)		1.0566	1.0618		1.0622	IV	1.0622	-0.66632	46.64	0.00023
H(39)		1.0239	1.0094		1.0123	IV	1.0123	-0.64011	46.85	0.00010
C(40)	5.5435	5.5404	5.5446		5.5479	IV	5.5479	-37.57651	49.82	0.00152
H(41)		1.0109	1.0041		0.9889	IV	0.9889	-0.64156	40.94	0.00002
H(42)		1.0077	1.0018		1.0054	IV	1.0054	-0.64366	46.00	0.00009
N(43)	8.3120	8.3099	8.2986		8.3026	I	8.3099	-55.01652	63.29	0.00018
H(44)	0.5090	0.5112	0.5253		0.5300	I	0.5112	-0.41237	22.67	0.00000
C(45)	5.4495	5.4486	5.4536		5.4527	IV	5.4527	-37.51104	58.19	0.00175
H(46)	0.9941	0.9954	0.9829		0.9837	IV	0.9837	-0.63340	43.07	0.00013
H(47)	0.9880	0.9891	0.9840		0.9864	IV	0.9864	-0.63068	45.83	0.00007
H(48)	0.9851	0.9872	0.9827		0.9858	IV	0.9858	-0.63045	45.56	0.00008
O(49)			9.2655	9.3017		II	9.2655	-75.33416	114.30	0.00010
H(50)	0.3253		0.3289	0.3293		II	0.3289	-0.31196	11.37	-0.00099
C(51)				5.7736		III	5.7736	-37.71143	57.76	0.00137

Table 6.1 (Continued)

Ω	$N(\Omega)$					frag	$N(\Omega)$	$E(\Omega)$	v(Ω)	$L(\Omega)$
	intact PEO	frag I	frag II	frag III	frag IV	used	assemb PEO	assemb PEO	assemb PEO	assemb PEO
H(52)				1.0625		III	1.0625	-0.66319	46.02	0.00003
H(53)				1.0875		III	1.0875	-0.66971	48.99	-0.00085
H(54)				1.0721		III	1.0721	-0.66157	50.82	0.00011
C(55)				5.8098		III	5.8098	-37.72067	50.80	0.00183
H(56)				1.0509		III	1.0509	-0.66734	46.24	0.00014
H(57)				1.0784		III	1.0784	-0.67468	46.63	0.00002
C(58)		5.7717		5.8090		III	5.8090	-37.69094	53.41	0.00230
H(59)		1.0451		1.0798		III	1.0798	-0.67528	47.02	0.00002
H(60)		1.0317		1.0675		III	1.0675	-0.66689	49.83	0.00010
C(61)	6.0231	6.0139		6.0177		III	6.0177	-37.86771	61.43	0.00015
C(62)		5.9632		5.9591		III	5.9591	-37.82078	77.01	0.00065
H(63)		1.0316		1.0592		III	1.0592	-0.66183	49.47	0.00011
C(64)		5.9786		5.9589		I	5.9786	-37.81338	80.36	0.00028
H(65)		1.0287		1.0493		I	1.0287	-0.64767	48.21	0.00007
C(66)	5.9897	5.9885		5.9546		I	5.9885	-37.81426	79.48	0.00068
H(67)		1.0399		1.0495		I	1.0399	-0.65170	49.50	0.00008
C(68)		5.9805		5.9583		I	5.9805	-37.81344	80.82	-0.00030
H(69)		1.0351		1.0491		I	1.0351	-0.64991	49.08	0.00008
C(70)		5.9640		5.9644		III	5.9644	-37.82527	75.78	0.00029
H(71)		1.0368		1.0523		III	1.0523	-0.66146	47.55	0.00000
SUMS							254.0199	-1510.68084	3821.54	0.02308

* Fragments are labelled as in figure 6.2. "Frag. used" indicates which fragment provided atom (Ω) used in the assembled PEO, $N(\Omega)$ is the atomic electron population (Eq. 2.62), v(Ω) is the atomic volume, $E(\Omega)$ is the atomic energy (Eq. 2.67), and $L(\Omega)$ is the integrated Laplacian (Eq. 2.66).

Table 6.2 Comparison of properties for some atoms in intact and assembled PEO

Ω	$q(\Omega)$ intact	$q(\Omega)$ assemb	$\|\mu(\Omega)\|$ intact	$\|\mu(\Omega)\|$ assemb	$\|Q(\Omega)\|$ intact	$\|Q(\Omega)\|$ assemb	$E(\Omega)$ intact	$E(\Omega)$ assemb	v(Ω) intact	v(Ω) assemb
H(1)	0.653	0.655	0.133	0.133	0.042	0.042	-0.3185	-0.3176	17.3	17.2
O(2)	-1.282	-1.282	0.330	0.332	0.806	0.804	-75.3652	-75.3741	117.3	117.5
C(8)	-0.025	-0.026	0.150	0.152	3.179	3.193	-37.8827	-37.8882	66.6	66.6
O(11)	-1.264	-1.262	0.234	0.238	0.884	0.880	-75.3553	-75.3627	97.8	97.7
C(13)	0.047	0.045	0.044	0.036	0.277	0.304	-37.8814	-37.8884	40.0	40.0
C(25)	0.628	0.614	0.661	0.663	0.883	0.895	-37.3737	-37.3719	35.6	35.8
O(27)	-1.274	-1.270	0.196	0.205	1.001	0.995	-75.3634	-75.3944	88.0	89.6
C(35)	0.003	-0.004	0.075	0.096	2.942	3.011	-37.8848	-37.8933	70.9	71.4
C(40)	0.456	0.452	0.522	0.526	1.120	1.151	-37.5684	-37.5765	49.9	49.8
N(43)	-1.312	-1.310	0.026	0.022	0.209	0.221	-55.0455	-55.0165	63.3	63.3
H(44)	0.491	0.489	0.158	0.159	0.077	0.078	-0.4110	-0.4124	22.5	22.7
C(45)	0.551	0.547	0.491	0.485	1.046	1.077	-37.5039	-37.5110	57.8	58.2
H(46)	0.006	0.016	0.098	0.098	0.422	0.419	-0.6383	-0.6334	43.6	43.1
H(47)	0.012	0.014	0.103	0.103	0.414	0.422	-0.6311	-0.6307	46.6	45.8
H(48)	0.015	0.014	0.102	0.102	0.418	0.422	-0.6299	-0.6305	46.5	45.6

Table 6.2 (Continued)

Ω	$q(\Omega)$ intact	$q(\Omega)$ assemb	$\|\mu(\Omega)\|$ intact	$\|\mu(\Omega)\|$ assemb	$\|Q(\Omega)\|$ intact	$\|Q(\Omega)\|$ assemb	$E(\Omega)$ intact	$E(\Omega)$ assemb	$v(\Omega)$ intact	$v(\Omega)$ assemb
H(50)	0.675	0.671	0.112	0.113	0.072	0.074	-0.3106	-0.3120	11.0	11.4
C(61)	-0.023	-0.018	0.108	0.081	2.492	2.555	-37.8847	-37.8677	61.4	61.4
C(66)	0.010	0.012	0.090	0.088	3.479	3.461	-37.8320	-37.8143	79.7	79.5
av|dev|*		0.004		0.005		0.017		0.0088		0.3

* av|dev| is the average magnitude of the difference between intact and assembled atomic property.

Table 6.3 Comparison of some molecular properties of intact PEO versus those obtained from fragments (values in au unless indicated otherwise)

molecular property	intact PEO (1)	assemb PEO (2)	difference (2)-(1)	%Error
Total number of electrons	254	254.0199	+0.0199	+0.01
Total charge	+1	+0.9801	-0.0199	-1.99
Total energy	-1510.703254	-1510.680842	+14.1kcal/mol	+0.001
Total volume	565.57Å^3	566.29Å^3	+0.73Å^3	+0.13

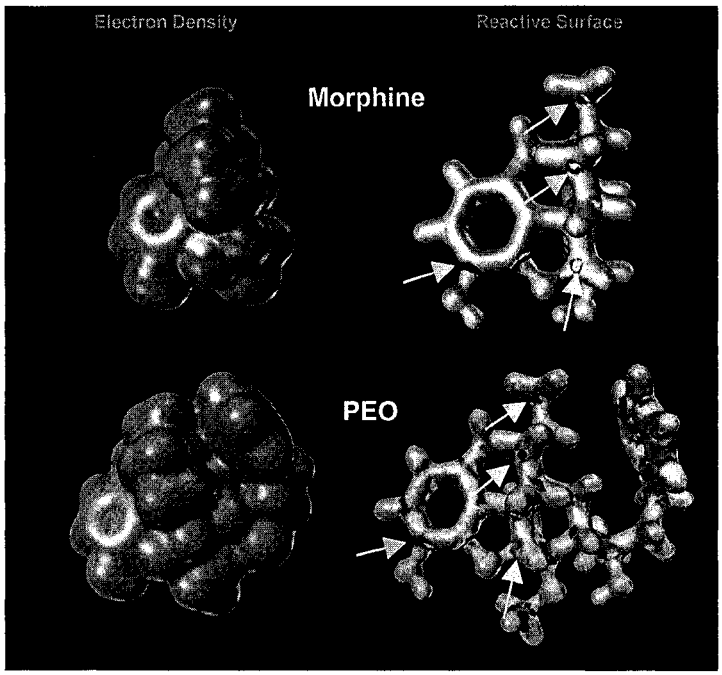


Figure 6.4. Electron density 0.002 au envelopes of morphine and PEO (left), and their outer zero-Laplacian envelope ($\nabla^2\rho(\mathbf{r})=0$), also known as the reactive surface, (right). The valence shell charge concentration (VSCC) exhibits lumps (regions of local charge concentration) and holes (regions of local charge depletion). A chemical reaction corresponds to the combination of a "lump" in the VSCC of the base with the "hole" in the VSCC of the acid. Similar regions of the morphine and of the PEO molecules have similar regions of local charge depletion and concentration (some regions of charge depletion are indicated by the arrows).

These problems are avoided when the atomic (*Matta and Bader, 2002b*) or BCP (*O'Brien and Popelier, 2001*) properties of QT-AIM are used in similarity or QSAR studies. The density of a particular grouping of atoms embedded in two different molecules can exhibit striking visual similarities, which are also reflected in displays of the Laplacian. As an example, the reader is asked to compare the isodensity envelope plots of $\rho(\mathbf{r})$ and of $\nabla^2\rho(\mathbf{r})$ of morphine and of PEO displayed in figure 6.4. The phenolic region, the region surrounding the quaternized nitrogen atom, and the 5-MR furan ring region look very similar in these two molecules. Such a qualitative similarity can be expressed quantitatively if the functional group of interest is "extracted" from the molecule at its bounding zero-flux surface. In figure 6.5, a comparative study of the charge distributions of the tyrosine residue in enkephalin, of morphine, and of PEO, serves as an example of the determination of similarity afforded by the use of the atoms of QT-AIM. (*Matta, 2001c*) The similarity of the phenolic moieties of enkephalin, PEO, and morphine is maximized when the charge distributions are compared in terms of the atomic boundaries and bond paths. The phenolic moieties in these three different molecules serve to illustrate how a set of connected zero-flux surfaces can be found to isolate within each molecule the region that exhibits maximum similarity accross the three molecules. (Figure 6.5). The apparent similarity of the atoms within this region in morphine and PEO is quantified by the corresponding high degree of similarity in the atomic properties, as summarized in table 6.4 which lists the charge $q(\Omega)$, the magnitude of the first moment $|\boldsymbol{\mu}(\Omega)|$, the magnitude of the quadrupolar polarisation $|\mathbf{Q}(\Omega)|$, the energy $E(\Omega)$ and volume $v(\Omega)$ for each atom Ω. Rather than assign a single number as a measure of the degree of similarity, one instead relates any observed differences in the physiological activities of the two molecules directly to the residual differences in their local properties. Thus the atoms of QT-AIM, in addition to identifying a common active site within a set of molecules, provide a maximal quantitative comparison of their properties. Figure 6.6 illustrates the atomic charge distributions in assembeled PEO, enkephalins, and morphine.

6.7.3 Sources of Error in the Buffered Fragments Approach

The use of truncated basis sets in SCF calculations leads to a basis set superposition error (BSSE). (*Boys and Bernardi, 1970; Liu and McLean, 1973; van Duijneveldt et al., 1994*) When the energy of

monomers *A* and *B* are calculated separately, the effective basis set used is smaller than the one used to estimate the energy of the complex *AB*. This results in an artificial lowering of the energy of the complex. In the fragment approach an intramolecular BSSE arises. When assembling a molecule from isolated fragments, one is assembling the total system from subsystems which are effectively described in terms of a smaller basis set than the one used to describe the intact molecule.

The inclusion of large buffering regions surrounding each fragment will reduce the imperfect transferability error, while the use of large basis sets reduces BSSE which is rooted in the basis set truncation. Minimizing the numerical integration errors is subject to a compromise between the required accuracy and the computational time and costs.

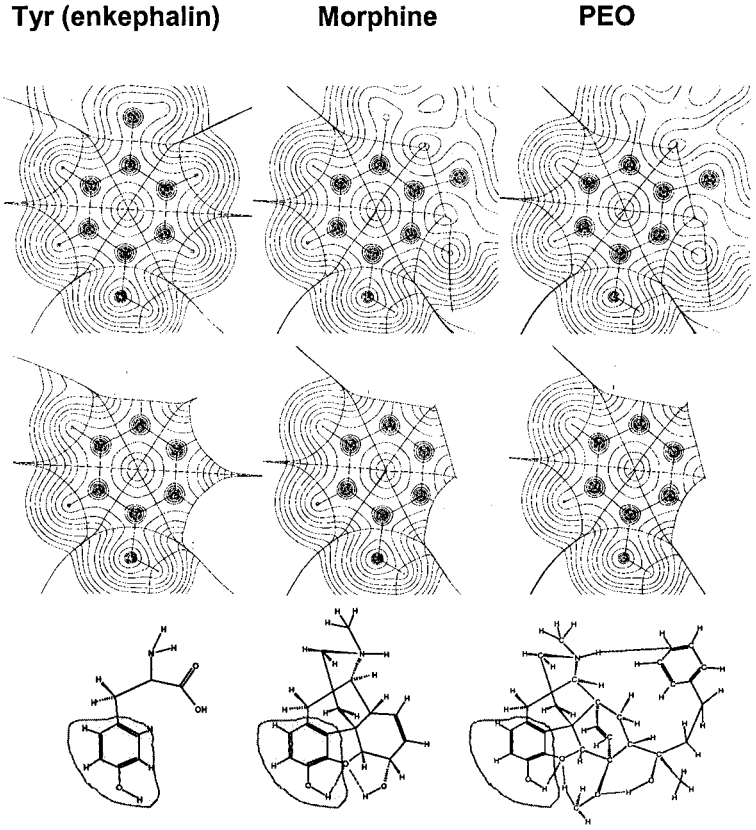


Figure 6.5. The top row illustrates the electron densities for portions of the enkephalin precursor tyrosine, for morphine, and for PEO. Note, as demonstrated in the second row, how the boundaries of the atoms enable one to uniquely isolate and define, as a bounded region of real space, the group that exhibits maximum similarity in all three molecules. The diargam at the bottom serves to locate this region in the three molecules.

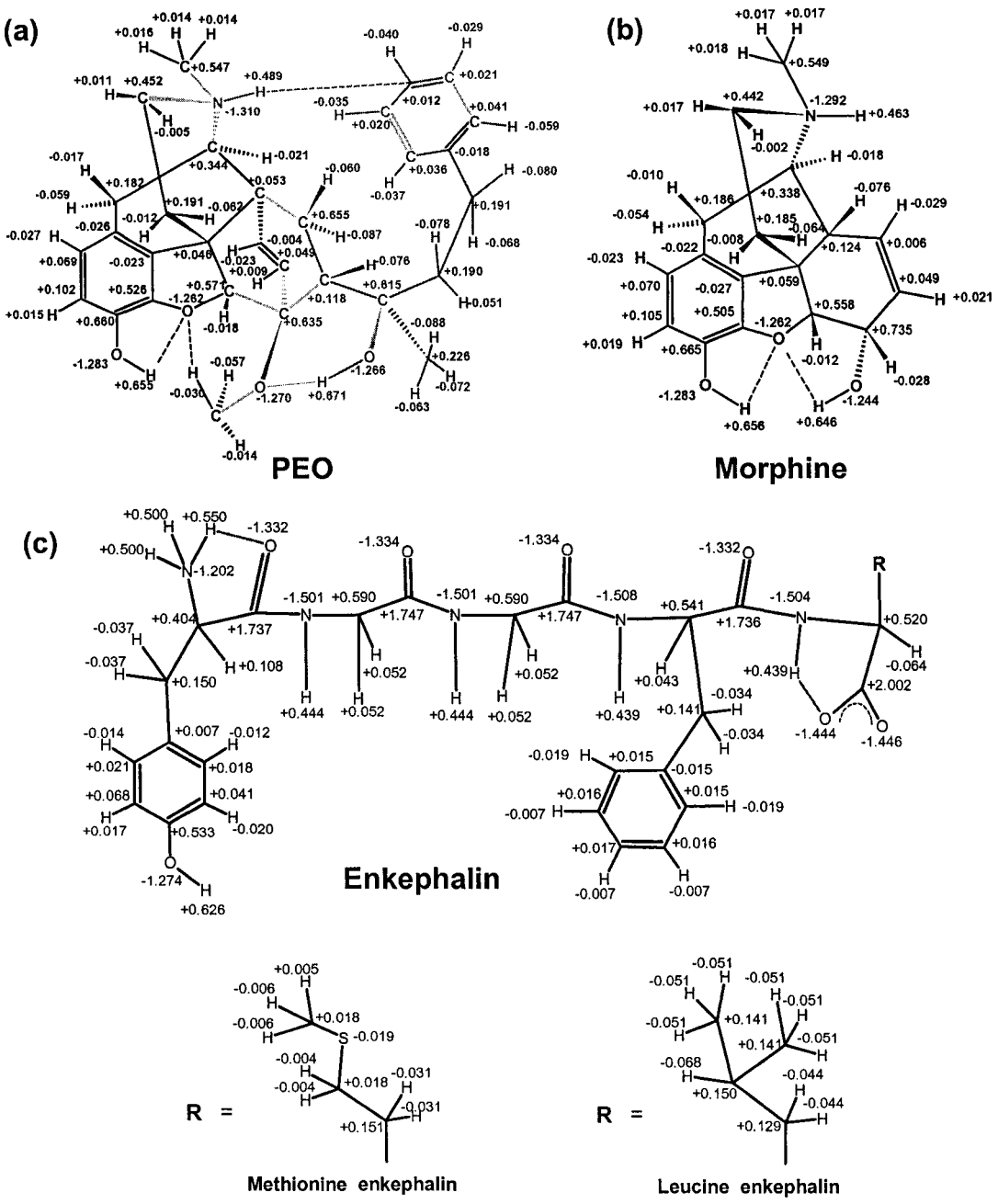


Figure 6.6. Atomic charges of quaternized PEO, quaternized morphine, and zwitterionic enkephalins. **(a)** Atomic charges of PEO. **(b)** Atomic charges of morphine (the sum of the atomic charges is 0.995 au). **(c)** Atomic charges of enkephalins (the sum of the atomic charges are +0.051 au and +0.043 au for Leu- and Met-enkephalin respectively). Note the similarity of the charges of equivalent atoms in these molecules.

Table 6.4 Comparison of atomic properties for maximal transferable atoms in morphine and PEO.*

Ω	$q(\Omega)$	$q(\Omega)$	$\|\mu(\Omega)\|$	$\|\mu(\Omega)\|$	$\|Q(\Omega)\|$	$\|Q(\Omega)\|$	$E(\Omega)$	$E(\Omega)$	$v(\Omega)$	$v(\Omega)$
	PEO	Morph	PEO	Morph	PEO	Morph	PEO	Morph	PEO	Morph
H(1)	0.655	0.656	0.133	0.133	0.042	0.043	-0.3176	-0.3171	17.2	17.2
O(2)	-1.283	-1.283	0.332	0.334	0.804	0.806	-75.3741	-75.3874	117.5	117.4
C(3)	0.660	0.665	0.758	0.760	3.135	3.128	-37.4492	-37.4509	57.4	57.4
C(4)	0.102	0.105	0.157	0.160	3.368	3.369	-37.7959	-37.8024	78.6	78.2
H(5)	0.015	0.019	0.098	0.098	0.454	0.452	-0.6291	-0.6274	46.0	45.7
C(6)	0.069	0.070	0.049	0.052	3.381	3.365	-37.7981	-37.8041	78.6	78.6
H(7)	-0.027	-0.023	0.098	0.097	0.480	0.479	-0.6466	-0.6449	48.5	48.3
C(8)	-0.026	-0.022	0.152	0.149	3.193	3.173	-37.8882	-37.8987	66.6	66.2
C(9)	-0.023	-0.027	0.187	0.158	2.987	3.062	-38.0415	-38.0415	58.3	58.0
C(10)	0.526	0.505	0.731	0.727	3.340	3.410	-37.6097	-37.6250	58.1	59.4
O(11)	-1.262	-1.253	0.238	0.242	0.880	0.833	-75.3627	-75.3687	97.7	97.0
C(12)	0.571	0.558	0.597	0.607	1.148	0.919	-37.5042	-37.4946	39.7	42.9
C(13)	0.046	0.059	0.036	0.048	0.304	0.393	-37.8884	-37.8550	40.0	41.4
C(14)	0.182	0.186	0.096	0.097	0.382	0.431	-37.7173	-37.7305	52.6	53.1
H(15)	-0.059	-0.010	0.105	0.105	0.401	0.401	-0.6455	-0.6416	46.4	46.5
H(16)	-0.017	-0.054	0.096	0.098	0.450	0.449	-0.6679	-0.6635	45.2	46.0
N(43)	-1.310	-1.292	0.022	0.041	0.221	0.301	-55.0165	-55.0374	63.3	63.8
H(44)	0.489	0.463	0.159	0.166	0.078	0.050	-0.4124	-0.4295	22.7	25.6
C(45)	0.547	0.541	0.485	0.490	1.041	1.097	-37.5110	-37.5141	58.2	58.9
av|dev|		0.012		0.006		0.041		0.009		0.7

* Numbering scheme is the same for these atoms in PEO and in morphine.

6.7.4 The biologically-active conformations of PEO and enkephalin.

The potential energy scan described in the computational section yielded, besides the global minimum **A**, two local minima which are labeled **B** and **C** in order of increasing energy. As suggested in (*Loew et al., 1991*), an important criterion for candidacy as a biologically-active conformation of enkephalin is the coincidence of the geometric relationships of certain functional groups with those between corresponding groups in the PEO or in the PET molecules. Following this suggestion, the x-ray crystallographic geometry of Leu-enkephalin was obtained from the *Cambridge Crystallographic Data Centre (U.K.)*. The geometry of enkephalin was visually modified using a molecular modeller to maximise the superimposition on PEO (**A**, **B**, and **C** geometries) while preserving the peptide bond geometries. Three geometries for enkephalin were obtained in this manner, each bearing a close geometric resemblance to a particular conformation of PEO (see figure 6.7). A future search for

geometric coincidence between each of these geometric pairs with known specific opioid receptor subtype agonists may help in assigning one geometric pair to each of the three receptor subtypes (μ, κ, and δ). The geometries of the three candidate biologically-active PEO/enkephalin conformation pairs are summarised in table 6.5 and their geometric coincidence illustrated in figure 6.7.

The scorpion conformation (**A**) of PEO allows the hydrogen bonding between the proton bonded to the quaternary nitrogen [H(44)] and the *para*-carbon atom of the phenyl on the side-chain. Similar hydrogen bonding been aromatic rings and proton donors has been reported for several systems and was found to play a crucial role in stabilizing local structures in proteins. [See for example: (*Chelli et al., 2002; Desiraju and Steiner, 1999; Steiner, 2002; Steiner and Koellner, 2001*), and references therein]. Since the atomic properties of the quaternized nitrogen atom and its attached hydrogen atom in enkephalin are very similar to those of the N(43)-H(44) group in PEO, a similar hydrogen bonding is anticipated in enkephalin when the Tyr-NH_3^+ group adopts a similar geometrical arrangement of with respect to the Phe aromatic ring. In PEO, $d_{C(66)\text{-}H(44)} = 2.93$Å, $d_{H(44)\text{-}N(43)} = 1.02$Å, and $\triangle_{N(34)\text{-}H(44)\text{-}C(66)} = 137.9°$. In enkephalin, when the Tyr-NH_3^+ group is rotated so that $\triangle_{N(Tyr)\text{-}H(Tyr)\text{-}C(Phe)} = 137.9°$**A**, the two other parameters also show a remarkable similarity to the corresponding values in PEO, these are: $d_{p\text{-}C(Phe)\text{-}H(Tyr)} = 3.15$Å and $d_{H(44)\text{-}N(43)} = 1.02$Å.

The close proximity of the Tyr and Phe aromatic ring of Leu-enkephalin in a tightly folded structure has been observed experimentally where the centroids of the two rings are separated by only 5.0Å in a nearly orthogonal orientation. (*Aubry et al., 1989*) Furthermore, the significance of hydrogen bonding between aromatic rings of Phe, Tyr and Trp residues and peptide backbone -NH groups in stabilising the peptide secondary structures has been recently emphasised. (*Tóth et al., 2001*) In their paper, Tóth *et al.* analysed 560 different proteins crystallographic structures statistically, and they reported a clustering of the aromatic side-chain angle χ1 around the range -50° to -100° for the hydrogen-bonds of the type Ar(*i*)-HN(*i*-3), an angle which in enkephalin **A** has the value of -40°. They also showed a significantly high propensity for the residue *i*-2 to be Gly, which is of course the case in both types of enkephalins. These considerations add weight to the proposed tightly folded biological

conformation **A** of enkephalins, a conformation which could be used as a reasonable initial guess for a full geometry optimisation in the future.

6.8 Conclusions

The feasibility of the theoretical assembling of large complicated molecules has been established and shown to provide an excellent approximation to the density and to the properties of the intact molecule. Thus, one can obtain accurate properties of large molecules without the need to run a direct SCF on the intact molecule, and instead run much faster calculations on model fragment molecules. The success of the method is the result of the simultaneous transferability as well as the additivity of the atomic properties, conditions that are closely met when the fragments are determined as proper open systems according to QT-AIM.

The definition of a pharmacophore as a proper open quantum system entails a modification of Carbó's similarity index to apply to a part of interest within a molecule, extracted at the zero-flux surface, as opposed to comparing the entire molecular densities. Even though this proposed mode of integration allows the comparison to focus on the biologically relevant parts of two molecules, it still suffers from the same alignment problem as well as from the dominant contribution arising from the cusps in the electron density at the position of the nuclei. The use of the atomic and bond properties derived from QT-AIM is a much richer alternative for the comparison of two molecules, or parts of two molecules, since one obtains a whole set of properties rather than just one number to define similarity. QT-AIM properties allow one to compare the two systems at any desired degree of resolution or detail.

Three different conformations for PEO have been identified and found to bear a striking geometric similarity to three corresponding enkephalin conformations. These three conformation pairs can be good candidates for biologically-active geometries, but further evidence is needed.

Table 6.5 Key geometric parameters defining three nearly coincident conformations (**A**, **B**, and **C**) of PEO and of enkephalins* (angles in (°) and distances in Å)

		A	**B**	**C**
PEO				
	$\theta 1$	-84	-84	-144
	$\theta 2$	93	121	93
	$\theta 3$	44	-16	44
	τ	77	64	89
	d	8.01	10.19	9.66
Enkephalins				
Tyr [1]	ϕ	73	73	73
	ψ	-97	-97	-97
	ω	-175	-175	-175
	$\chi 1$	178	178	178
	$\chi 2$	46	46	46
	$\chi 3$	177	177	177
Gly [2]	ϕ	-151	-151	-151
	ψ	27	27	27
	ω	-164	-164	-164
Gly [3]	ϕ	-140	-140	-140
	ψ	-21	-21	-84
	ω	-170	-170	-170
Phe [4]	ϕ	-82	-82	-82
	ψ	5	5	5
	ω	179	179	179
	$\chi 1$	-40	-100	93
	$\chi 2$	-1	-1	11
	$\chi 3$	179	179	179
Met or Leu [5]	ϕ	-73	-73	-73
	ψ	-39	-39	-39
	τ	62	108	104
	d	7.66	10.02	10.24

* Conformations **A**, **B**, and **C** are defined in the text, they correspond to those in figure 6.7. Side-chain angles for Met or Leu are omitted since their conformation was not studied in detail. The angles θ1-3 are defined in the text and in figure 6.1; peptide torsion angles are named in accordance with IUPAC-IUB conventions. τ is the dihedral angle between the two aromatic ring planes and d the distance between their centroids.

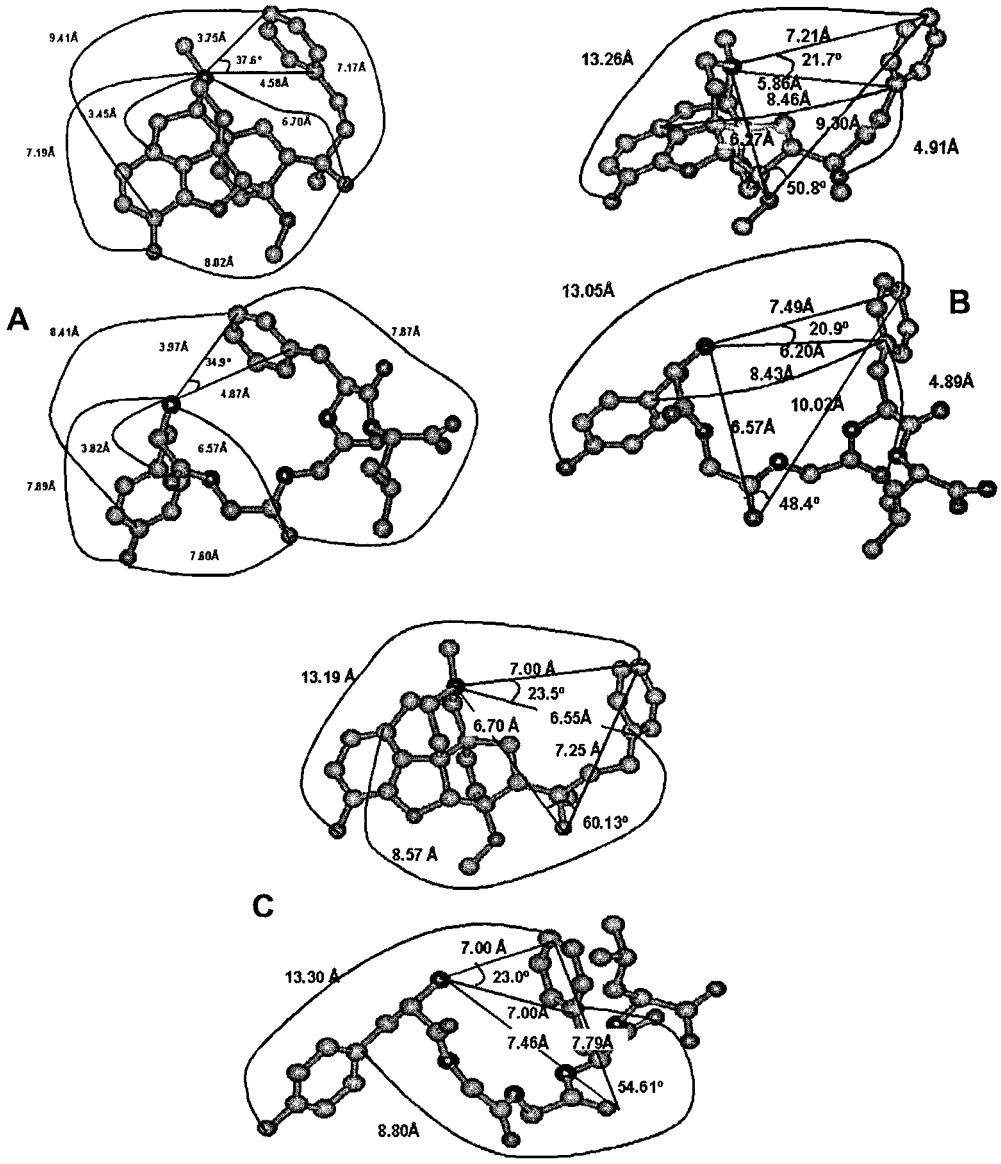


Figure 6.7. Three coincident conformations (**A**, **B**, and **C**) of PEO (top) and Leu-enkephalin (bottom). Some striking geometrical coincidences are indicated. These are candidates for three biologically active conformations of opioids. (Distances in Å, angles in (°)).

Chapter 7

Properties of Atoms in Crystals and Extended Systems: Dielectric Polarization

7.1 Statement of the Problem

It has been previously shown that a change in the polarization of a dielectric can be considered a consequence of a geometric quantum phase and obtainable from a Berry phase in a parameter space. It is the purpose of this chapter to demonstrate that such a possibility does not exclude a real space description, stated in terms of the charge distributions of the system's composite atoms or cells. It is shown that the polarization of a molecule or an extended system, permanent or induced can, like all measurable properties, be equated to a sum of atomic contributions (Eq. 2.49). The cells are defined as bounded regions of real space whose properties are described by the physics of a proper open system, a description that applies to any system regardless of the nature of the interactions between the atoms. This approach necessarily leads to the inclusion of a contribution to the polarization arising from the transfer of charge across the boundary of a cell, in addition to that from the cell's internal polarization, thereby correcting the textbook description of polarization that considers only the latter contribution. It is shown that a neutral repeating cell in a dielectric behaves as an atomic capacitor which mimics the macroscopic behaviour, with the internal charge transfer leading to the creation of what Feynman terms the 'surface polarization charge'. This chapter is based on the paper (*Bader and Matta, 2001b*)

7.2 Introduction

Martin (*Martin, 1974*) has shown that the polarization of a crystal cannot be derived solely in terms of the polarization density in a unit cell in an infinite periodic crystal, the procedure followed by Bennett and Maradudin (*Bennett and Maradudin, 1972*) in calculations of the transverse effective charges for crystals possessing the zinc-blende structures. Their approach is similar to the one presented in the textbook descriptions of polarization. (*Ashcroft and Mermin, 1976; Kittel, 1996*) Martin demonstrated that one must include a surface term that accounts for the charge transferred across the boundary of the cell, no matter how the cell itself is chosen or defined. He argues that it will always be simpler to utilize the long-wavelength limit of the finite-wave-vector expressions to calculate the electric polarization in a crystal and avoid what he considers to be an arbitrary division into cell and surface contributions.

The modern theory of polarization was developed by King-Smith and Vanderbilt (*King-Smith and Vanderbilt, 1993*) using an expression for the change in **P** with respect to a parameter derived heuristically by Resta. They implemented the expression through the introduction of an adiabatic change in the Kohn-Sham Hamiltonian (Eq. 1.55) of a solid, expressing it in terms of the valence-band wave functions of the initial and final Hamiltonians. They interpret **P** as being proportional to the displacement in the centre of charge of the Wannier functions induced by the adiabatic change in the Hamitonian and note that Resta's expression can be interpreted as the change in Berry phase for fictitious adiabatic evolution of the cell periodic wave function around a loop of a contour integration in parameter space. Resta (*Resta, 1993*) argued that the measurable change in the macroscopic polarization of a crystal **P**, between any two crystal states cannot be recovered from periodic charge densities wherein any phase information is deleted, arguing instead that **P** is a quantum phenomenon governed by a gauge invariant phase feature of the crystal wave functions. The work of these authors (*King-Smith and Vanderbilt, 1993; Resta, 1993*) shows that the observable **P**

can be considered to be a consequence of a geometric quantum phase and can be obtained from a Berry phase (*Berry, 1984*) in a particular parameter space.

Resta (*Resta, 1998*) has revisited these problems from another point of view by replacing the usual single-particle position operator with a many-body operator, relating this approach to the Berry phase theory of polarization. He has recently reviewed the observable consequences of the Berry phase of a many-electron wave function. (*Resta, 2000*)

It is the purpose of the present chapter to point out an alternative approach to the general problem of defining and calculating macroscopic polarization or magnetization of extended systems, one that uses the densities and currents associated with the individual periodic cells within a crystal whether they be 'ionic', 'covalent' or 'molecular'. This is accomplished using the physics of an open system, (*Bader, 1994*) an approach that leads naturally to the inclusion of surface terms that account for the transfer of charge or to the flow of current across cell or atomic boundaries. The contribution of a surface flux to the expectation value of an observable is the very property that distinguishes the physics of an open system from that of the isolated, total system of which it is a part. This physics, and the unique identification of an atom in a molecule or of a Wigner-Seitz cell in a periodic crystal with a bounded region of real space whose properties are described by the physics of a proper open system removes the objections raised by Martin in his critique of the cell approach. Ashcroft and Mermin (*Ashcroft and Mermin, 1976*) describe the Born-von Karman boundary condition which forms the basis for obtaining the wave function for a periodic lattice, as one in which the flux in the electrons through each face of a cell of macroscopic dimensions is balanced by a corresponding flux through an opposite face, thereby describing in effect, an open system. It will be argued that the physics of an open system provides a physically meaningful, if not the natural, approach to the physics of the solid state.

7.3 Theory: Definition of Charge Displacement for an Open System

The essential problem in defining macroscopic polarization or magnetization concerns the determination of the expectation value of the electronic position operator. This is equivalent to averaging of densities weighted by the electronic position operator **r**. No difficulties are encountered for a finite system where the wavefunction vanishes exponentially at infinity and the expectation value is given by:

$$\langle \mathbf{r} \rangle = \int d\tau \left[\sum_i^N \mathbf{r}_i \Psi^* \Psi \right] = \int d\mathbf{r}\ \mathbf{r}\, \rho(\mathbf{r}). \tag{7.1}$$

The displacement of the centroid of negative charge determined by Eq. 7.1 relative to the centroid of positive charge, both measured from a common origin, determines the dipole moment of a neutral molecule.

It is shown here that the expectation values of properties dependent upon the position operator are well-defined within the physics of an open system, electronic polarization by the density $\mathbf{r}\,\rho(\mathbf{r})$, magnetization by $\mathbf{r}\times\mathbf{J}(\mathbf{r})$, for both finite and extended systems. (*Bader and Keith, 1993*)

7.3.1 Systems with a Permanent Dipole

To express $\boldsymbol{\mu}$, the dipole moment of a molecule or the polarization of an extended system in terms of atomic or group contributions it is necessary to introduce a local coordinate system for each atom. One makes the substitution $\mathbf{r} = \mathbf{r}_A + \mathbf{X}_A$, in Eq. 7.2, where $\mathbf{r}_A$ is referenced to the nucleus of A with position coordinate $\mathbf{X}_A$ and charge Z_A,

$$\boldsymbol{\mu} = \int d\mathbf{r}\ \mathbf{r}\, \rho(\mathbf{r}) + \sum_A \mathbf{X}_A Z_A \tag{7.2}$$

to obtain Eq. 7.3 (*Bader, 1990*)

$$\boldsymbol{\mu} = \sum_A \left[-\int_A d\mathbf{r}\ \mathbf{r}_A\, \rho(\mathbf{r}) + \mathbf{X}_A \left(Z_A - \int_A d\mathbf{r}\ \rho(\mathbf{r}) \right) \right]$$

$$\boldsymbol{\mu} = \sum_{A} \left[\mathbf{M}(A) + \mathbf{X}_A q(A) \right], \tag{7.3}$$

where the net charge on atom A is given by Eqs. 2.62 and 2.63. The dipole moment is expressed as a sum of atomic dipolar polarizations (Eq. 2.68), which is given the symbol ($\mathbf{M}(A)$) in this chapter, plus a term resulting from the transfer of electronic charge between the atoms. Each charge transfer term $\mathbf{X}_A q(A)$ depends upon the choice of origin, a dependence removed by introducing terms that take into account the explicit transfer of charge from an atom to each of its bonded neighbours, that is to each group with which it shares an interatomic surface. (*Bader et al., 2000; Bader and Keith, 1993*) As a result, the molecular dipole moment $\boldsymbol{\mu}$ is expressible as a sum of atomic contributions ($\boldsymbol{\mu}$(A)) defined by:

$$\boldsymbol{\mu} = \sum_{A} \boldsymbol{\mu}(A) = \sum_{A} \left\{ \mathbf{M}(A) + \sum_{B \neq A} \left[\mathbf{X}^c(A \mid B) - \mathbf{X}_A \right] q(B) \right\}, \tag{7.4}$$

where $\mathbf{X}^c(A|B)$ is the coordinate of the bond cp whose associated trajectories of $\nabla\rho(\mathbf{r})$ define the interatomic surface $S(A|B)$, where B runs over all of the atoms or groups linked to and sharing a surface with atom A. Since the system is electrically neutral, one has $q(A) = -\sum_{B \neq A} q(B)$ and because of this, the term $-\sum_{A} \sum_{B \neq A} \mathbf{X}_A q(B) = \sum_{A} \mathbf{X}_A q(A)$ in Eq. 7.3 and the second product of sums in Eq. 7.4 $[\sum_{A} \sum_{B \neq A} \mathbf{X}^c(A \mid B) q(B)]$ vanishes. The contributions to the sum $\sum_{B \neq A}$ describe the manner in which the electronic charge lost or gained by A is shared between its bonded neighbours B. This term when summed over all the atoms, yields the charge transfer term $\sum_{A} \mathbf{X}_A q(A)$ in Eq. 7.3. The dependence of $\boldsymbol{\mu}(A)$ on the charges of neighbouring groups, $q(B)$, can be removed by replacing each $q(B)$ of Eq. 7.4 with the corresponding interatomic surface integral of the flux in the total electrostatic field $\mathbf{E}(\mathbf{r})$ due to the electronic as well as to the nuclear charge contained within the boundaries of the atom.

The molecular electrostatic field $\mathbf{E}(\mathbf{r})$ is related to the total charge density $\rho(\mathbf{r})$ by Maxwell's equation, the differential form of Gauss's law of electrostatics:

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = 4\pi\ \rho(\mathbf{r})\,. \tag{7.5}$$

Integrating Eq. 7.5 over an arbitrary volume Ω, one obtains:

$$\int_{\Omega} d\mathbf{r}\ \nabla \cdot \mathbf{E}(\mathbf{r}) = 4\pi \int_{\Omega} d\mathbf{r}\ \rho(\mathbf{r}) = 4\pi\, q(\Omega)\,. \tag{7.6}$$

Using the divergence theorem, Eq. 7.6 can be transformed into a surface integral over the surface bounding the volume Ω:

$$\oint dS(\Omega)\ \mathbf{n}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}) = 4\pi\, q(\Omega)\,. \tag{7.7}$$

Using Eq. 7.7, and integrating over the volume of an atom A, the charge on the atom will be given by the flux in $\mathbf{E}(\mathbf{r})$ through its atomic surface $S(A)$. The zero-flux surface $S(A)$ is in turn given by a union of interatomic surfaces $S(A|B)$, there being one such surface for each group B linked to atom A by a bond path. Thus $q(A)$ can be expressed by: (*Bader, 2002*)

$$q(A) = -\sum_{B \neq A} q(B) = \frac{1}{4\pi} \sum_{B \neq A} \oint dS(A \mid B; \mathbf{r})\ \mathbf{n}(\mathbf{r}) \cdot \mathbf{E}(\mathbf{r}) = -\sum_{B \neq A} Q(A \mid B) = \sum_{B \neq A} Q(B \mid A)\,. \tag{7.8}$$

The above result can be utilized to rewrite Eq. 7.4 yielding:

$$\boldsymbol{\mu} = \sum_{A} \boldsymbol{\mu}(A) = \sum_{A} \left\{ \mathbf{M}(A) + \sum_{B \neq A} \left[\mathbf{X}^{c}(A \mid B) - \mathbf{X}_{A} \right] \times \left[-\frac{1}{4\pi} \oint dS(A \mid B; \mathbf{r}_{s})\ \mathbf{E}(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r})_{B} \right] \right\}. \tag{7.9}$$

The contribution of each atom to the dipole moment $\boldsymbol{\mu}$ is given by $\mathbf{M}(A)$, the polarization of its density relative its nucleus, together with the dipoles resulting from the flux in the electric field through each of its interatomic surfaces, measured relative to its nuclear position, that is, as a result of the corresponding charge transfers. One must use Eq. 7.9 for an extended lattice, since in this case the charge transferred across each interatomic surface is explicitly defined only through its use. It is useful to express Eq. 7.9 in the following more compact form:

$$\boldsymbol{\mu} = \sum_{A} \boldsymbol{\mu}(A) = \sum_{A} \left\{ \mathbf{M}(A) + \sum_{B \neq A} \left[\mathbf{X}^{c}(A \mid B) - \mathbf{X}_{A} \right] Q(A \mid B) \right\}, \tag{7.10}$$

wherein the surface integral of Eq. 7.9, including the factor of -(1/4π), determining the charge transferred across $S(A|B)$, is denoted by $Q(A|B)$, with the surface normal directed from B to A. Thus

$Q(A|B) = -\,Q(B|A) = q(B)$ and one has $\sum_{B \neq A} Q(B \mid A) = q(A)$. One may combine the contributions from two atoms that share a common surface $A|B$, that is, two atoms that are linked by a bond path and are thus bonded to one another. The addition of the terms $Q(A|B)$ and $Q(B|A)$ associated with the interatomic surface $A|B$ reduces to $(\mathbf{X}_A - \mathbf{X}_B)\, Q(B|A)$. In this manner one may obtain group or cell contributions. A program FRAGDIP (*Matta, 2001b*) is available to compute the charge transfer and group contributions to the dipole; a complete listing of FRAGDIP is given in appendix 7.1. Popelier has recently devised an algorithm for the evaluation of the surface term $Q(B|A)$ in the isolated molecule case. (*Popelier, 2001*)

The expression for the contribution of an individual atom or unit cell to the polarization given in Eq. 7.9 is formally analogous to that obtained by Martin for the macroscopic polarization **P** expressed in terms of cell contributions. (*Martin, 1974*) The polarization density $\mathbf{P}(\mathbf{r})$ is related to the density by $\nabla\cdot\mathbf{P}(\mathbf{r}) = -\,\rho(\mathbf{r})$. The use of this expression in an integration by parts of the cell integral of $\mathbf{P}(\mathbf{r})$ yields the term $\int_{\mathbf{r}} \rho(\mathbf{r})\, d\mathbf{r}$, corresponding to the cell polarization $\mathbf{M}(A)$, together with the surface term $\oint_{\mathbf{r}} \mathbf{r}\left[\mathbf{n}\cdot\mathbf{P}(\mathbf{r})\right] dS$ where $\mathbf{n}\cdot\mathbf{P}(\mathbf{r})$ is termed the surface charge density, the charge transferred across the cell boundary. (*Feynman et al., 1964*) The surface integral of the electric field appearing in the open system expression, Eq. 7.9, yields the charge transferred across the associated interatomic surface. Thus this term has the same physical meaning as the surface term obtained from general considerations by Martin and both expressions relate the atomic contribution to **P** to the sum of a basin polarization and a term resulting from the charge transferred across the cell boundary. In the discussion which follows we shall refer to $Q(A|B)$ as the surface transfer charge, equal to the charge transferred from A to B across the surface $A|B$.

Eq. 7.3 has been applied in the determination of the dipole moments of the individual molecular constituents in crystals of urea, water and *p*-nitroaniline. (*Gatti et al., 1994; Gatti et al., 1995; Volkov et al., 2000*) Hartree-Fock level wave functions for these periodic systems were calculated using CRYSTAL (*Dovesi et al., 1992*) and the resulting densities analysed using the

interfaced program TOPOND. (*Gatti, 1998*) The use of Eq. 7.3 equates the dipole for each individual molecule to the sum of the atomic polarizations and the sum of the interior charge transfer terms, omitting the contributions from the surface transfer charges between the molecules. Such terms do not in general vanish even for molecules or cells which bear no net charge, the sole requirement being that the flux out equal the flux in for each cell, a property illustrated for the systems studied here. The calculations do demonstrate that the electron density of a periodic dielectric can be analysed to obtain the topological atoms and their properties. (*Gatti et al., 1994*)

It is important to note that Eqs. 7.4, 7.9, and 7.10 do not require the translational symmetry of a periodic crystal for their application. Proper open systems are always uniquely defined in terms of the zero-flux boundary condition Eq. 2.6: for individual atoms in molecules, for the atomic cells in ionic or covalent crystals and for the individual molecules and their constituent atoms in a molecular crystal. In their book on solid state physics, Ashcroft and Mermin (*Ashcroft and Mermin, 1976*) discuss the apparent difficulty encountered in applying the cell concept to 'covalent crystals', crystals in which "appreciable electronic charge density resides between ions (forming the so-called covalent bonds)". They state that in this situation one must invoke band theory and forego the cell model or develop a phenomenology of "bond polarizabilities" and offer as examples crystals with the diamond, zinc-blende and wurtzite structures. This difficulty is not encountered when the atoms in a crystal are treated as open systems. The zincblende structures for example, were among the first solids to be treated by the method of atoms defined as open systems, serving to illustrate the topological definition of a Wigner-Seitz cell. (*Zou and Bader, 1994*)

The partitioning of the dipole moment defined in Eqs. 7.4, 7.9, and 7.10 not only gives physically anticipated transferable group contributions to a dipole moment, but when applied to the first-order field induced dipole, the atomic contributions define and recover the additive group contributions to the molecular polarizability that are experimentally measurable. (*Bader et al., 1992*) The same procedure enables one to partition the transition probability for dipole-induced transitions into atomic or group contributions. (*Bader et al., 2000*) An expression of the same form as Eqs. 7.4

and 7.9 defines the group contributions to the magnetic susceptibility in terms of the first-order induced current density $\mathbf{J}^{(1)}(\mathbf{r})$. (*Bader and Keith, 1993*) The basin polarization of Eq. 7.4 is replaced with the atomic average of the basin magnetization density $\mathbf{r}_A \times \mathbf{J}^{(1)}(\mathbf{r})$ and the atomic charge by $\mathbf{J}^{(1)}(B)$, the atomic average of the induced current. Just as the charges of the neighbouring groups can be replaced by a surface integral of the flux in the associated electric field in Eq. 7.9, the atomic averages $\mathbf{J}^{(1)}(B)$ can be replaced by a surface integral of the flux in the position weighted current, this latter step being a consequence of $\sum_{B \neq A} \mathbf{J}^{(1)}(B) = -\mathbf{J}^{(1)}(A)$. The resulting atomic contributions recover the experimental additive group contributions assigned by Pascal. (*Pascal et al., 1961*) In addition, the quantum theory of atoms in molecules leads to a method of calculating $\mathbf{J}^{(1)}(\mathbf{r})$ in a divergence-free manner, using a gauge origin that changes continuously with the position coordinate $\mathbf{r}$, reflecting its association with a given nucleus. (*Keith and Bader, 1993*) In this manner one is able for the first time, to calculate and display the field induced currents, including the 'ring' currents induced in aromatic molecules by a field applied perpendicular to the molecular plane. Such currents lead to an enhancement in the measured susceptibilities and were termed 'aromatic exaltations' by Pascal. The exaltation assigned to the benzene ring in Pascal's experimental additivity scheme is -15×10^{-6} emu and this value is precisely six times the amount by which the calculated atomic susceptibility for a carbon atom in benzene exceeds that for a correspondingly conjugated atom in non-aromatic butadiene. These examples are given to demonstrate that the atomic partitioning recovers the experimentally measurable group contributions to properties that involve averages over the position coordinate and thus appear to have an inherent origin dependence. The very existence of the experimental additivity schemes demonstrates that such a partitioning is both possible and necessary.

7.4 Model: Atomic Polarization in Molecular Chains

The atomic contributions to the polarization and changes in polarization of molecular chains of increasing length are determined to demonstrate the insensitivity of the contributions from interior

atoms to the size of the system. This result is a consequence of the short-range nature of the one-electron density matrix whose diagonal elements determine the electron density and which, through the stress tensor $\bar{\sigma}(\mathbf{r})$, determines the electronic kinetic and potential energy densities, as well as the current density. It is the short-ranged nature of the one-electron density matrix that underlies the working hypothesis of chemistry and the physics of any system, molecule or crystal: that atoms and functional groupings of atoms exhibit characteristic properties. (*Bader, 1990; Bader, 1995*) The extent of the transferability of the charge distribution of a given grouping of atoms and its prevalence throughout matter is not generally appreciated. Indeed, density functional theory would lead one to anticipate otherwise, since the basic tenet of this theory is that the electron density is uniquely and functionally related to the external potential. (*Hohenberg and Kohn, 1964*) Thus different systems should necessarily exhibit different charge distributions, even for example, interior atoms at different sites within a crystal or two different macroscopic samples of the same material. While density functional theory would appear to exclude perfect transferability of the charge density of an atom in two different systems or at different sites within a single system, it does not place any limit on how small any difference must be. The model systems investigated here demonstrate anew the observation gleaned from QT-AIM , that a given grouping of atoms can exhibit what is an essentially transferable charge distribution that contributes identical additive amounts to each of a system's properties, independent of its size.

7.4.1. Computational Methods

Consider a polar molecule consisting of a chain of n alternating atoms or groups A and B, as exemplified by the polyacetals in figure 7.1, with A = CH_2 and B = O. Each end is capped by a hydrogen atom to yield $H|CH_2O|_nH$, the vertical bars representing the interatomic surfaces defining each repeating $|CH_2O|$ group. The chain is prevented from folding by requiring that the C-O-C and O-C-O angles be equal, this angle and the remaining geometrical parameters assuming values determined by an energy optimization. A normal value for the C-O-C angle in an acetal is 112°, and the optimized

angle for the chains is 109.4°, a value close to the tetrahedral angle found for a saturated carbon atom. To contrast the behaviour of a polar system with an ionic one, we also consider the properties of an optimized chain of nine sodium fluoride units constrained to be linear, capped with sodium and fluoride atoms, the system $Na|FNa|_9F$, figure 7.2.

The density and properties are obtained from a single determinant self consistent field (SCF) wave function using the 6-31G** basis set in Gaussian94. (*Frisch et al., 1995*) This basis set yields geometries in good agreement with experiment and provides an adequate account of the atomic polarizations. For example, the experimental and calculated dipole moments of methanol at its optimized geometry (the model system for n = 1) are 0.67 and 0.72 au respectively, while the corresponding values for NaF are 3.21 and 3.08 au. The sole purpose of the models is to demonstrate the manner in which the atomic contributions determine the polarization of the system and to determine the dependence of the contributions on the chain length. The atomic properties are obtained using the AIMPAC series of programs. (*Biegler-König et al., 1982*) The molecular sums of the atomic integrations of the density and of the kinetic energy density yield respectively, the total number of electrons N and, via the atomic virial theorem, the total energy E. These sums are given in table 7.1, together with the integration errors. While the absolute error increases with the size of the molecule - there are 34 separate atomic integrations for n = 8 - the relative error never exceeds 0.02% in N and 0.5 % in E. The interatomic surfaces in the ionic system $(NaF)_{10}$ are of a simple form, figure 7.2, and the atomic properties are determined with high accuracy. The accuracy of an atomic integration can be gauged from the basin integration of $L(\mathbf{r})$, (Eq. 2.66). In the ionic system the values of L(Na) and L(F) never exceed 5×10^{-4} au while in the polyacetals, the values of L(C) and L(O) for interior atoms approach 4×10^{-3} au.

7.4.2 Atomic Properties as a Function of Chain Length

The interatomic surfaces shared between the terminal hydrogen atoms and the terminal $|CH_2O|$ groups define the boundaries of the n–membered chain, $|CH_2O|_n$. The dependencies of the charge, energy and

volume of a $|CH_2O|$ group on its proximity to the boundary surfaces are considered first to demonstrate the rapid decay in surface effects (these quantities have been defined in chapter 2).

A $|CH_2O|$ group may be considered as the unit cell in a one-dimensional system. Since the system does not possesses a net charge, the repeating unit cell in the interior of the system should exhibit a zero charge once the effects of the surface cells have decayed to zero. Because the zero-flux surfaces define the natural atomic boundaries, the condition of zero cell charge is rapidly attained. Even in the system with $n = 4$, the two interior cells have net charges, denoted by q(cell) of 0.01 e, (see table 7.2 which lists the properties for the two innermost cells in each chain). For $n \geqslant 6$, a single cell at each terminus suffices to yield zero net charges to within the integration error for the remaining cells. The condition of cell neutrality implies that the electron density distributions of interior cells are superimposable, a superposition that includes their interior as well as their bounding interatomic surfaces. The transferable nature of the |C| and |O| atoms and the associated C|O surfaces are qualitatively apparent in figure 7.1. As a consequence, charge neutrality results in all of the properties of the cell exhibiting the condition of essential transferability within computational accuracy, as demonstrated by the data in table 7.2. Two neighbouring interior cells in a given chain have the same values for their volumes, v(cell), for the magnitudes of their atomic polarizations $|\mathbf{M}(\text{cell})|$, dipoles $|\boldsymbol{\mu}(\text{cell})|$ and energies E(cell). The small variation in the values of the cell properties between chains illustrates their near vanishing dependence on chain length, the values averaged over all cases in table 7.2 exhibiting standard deviations of 0.3 % for the cell polarization, 0.5% for the cell dipole and a vanishingly small deviation for the energy. Thus the transferability visually apparent in figure 7.1 is made quantitative.

The $|CH_2|$ and |O| groups of a transferable cell possess charges of magnitude 1.297 ± 0.002 e. The dipole moment is transverse to the chain length, as are the group contributions resulting from the charge transfer, Eq. 7.4. Both groups are polarized in a direction opposed to the direction of charge transfer and the polarization vectors are also transverse to the chain length. Hence there is no

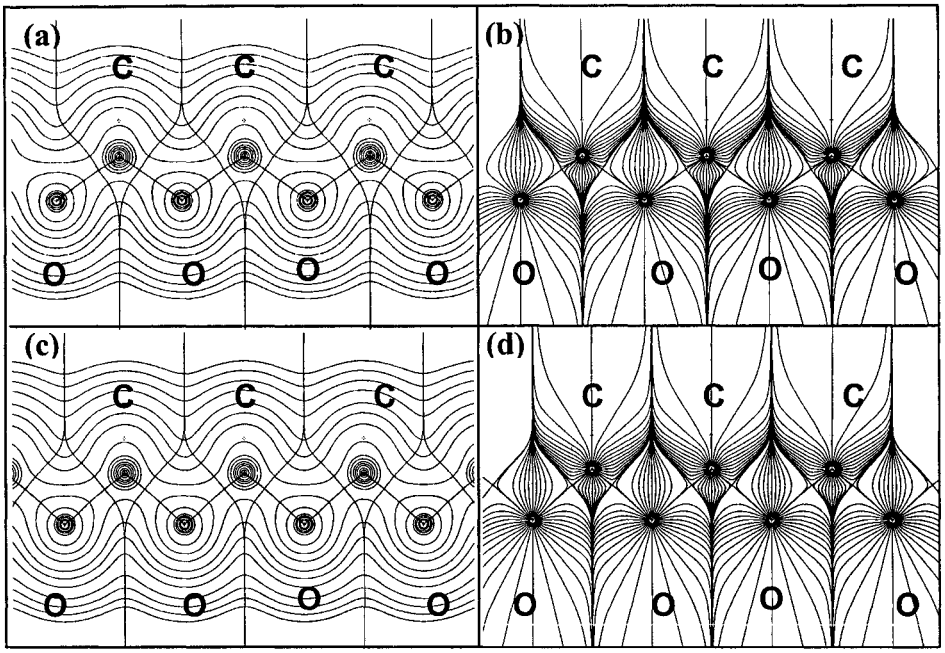


Figure 7.1 **(a)** Contour map of the electron density for the central portion of $H|CH_2O|_8H$ overlaid with interatomic surfaces and bond paths and **(b)**, the associated gradient vector field. The projected positions of the out-of-plane hydrogen atoms are indicated by the open cross appearing in the basin of each C atom. The maps of neighbouring $|CH_2|$ and |O| groups are superimposable. The values of the density and its Laplacian at the bond critical points are typical of shared interactions with respective ρ_b and $\nabla^2\rho_b$ values of 0.285 and -0.289 au for C|O and 0.300 and -1.172 au for C|H, values to be contrasted with those for the ionic system $(NaF)_{10}$. Each atom is bounded by a pair of mirror image surfaces, each $|CH_2O|$ cell by a conjugate pair of surfaces. The density of the compressed molecule and its associated gradient vector field are displayed in **(c)** and **(d)** respectively. Note that the C-O bond paths exhibit a slight curvature, a reflection of the strain introduced by the compression of the chain.

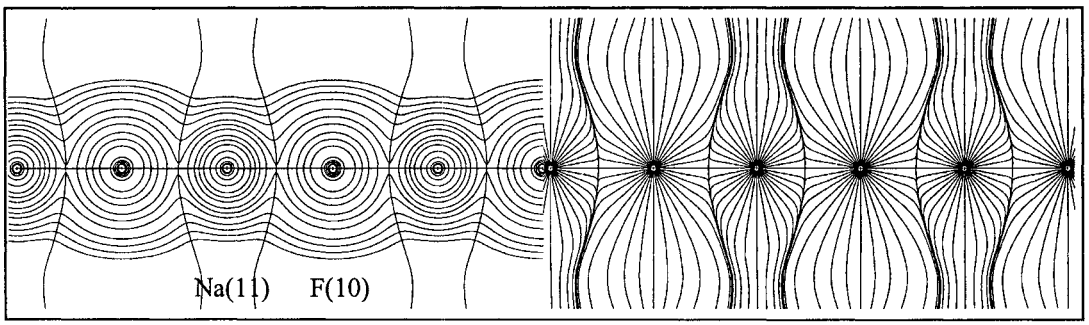


Figure 7.2 **(a)** Display of the electron density for the central portion of the linear $(NaF)_{10}$ system overlaid with the interatomic surfaces and bond paths and **(b)** the associated gradient vector fireld. The values of the density ρ_b and its Laplacian $\nabla^2\rho_b$ at a bond critical point are 0.041 and +0.35 au, respectively, values typical of an "ionic" interaction, one dominated by the consequences of the Pauli principle for the approach of two closed-shell systems.

Table 7. 1. Comparison of self-consistent-field (SCF) values with atomic sums (values in au)

System†	SCF E_{total}	$\sum_{\Omega} E(\Omega)$	|error| in E_{total}	$\sum_{\Omega} N(\Omega)$	N	|%error| in N_{total}	SCF $\|\mu\|$	$\sum_{\Omega} \|\mu(\Omega)\|$	|%error| in $\|\mu\|$
1r	-115.04671	-115.04729	0.00058	17.998	18	0.01%	0.7214	0.7195	0.27%
2r	-228.92885	-228.92991	0.00107	33.996	34	0.01%	1.2785	1.2794	0.07%
2c	-228.90666	-228.90718	0.00052	33.998	34	0.01%	1.4729	1.4794	0.44%
3r	-342.80902	-342.81075	0.00173	49.992	50	0.02%	1.8309	1.8378	0.37%
3c	-342.77280	-342.77434	0.00154	49.993	50	0.01%	2.1245	2.1400	0.73%
4r	-456.68896	-456.68584	0.00312	66.010	66	0.01%	2.3819	2.3711	0.45%
4c	-456.63865	-456.63900	0.00035	65.998	66	0.00%	2.7734	2.7812	0.28%
6r	-684.44856	-684.45272	0.00416	97.981	98	0.02%	3.4819	3.5331	1.47%
6c	-684.37009	-684.37474	0.00466	97.977	98	0.02%	4.0674	4.1521	2.08%
8r	-912.20803	-912.21380	0.00577	129.974	130	0.02%	4.5803	4.6819	2.22%
8c	-912.10140	-912.10853	0.00713	129.965	130	0.03%	5.3588	5.5338	3.27%
NaF	-261.30170	-261.30173	0.00003	19.999	20	0.00%	3.0810	3.0831	0.07%
$(NaF)_{10}$	-2613.70350	-2613.70299	0.00051	200.000	200	0.00%	36.9385	36.8892	0.13%

† For polyacetals, the systems are labeled nr or nc, where n is the number of cells denoted by the vertical bars in $H|CH_2O|_nH$, r stands for relaxed geometry and c for compressed geometry.

Table 7.2. Transferability of cell properties (values in au)

System†	cell	q(cell)	|M(cell)|	|μ(cell)|	E(cell)	v(cell)‡
n=4	2	-0.00980	1.19868	0.56488	-113.881152	228.32
	3	-0.00923	1.20837	0.55883	-113.881330	227.89
n=6	3	0.00120	1.21491	0.56181	-113.881344	227.31
	4	0.00156	1.21585	0.56155	-113.881380	227.36
n=8	4	0.00259	1.21698	0.56370	-113.881561	227.23
	5	0.00267	1.21734	0.56377	-113.881522	227.30
n=21	11	-0.00344	1.21570		-113.879348	228.18
	12	-0.00257	1.21448		-113.879662	227.87
Average		-0.00213	1.21279	0.56242	-113.880912	227.68
Standard deviation		0.00476	0.00594	0.00198	0.000825	0.41

† n is the number of cells denoted by the vertical bars in $H|CH_2O|_nH$, the polyacetals in this tables all have relaxed geometry. The cell numbering starts at the methyl end of the system.
‡ Volume integrated to the 0.001 au isodensity envelope.

component of the atomic polarization parallel to the chain and each group is bounded by a pair of mirror image interatomic surfaces, shared with a pair of |O| groups for the former and with a pair of $|CH_2|$ groups for the latter (see figure 7.1). Thus the values of ρ_b, the electron density at a bond critical point whose associated trajectories define the interatomic surface, are the same for both surfaces separating the carbon of a $|CH_2|$ group from its two neighbouring oxygen atoms, as are the values of

its Laplacian $\nabla^2\rho_b$ and kinetic energy density. Thus the charge of the group is a result of an equal transfer of charge across both of its bounding surfaces and the surface transfer charge Q(C|O) for both surfaces of the $|CH_2|$ group have the value - (1.297/2)e = - 0.649 e and Q(O|C) for each surface of |O| equals + 0.649 e. Since the cell is electrically neutral, the surface transfer charge for a $|CH_2O|$ cell equals - 0.649 e neighbouring the positively charged $|CH_2|$ group and + 0.649 e neighbouring the negatively charged |O|. Each unit cell is bounded by a pair of identical *conjugate surfaces*. There is a small difference of 0.010 e in the net charges of the two end groups, the $HCH_2O|$ group possessing a charge of - 0.642 ± 0.003 e and the $|CH_2OH$ group a charge of +0.652 ± 0.005 e, the average being taken over all systems with $n \geqslant 2$. This difference is accommodated by the first of each of the interior groups and their interior surface charge transfers equal the values quoted above for Q(C|O) and Q(O|C).

Feynman (*Feynman et al., 1964*) refers to the field induced charge density on the two surfaces bounding a neutral dielectric as the *surface polarization charge*. One may view each neutral unit cell in a polar or ionic molecule or crystal as an atomic capacitor, the field of its neighbours inducing a surface polarization charge corresponding to the equal and opposite transfer of charge across each of its conjugate surfaces. The existence of conjugate surfaces plays an important role in the structures of biopolymers. For example, for a polypeptide to consist of a chain of transferable amino acid residues, each residue identified by the side chain R in |NHCHRC(=O)| must be bounded by a pair of conjugate, that is identical, amidic surfaces requiring that a formally neutral residue bear a zero charge. Equal and opposite surface charge transfers of ± 0.5 e are indeed found for the conjugate pair of amidic surfaces for an amino acid residue in the tri-peptide $H_2NCH_2C(=O)$ | $NHCHRC(=O)$ | $NHCH_2C(=O)OH$ where it is linked to two glycyl groups. (*Bader and Martín, 1998*) The residue is neutral with compensating surface transfer charges of ∓ 0.5 e. Because of this conjugate property of the amidic surfaces, one may construct a polypeptide by linking the amidic zero-flux surface of predetermined amino acid residues. (*Chang and Bader, 1992*)

The condition of cell neutrality in the Na|FNa|$_9$F chain is attained more rapidly from the Na terminus, with q(Na) = + 0.9550 e, than from the F terminus with q(F) = - 0.9418 e. Groups three, four and five removed from Na all possess a net charge of ± 0.0002 e, which is within the integration error of 0.0004 e for the total molecule (table 7.1). The atomic charges in the zero charge groups are ± 0.928 e. From group six onwards, the cell charge increases from -0.0005 e to -0.0167e for group nine. Each atom has an atomic polarization directed along the molecular axis in a direction opposed to the direction of charge transfer, with the terminal values being **M**(Na) = - 0.0095 au and **M**(F) = - 0.1457 au, fluorine being more polarizable than Na, as anticipated. The cell polarization decreases in magnitude for the interior groups, to - 0.0035 ± 0.0005 au for groups three to five, but for this chain length, does not vanish as it would for a chain of infinite length.

In this linear system of atoms with charges of alternating sign, ± $q(A)$, the interatomic surfaces and their associated surface transfer charges must exhibit a corresponding alternation, since the surface of $|A|$ bordering an even number of atoms will experience a near vanishing field while the other will experience one equivalent to approximately ± $q(A)$. Thus |F(2)| , the fluorine bordering the terminal sodium Na(1)|, has a surface transfer charge Q(F(2)|Na(1)) = q(Na(1)) = + 0.955 e. Since q(Na1) > |q(F2)| by 0.031 e, the other surface transfer charge of |F2|, the term Q(F2|Na2), will equal - 0.031 e. The surface transfer charges for a central neutral cell |F|Na| are ± 0.963 e with the inner field Q(F|Na) = - Q(Na|F) = - 0.035 e. One may choose to take the nearly neutral NaF| and |FNa as the terminal groups, rather than the individual Na and F atoms. Then each repeating |NaF| cell is bounded by surface transfer charges of ± 0.035 e with inner surface transfer charges of ± 0.963 e. However, the properties of a |FNa| group are the same as those for a |NaF| group, including the charge transfer contributions to the group dipole. This is a consequence of including the contributions from surface transfer charges for the interatomic surfaces that bound the cell. The results are independent of how the individual cells are defined, as long as they exhaust the system's space.

The alternating behaviour exhibited by the surface transfer charges is reflected in a corresponding alternation in the interatomic surfaces and Na-F bond lengths, similar to a Peierls-like

distortion. The links to each end of a unit cell are shorter than the bond linking F to Na within the cell. The maximum difference is for the Na1-F2 link, which is shorter by 0.17 au than the succeeding one. The difference decreases along the chain and attains a minimum value for the zero-charge groups where the central bond length of 3.794 au exceeds those linking the cell to its neighbours by 0.013 au. Unlike the case of the transferable groups $|CH_2|$ and |O| in the polyacetal, the two surfaces bounding a |F| or |Na| atom are not mirror images of one another, the 0.04 au contour being more pinched towards the bond critical point in one than in the other (figure 7.2).

7.4.3 Molecular Dipole and Group Contributions

Table 7.1 lists the values of the molecular dipole moments of the molecules determined in the SCF calculations, compared to the values obtained as the sum of the group contributions using Eq. 7.9. The small differences in the two sets of values are a result of accumulated errors in the numerical integrations over each atomic basin. The differences increase with the size of the system. The use of Eq. 7.9 requires a knowledge of all the atomic charges in the system to determine each atomic contribution to the dipole. This is prohibitive for the case $n = 21$ and consequently only the atoms in two central and the two terminal groups were integrated and the group dipoles are not reported for this case .

The magnitudes of the molecular dipole moments of the polyacetals increase incrementally by 0.551 ± 0.008 au (1 au = 2.5418 debyes) for each $|CH_2O|$ cell beginning with $n = 2$. The dipole moment is directed perpendicular to the chain length, lying parallel to the bisector of a C-O-C or O-C-O with the sense C^+O^-. Beginning with $n = 4$, each of the interior $|CH_2O|$ cells makes a contribution of 0.562 au to the magnitude of the molecular dipole moment (table 7.2) and its moment is directed along the bisector of the C-O-C angle to within a degree. This value is somewhat greater than the average increment because the directions of the moments of the terminal groups differ somewhat from those of the interior groups. The $|CH_2O|$ cell dipole is dominated by the charge transfer contribution, its magnitude being 1.779 au, a contribution reduced by the oppositely directed atomic polarization term

of magnitude 1.216 au. The charge transfer and atomic polarization contributions are oppositely directed for both groups and the corresponding magnitudes may be subtracted to obtain the group dipole, as displayed in figure 7.3. The moment of an interior $|CH_2|$ is dominated by the atomic polarization term of 0.901 au which is reduced by the charge transfer term of 0.403 au to give the group moment of magnitude 0.498 au. The moment of |O| is dominated by the charge transfer term of magnitude 1.378 au. The magnitude of the atomic polarization of |O| is 0.317 au, to yield a dipole of magnitude 1.062 au. The polarization per unit cell equals 2.47×10^{-3} au or 4.24×10^{-2} debyes/Å^3. A crystal composed of such chains of polarized repeating groups would exhibit a permanent dipole moment, it would be pyroelectric.

The system $Na|NaF|_9F$ is linear and the cell dipoles are additive. The average dipole of the nine interior cells is 3.73 au. The three transferable interior cells with zero net charge, cells 3, 4 and 5, make additive contributions of 3.77au. This value equals the charge transfer contribution since the atomic polarizations contribute less than - 0.004 au. This value is considerably smaller in magnitude than the -0.232 au for an isolated NaF molecule, a contribution that significantly reduces the charge transfer contribution of 3.32 au to the dipole of NaF. Defining the end groups to be near neutral NaF| and |FNa cells and the repeating interior cells as |NaF| leaves the description of the system unchanged, as the interior |FNa| and |NaF| cells make the same contributions to all of the molecule's properties. In a face centred cubic lattice representative of the sodium chloride structure, the charge transfer contribution of each |NaF| unit cell would be equally divided into three Cartesian components using Eq. 7.10, but these components would each be cancelled by an oppositely directed moment of a neighbouring cell. The crystal would bear no net surface charge when bounded by [100] lattice planes, the thermodynamically stable surface.

7.4.4 Compression Induced Change in Molecular Polarization

A compression of the polyacetals is used to illustrate the use of cell contributions in the determination of changes in the polarization induced by changes in the cell dimensions that can be brought on by

changes in pressure (piezoelectric effect) or temperature (ferroelectric effect). Each of the polyacetals was compressed by reducing the C-O-C and O-C-O angles, that are constrained to be equal, by 10° and optimizing the remaining geometrical parameters. This compression increases the dipole moment, its magnitude increasing by 0.10 au for each additional cell (table 7.1). The major geometrical changes caused by the compression are a lengthening of the C-O bonds by 0.06 au, a shortening of the C-H bonds by 0.01 au and an opening of the H-C-H angle from 108.0° to 110.7°.

The compression results in a small reversal in the charge transfer, with the less electronegative carbon gaining 0.091 e from oxygen and 0.012 from each hydrogen. The result of this decreased charge transfer and accompanying geometrical changes is a reduction of 0.10 au in the magnitude of the $|CH_2|$ dipole while leaving that of |O| essentially unchanged. Thus the cell dipole is again dominated by the charge transfer contribution from |O|, the magnitude of the cell dipole increasing because of the reduction in the magnitude of the $|CH_2|$ group dipole. The relative magnitudes and directions of the $|CH_2|$ and |O| group dipoles and their charge transfer and atomic polarization contributions are indicated in figure 7.3, which can be contrasted with those for the uncompressed groups. The polarization per unit cell Ω equals 2.99×10^{-3} au or 5.13×10^{-2} debyes/Å^3 yielding $\Delta\mu(\Omega)/\Delta v = 0.52 \times 10^{-3}$ au or 0.89×10^{-2} debyes/Å^3.

7.5 Discussion and Conclusions

The polarization of any system can be expressed as a sum of contributions from its constituent atoms. The contribution consists of two terms: an atomic polarization describing the displacement of the centroid of the electronic charge relative to its nucleus and a surface charge transfer that results from the transfer of charge between the atom and each of its bonded neighbours. For a neutral repeating cell of atoms in a molecule or a crystal, this term can be interpreted as a surface polarization charge, resulting from the transfer of charge between the atoms within the cell. The rapid attainment of cell neutrality and the associated transferability of the cell's charge distribution and properties, even in

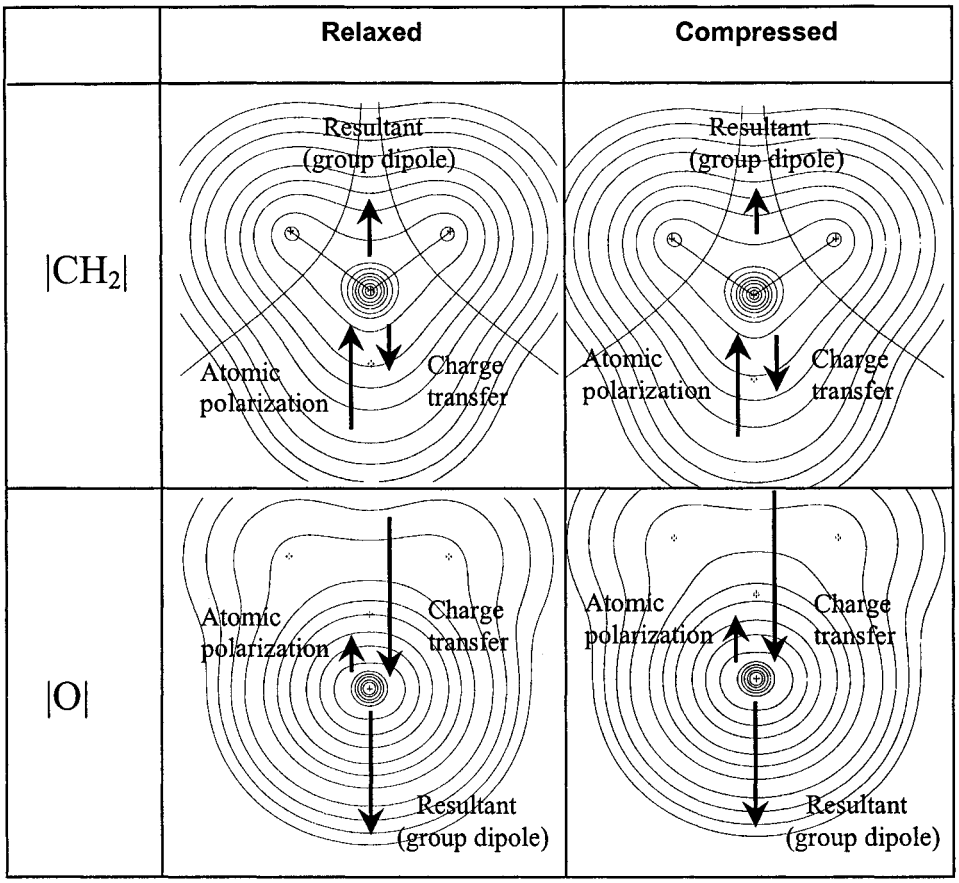


Figure 7.3 Contour maps of the density distributions of a central $|CH_2|$ and |O| groups for the relaxed and compressed geometries of the system with $n = 8$ in the plane perpendicular to the molecular chain. The arrows indicate the direction and magnitude of the atomic polarization and charge transfer contributions to the group dipole. The dipole moment of the $|CH_2|$ group is dominated by the atomic polarizations and that of the |O| group by the charge transfer terms. This picture succinctly summarizes the physics underlying the system's calculated dipole moment and its change upon compression.

systems of finite length, demonstrates the additivity of properties that enables one to relate the polarization or change in polarization of a crystal to the sum of contributions from the basin and surface properties of the individual unit cells that comprise the system.

All measurable properties of a system can be equated to a sum of atomic contributions. Demonstrating how this can be done for the property of polarization is secondary to illustrating how this property, like all others, is not only additive but transferable. The transferable $|CH_2O|$ and |NaF|

cells provide further examples of the fundamental observation underlying chemistry, that the properties of a proper open system, whether determined by the first-order density matrix, the pair density or field induced charge and current densities, are all functionally related to its form in real space, that is, to its distribution of charge. When the charge distribution of the $|CH_2O|$ group is transferable between chains of different lengths or between different sites within a given chain, so are its properties and each group makes identical contributions to every property of the chain.

Not only can one relate the polarization of a system to that of its atomic constituents, one can obtain a mechanistic explanation of the origin and magnitude of the polarization and its relation to other properties. While this aspect has not been dealt with in depth in this chapter, such analysis can relate a change in polarization to the change in the energies of the individual atoms. In the present case, for example, the increase in energy on compression of a polyacetal chain is primarily confined to the oxygen, the stability of the carbon increasing with the transfer of density to its basin. The possibility of spatially partitioning the energy into additive contributions is a consequence of the atomic statement of the virial theorem wherein the virial field (the virial of the Ehrenfest force exerted on the electron density) reduces the many-particle interactions to a dressed density in real space. The virial field provides the most short-range possible description of the interactions experienced by an element of the density, a consequence of the electronic virial containing the nuclear-nuclear energy of repulsion that is described locally by the virial of the Hellmann Feynman force exerted on the density. While the individual contributions to the potential energy are long-range, as for example, the exchange and correlation energy densities of density functional theory, the virial field exhibits only the resultant change, one that reflects the corresponding chemical changes in the form of the atom determined by $\rho(\mathbf{r})$. Figure 7.4 illustrates the transferable nature of the kinetic energy density and the virial field for cells in the polyacetal chain whose forms parallel the transferability of the charge distribution. The kinetic energy density integrates to the negative of the atom's total energy, the virial field to twice its energy.

The arguments presented in this chapter challenge Resta's (*Resta, 2000*) statement that the solution to the problem of calculating a change in the macroscopic polarization of a dielectric can be achieved only by shifting the problem from one stated in terms of the density to one stated in terms of the phase of the wave function. It does so by demonstrating that the polarization or its change can be determined in terms of the properties of the atomic components of the charge distribution when proper account is taken of the surface charge density that arises as a consequence of the transfer of charge between the atoms. Both approaches agree that the solution to the problem is independent of any periodicity in the density. In Resta's approach, the change in polarization involves a change in a parameter appearing in the Hamiltonian. To quote Resta: "the parameter schematizes a kind of coupling with other variables not included in the Hilbert space, or more generally with 'the rest of the universe' to use Berry's words." (*Resta, 2000*) The advantage of using the physics of an open system is that all of the effects of the surroundings are included, made manifest by fluxes in its bounding surface and including for example, the effect of charge transfer. A parametric induced change in the system may be treated directly in terms of the corresponding changes in the properties of the observable electron or current densities, two fields introduced by Schrödinger in his fourth paper, together with the equation of continuity that relates them. (*Schrödinger, 1926b*) He strongly advocated their use in the interpretation and understanding of the electronic and magnetic properties of matter, arguing equally against the use of the wave function for this purpose. While lacking the abstract elegance of the Berry approach that uses the phase of the wavefunction, the present method has the advantage of focussing on the observable physics of the problem.

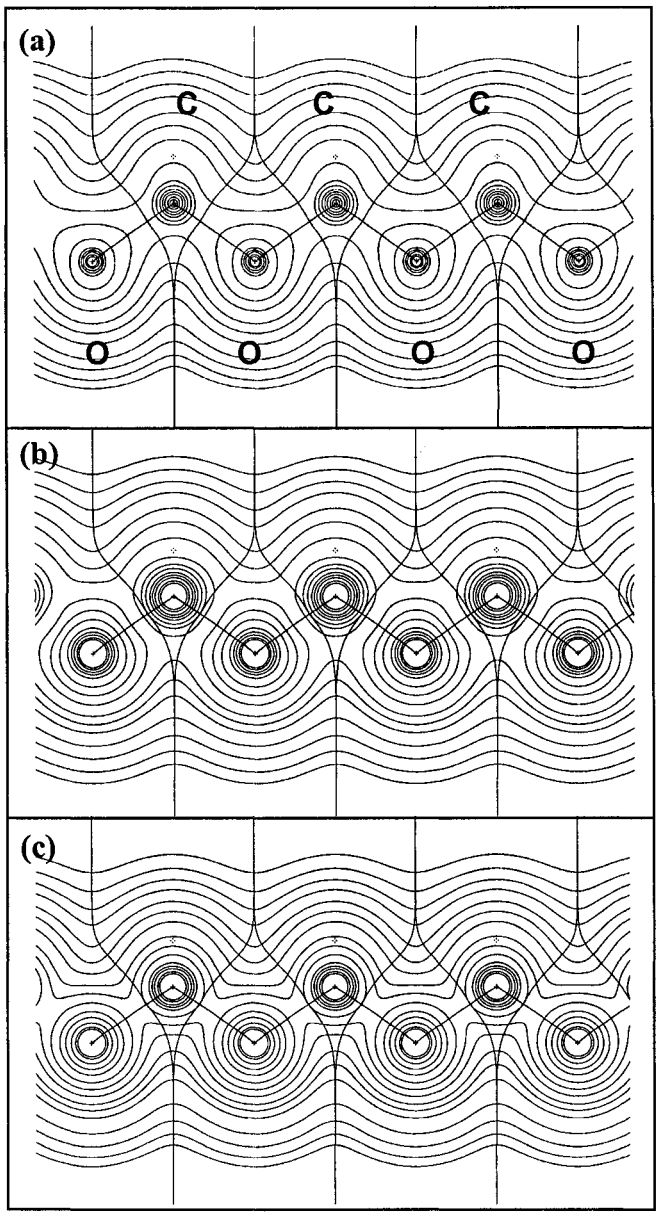


Figure 7.4 Contour maps of: **(a)** the density, **(b)** the kinetic energy density, and **(c)** the virial field density for a portion of $H|CH_2O|_8H$ overlaid with interatomic surfaces and bond paths, as determined by the gradient vector field of the electron density. The virial field integrates to twice the total energy E, the kinetic energy density to the negative of E. The virial field is topologically homeomorphic with the electron density, exhibiting the same type and number of critical points and the same structure diagram. Both the kinetic energy density and virial field mirror the essentially perfect transferability of the electron density for each atom in the system. Atoms that look the same, i.e. that have the same electron density distribution, possess the same properties, whether free or bound.

APPENDIX 7.1: Listing of the FRAGDIP Program (QCPE 0801)

The following program, written by the author, computes the group contributions to the molecular dipole moment and was used to generate the results described in the present chapter. The program consists of *ca.* 390 lines of code written in the Turbo Pascal Language. (*Matta, 2001b*)

```
program dipole_from_fragments;

{$N+}

uses
        wincrt;
const
        MaxNum=50;
type
          IntArray=array [1..MaxNum] of Integer;
          IntArray2D=array [1..Maxnum,1..MaxNum] of integer;
          ExtendedArray=array [1..MaxNum] of Extended;
var
          SumQ,SumE,SumVol,molecX,molecY,molecZ,molecMag:Extended;
          SumQrunning,SumErunning,SumVolrunning:Extended;
          Atom,AtomR,n,ncp,BondNumber,R,NR:IntArray;
          counter, counter1, counter11,counter111: Integer;
          counter2, counter6,MaxCounter: Integer;
          ii,NumAtoms,NumBonds,NumResiduals,i,j,k,newindex: integer;
          x,y,z,dx,dy,dz,q,E,Vol:ExtendedArray;
          n1,x1,y1,z1,dx1,dy1,dz1,q1:ExtendedArray;
          cpx,cpy,cpz,qR:ExtendedArray;
          xContrib,yContrib,zContrib:ExtendedArray;
          xRContrib,yRContrib,zRContrib:ExtendedArray;
          yesORno:char;
          InFileName1, InFileName2, InFileName3, OutFileName, Title: String;
          InFile1, InFile2, InFile3, OutFile: Text;
          bol6: boolean;
          xContribTotal,yContribTotal,zContribTotal,xmu,ymu,zmu:Extended;
          xRContribTotal,yRContribTotal,zRContribTotal:Extended;
          Xtotal,Ytotal,Ztotal,magnitude:ExtendedArray;
        begin
          Writeln;
          Writeln ('================================================');
          Writeln (' FRAGMENTS DIPOLE MOMENTS COMPUTATION (FRAGDIP)');
          Writeln ('================================================');
          Writeln;
{Input section}
          counter6:=0;
          Writeln (' Info file contains the following field:');
          Writeln (' serial numbers, atomic coordinates, dipoles components, and charge:');
          Write (' INFO FILE NAME WITH EXTENSION: --> ');
          Readln (InFileName1);
          Assign(InFile1, InFileName1);
          Writeln;
          Writeln (' Bond critial points file contains the following field:');
          Writeln (' serial numbers, critical points coordinates:');
          Write  (' BCP FILE NAME WITH EXTENSION: --> ');
          Readln (InFileName2);
          Assign(InFile2, InFileName2);
          Writeln;
          Write  (' RESIDUES FILE NAME WITH EXTENSION: --> ');
```

```
            ReadIn (InFileName3);
            Assign(InFile3, InFileName3);
            Reset (InFile3);
            Writeln;
            Write (' Output file name with extension: --> ');
            ReadIn (OutFileName);
            Writeln;
            Writeln ('*********************************************');
{Prepare output file}
            Assign (OutFile,OutFileName);
            ReWrite (OutFile);
            Writeln (OutFile);
            Writeln (OutFile,'================================================');
            Writeln (OutFile,' FRAGMENTS DIPOLE MOMENTS COMPUTATION (FRAGDIP)');
            Writeln (OutFile,'================================================');
            Writeln (OutFile);
            Writeln (OutFile,'Input atomic info file: ',InFileName1);
            Writeln (OutFile,'Input bond critical point (bcp) info file: ',InFileName2);
            Writeln (OutFile,'Residues information from info file: ', InFileName3);
            Writeln (OutFile);
{Main loop: compute the dipole of a FRAGMENT in a molecule}
           counter:=1;
            molecX:=0.0;
            molecY:=0.0;
            molecZ:=0.0;
            molecMag:=0.0;
           SumQrunning:=0.0;
           SumErunning:=0.0;
           SumVolrunning:=0.0;
repeat
            Xtotal[counter]:=0.0;
            Ytotal[counter]:=0.0;
            Ztotal[counter]:=0.0;
{Read atoms composing the fragment}
            Writeln (OutFile,'FRAGMENT ',counter);
            Writeln (OutFile,'============');
            NumAtoms:=0;
            Write (' How many ATOMS are in FRAGMENT',counter,' ? --> ');
            ReadIn (NumAtoms);
            NumResiduals:=0;
            Write (' How many R GROUPS are connected to FRAGMENT',counter,'  ? --> ');
            ReadIn (NumResiduals);
            Writeln ('-----------------------------------------------');
            Writeln (OutFile,'There are ',NumAtoms,' atom(s) in FRAGMENT',counter,'   connected to
                  ',NumResiduals,' residual group(s)');

            i:=0;
              for i:=1 to NumAtoms do
                 begin
                       Writeln (' Enter an atom number:');
                             if i=1 then
                             Writeln (' (IMPORTANT:First atom is the one connected to the rest of the molecule)');
                             Write ('  Atom[',i,']---> ');
                             ReadIn (Atom[i]);
                             Writeln ('-----------------------------------------------');
                 end;
{Display and print the atom numbers of atoms in the FRAGMENT and calculate the fragment total charge, energy
and volume }
            Write ('FRAGMENT', counter,' consist of the following atoms:  ');
            Write (OutFile,'FRAGMENT', counter,' consist of the following atoms:  ');
            i:=0;
            SumQ:=0;
            SumE:=0;
            SumVol:=0;
```

```
                for i:=1 to NumAtoms do
                  begin
                    Write (' ');
                    Write (Atom[i]);
                    Write (' ');
                    Write (OutFile,' ');
                    Write (OutFile,Atom[i]);
                    Write (OutFile,' ');
                      Reset (InFile1);
                      counter111:=0;
                       While counter111<(Atom[i]-1) do
                         begin
                           ReadIn (InFile1);
                           counter111:=counter111+1;
                         end;
                      ReadIn (InFile1,n[i],x[i],y[i],z[i],dx[i],dy[i],dz[i],q[i],E[i],Vol[i]);
                      SumQ:=SumQ+q[i];
                      SumE:=SumE+E[i];
                      SumVol:=SumVol+Vol[i];
                    end;
                      SumQrunning:= SumQrunning+SumQ;
                      SumErunning:= SumErunning+SumE;
                      SumVolrunning:= SumVolrunning+SumVol;
                  WriteIn;
                  WriteIn ('Charge of the Fragment = ',SumQ:12:6);
                  WriteIn ('Energy of the Fragment = ',SumE:12:6);
                  WriteIn ('Volume of the Fragment = ',SumVol:12:4);
                  WriteIn (OutFile);
                  WriteIn (OutFile,'Volume of the Fragment = ',SumVol:12:4);
                  WriteIn (OutFile,'Energy of the Fragment = ',SumE:12:6);
                  WriteIn (OutFile,'Charge of the Fragment = ',SumQ:12:6);
                  WriteIn;
{Input information about the residuals attached to the fragment}
              j:=0;
                for j:=1 to NumResiduals do
                  begin
                    Write (' Enter the bcp between atom ', Atom[1],' of fragment',counter,' and R[',j,'] : ');
                    Read (ncp[j]);
                    WriteIn;
                    Write (' R[',j,'] consists of how many atoms? --> ');
                    ReadIn (NR[j]);
                    WriteIn;
                    WriteIn ('.....');
{Reset (InFile3);}
{Read atoms in residual and sum their charges to obtain the total charge on the residue}
                  k:=0;
                  qR[j]:=0.0;
                    for k:=1 to NR[j] do
                                begin
                                  WriteIn (' Enter the atoms of R[',j,']');
                                  Write   (' AtomR[',j,k,']---> ');
                                  ReadIn (InFile3, AtomR[k]);
                                  Write (AtomR[k]);
{Charge on residue Rj}
                                  Reset (InFile1);
                                  counter11:=0;
                                            While counter11<(AtomR[k]-1) do
                                              begin
                                                ReadIn (InFile1);
                                                counter11:=counter11+1;
                                              end;
                                  ReadIn (InFile1,n1[k],x1[k],y1[k],z1[k],dx1[k],dy1[k],dz1[k],q1[k]);
                                  qR[j]:=qR[j]+q1[k];
                                  WriteIn;
```

```
                    Writeln ('Total charge on residue R[',j,'] = ',qR[j]:12:6);
                    Writeln;
                    Writeln ('.....');
                  end;
    Writeln (OutFile,'Total charge on residue R[',j,'] = ',qR[j]:12:6);
    Writeln;
 end;
{Read the atomic coordinates (x,y,z), the atomic dipole moment components (dx, dy, dz), and
the atomic charges of i atoms from the inf file}
        i:=0;
              for i:=1 to NumAtoms do
               begin
                Reset (InFile1);
                counter1:=0;
                 While counter1<(Atom[i]-1) do
                  begin
                   Readln (InFile1);
                   counter1:=counter1+1;
                  end;
                Readln (InFile1,n[i],x[i],y[i],z[i],dx[i],dy[i],dz[i],q[i]);
                {Writeln (n[i],x[i],y[i],z[i],dx[i],dy[i],dz[i],q[i]);}
               end;
{Read the coordinates of the bond critical points (cpx,cpy,cpz) from the bcp file}
        j:=0;
              for j:=1 to NumResiduals do
                 begin
                  Reset (InFile2);
                  counter2:=0;
                   While counter2<(ncp[j]-1) do
                     begin
                      Readln (InFile2);
                      counter2:=counter2+1;
                     end;
                  Readln (InFile2,ncp[j],cpx[j],cpy[j],cpz[j]);
                 end;
{Calculate the contributions of FRAGMENT's atoms DIPOLES}
        xmu:=0.0;
        ymu:=0.0;
        zmu:=0.0;
        i:=0;
              for i:=1 to NumAtoms do
                 begin
                  xmu:=xmu+dx[i];
                  ymu:=ymu+dy[i];
                  zmu:=zmu+dz[i];
                  Writeln (xmu:10:5, ymu:10:5, zmu:10:5);
                 end;
        Writeln ('TOTALS mu',xmu:10:5, ymu:10:5, zmu:10:5);
        xContribTotal:=0.0;
        yContribTotal:=0.0;
        zContribTotal:=0.0;
        i:=0;
        if NumAtoms >=2 then
              for i:=2 to NumAtoms do
                 begin
                  xContrib[i]:=(x[i]-x[1])*q[i];
                  yContrib[i]:=(y[i]-y[1])*q[i];
                  zContrib[i]:=(z[i]-z[1])*q[i];
                  xContribTotal:=xContribTotal+xContrib[i];
                  yContribTotal:=yContribTotal+yContrib[i];
                  zContribTotal:=zContribTotal+zContrib[i];
                  Writeln (xContrib[i]:10:5, yContrib[i]:10:5, zContrib[i]:10:5);
                 end;
        Writeln ('TOTALS',xContribTotal:10:5,yContribTotal:10:5,zContribTotal:10:5);
```

```
{Calculating the contribution of FRAGMENT-to-Rs contributions}
            xRContribTotal:=0.0;
            yRContribTotal:=0.0;
            zRContribTotal:=0.0;
            j:=0;
                        for j:=1 to NumResiduals do
                          begin
                            xRContrib[j]:=(cpx[j]-x[1])*qR[j];
                            yRContrib[j]:=(cpy[j]-y[1])*qR[j];
                            zRContrib[j]:=(cpz[j]-z[1])*qR[j];
                            xRContribTotal:=xRContribTotal+xRContrib[j];
                            yRContribTotal:=yRContribTotal+yRContrib[j];
                            zRContribTotal:=zRContribTotal+zRContrib[j];
                            Writeln (xRContrib[j]:10:5, yRContrib[j]:10:5, zRContrib[j]:10:5);
                          end;
            Xtotal[counter]:= xContribTotal+xRContribTotal+xmu;
            Ytotal[counter]:= yContribTotal+yRContribTotal+ymu;
            Ztotal[counter]:= zContribTotal+zRContribTotal+zmu;
            magnitude[counter]:=SQRT(sqr(Xtotal[counter])+sqr(Ytotal[counter])+sqr(Ztotal[counter]));
            Writeln ('Fragment group dipole components:');
            Writeln (Xtotal[counter]:12:6,' ',Ytotal[counter]:12:6,' ',Ztotal[counter]:12:6,' ',magnitude[counter]:12:6);
            Writeln (OutFile,'Fragment group dipole components:');
            Writeln (OutFile,'     x          y          z      dipole magnitude');
            Writeln (OutFile,Xtotal[counter]:12:6,' ',Ytotal[counter]:12:6,' ',Ztotal[counter]:12:6,'
          ',magnitude[counter]:12:6);
            Writeln (OutFile);
            Writeln (OutFile,'Fragment group dipole components (Normalized):');
            Writeln (OutFile,'     x          y          z      dipole magnitude');
            Writeln (OutFile,Xtotal[counter]/magnitude[counter]:12:6,
            ' ',Ytotal[counter]/magnitude[counter]:12:6,' ',
            Ztotal[counter]/magnitude[counter]:12:6,' ',magnitude[counter]/magnitude[counter]:12:6);
            Writeln (OutFile,'-------------------------------------------------------------------------');
            Writeln (OutFile);
{Calculate the molecular dipole}
            molecX:=molecX+Xtotal[counter];
            molecY:=molecY+Ytotal[counter];
            molecZ:=molecZ+Ztotal[counter];
            molecMag:=SQRT(sqr(molecX)+sqr(molecY)+sqr(molecZ));
{Update the counter after the MAIN LOOP}
            counter:=counter+1;
            write ('Do you want more calculations (y/n):');
            readln (yesORno);
            bol6:=yesORno='n';
                        until bol6;
                          MaxCounter:=counter;
                          Writeln (OutFile,'=====================================================');
                          Writeln (OutFile);
                          Writeln (OutFile,'Sums of fragments dipole components contributions* :');
                          Writeln (OutFile,'          x              y              z       dipole magnitude');
                          Writeln (OutFile,molecX:12:6,' ',molecY:12:6,' ',molecZ:12:6,' ',molecMag:12:6);
                          Writeln (OutFile);
                          Writeln (OutFile,'* These sums equal to the molecular values only if all the fragments');
                          Writeln (OutFile,'  are included in the calculation.');
                          Writeln (OutFile,'=====================================================');
                          Writeln (OutFile);
                          Writeln (OutFile,'SUMMARY');
                          Writeln (OutFile,'=======');
                          Writeln (OutFile,'          x              y              z       dipole magnitude');
              for counter:=1 to (MaxCounter-1) do
                begin
                  Write (OutFile,'FRAGMENT',counter,'  ');
                  Writeln (OutFile,Xtotal[counter]:12:6,' ',Ytotal[counter]:12:6,' ',Ztotal[counter]:12:6,'
          ',magnitude[counter]:12:6);
                end;
```

```
        Writeln (OutFile);
        Write (OutFile,'Molecule (a.u.) ');
        Writeln (OutFile,molecX:12:6,' ',molecY:12:6,' ',molecZ:12:6,' ',molecMag:12:6);
        Write (OutFile,'Molecule (D) ');
        Writeln (OutFile,2.5418*molecX:12:6,' ',2.5418*molecY:12:6,' ',2.5418*molecZ:12:6,'
      ',2.5418*molecMag:12:6);
        Writeln ('TOTAL CHARGE = ',SumQrunning:12:6);
        Writeln ('TOTAL ENERGY = ',SumErunning:12:6);
        Writeln ('TOTAL VOLUME = ',SumVolrunning:12:6);
        Writeln (OutFile);
        Writeln (OutFile,'TOTAL CHARGE = ',SumQrunning:12:6);
        Writeln (OutFile,'TOTAL ENERGY = ',SumErunning:12:6);
        Writeln (OutFile,'TOTAL VOLUME = ',SumVolrunning:12:6);
        Writeln (OutFile);
        Writeln (OutFile,'Normalized fragments dipoles');
        Writeln (OutFile,'        x          y          z       dipole magnitude');
              for counter:=1 to (MaxCounter-1) do
                begin
                  Write (OutFile,'FRAGMENT',counter,'  ');
                  Writeln (OutFile,Xtotal[counter]/magnitude[counter]:12:6,
                  ' ',Ytotal[counter]/magnitude[counter]:12:6,' ',
                  Ztotal[counter]/magnitude[counter]:12:6,' ',magnitude[counter]/magnitude[counter]:12:6);
                end;
        Writeln (OutFile);
        Writeln (OutFile, 'Normal termination of FRAGDIP.');
        Close (InFile1);
        Close (InFile2);
        Close (OutFile);
        DoneWinCrt
end.
```

PART IV:

ELECTRON DELOCALIZATION AND PROBLEMS IN CHEMICAL BONDING AND REACTIVITY

Chapter 8

Aromaticity in Polycyclic Hydrocarbons Revisited in Terms of the Electron Density and of the Pair Density

> "... six electrons are able to form a group which resists disruption, and may be called the *aromatic sextet* ... The circle in the ring symbolises the view that six electrons in the benzene molecule produce a stable association which is responsible for the *aromatic* character of the substance."
>
> Armit and Robinson (*Armit and Robinson, 1925*).

8.1 Statement of the Problem

This chapter addresses the question of whether one can formulate a theory of aromaticity in real three-dimensional space. Recasting the problem in terms of observables that reside in real-space would provide a framework in which the contribution of an atom (or a group of atoms) to the aromatic character of a molecule can be understood. In other words, aromaticity can be understood "locally" at the level of an atom (or a group of atoms) in a molecule. A description of aromaticity in terms of observables such as the electron density or the pair density leads to its roots in physics instead of relying on models based on a particular choice of orbitals or on graph theoretic methods.

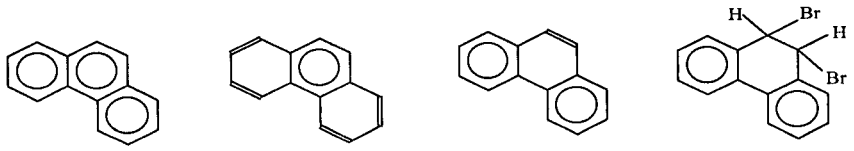
It is shown here how the physical description of aromaticity can be used to explain and predict a host of experimental properties of aromatic molecules. It is shown how the empirical Clar sextet theory (a set of rules derived from the known reactivity of polycyclic hydrocarbons), resonance energies, and NMR spin-spin coupling constants, can all be predicted accurately using properties

derived from the pair density. Further, the unexpected stability of the angular isomer phenanthrene over its linear isomer anthracene can be traced to its atomic origin where two hydrogen atoms involved in a dihydrogen bond located in the bay region are found to be responsible for the anomaly. This chapter is based on (*Matta and Hernández-Trujillo, 2002; Matta et al., 2002*).

8.2 Introduction

Aromatic compounds are unsaturated cyclic molecules characterised by a number of chemical and physical properties. Typically, aromatic molecules react by substitution rather than by addition reactions which are characteristic of unsaturated aliphatic compounds. Aromatic molecules are known to have an unusual stability that explains their lower than expected heats of combustion and of hydrogenation. Geometric (alternation of bond lengths) and spectroscopic properties have been used as indicators of aromatic character. With the exception of benzene, aromatic molecules exhibit different degrees of aromatic character. In particular, polycyclic aromatic hydrocarbons display this differential aromaticity within one and the same molecule. It is also clear that many polycyclic aromatic compounds are more reactive than benzene, and that many of them react with electrophilic reagents by addition rather than by substitution. (*Badger, 1969*) In order to rationalise the reactivity of aromatic compounds, patterns of unsaturation are assigned. The unsaturation pattern of benzene, the aromatic molecule par excellence, has traditionally been depicted by a circle (*Armit and Robinson, 1925*) to indicate the delocalization of six π-electrons (a sextet) over a rigid σ-framework. It is less straightforward, however, to assign unsaturation patterns in condensed polycyclic aromatic hydrocarbons. (*Clar, 1972*) Consider phenanthrene for example:

Scheme 8.1



I II III IV

Structure I is incorrect since it implies 18 π-electrons while phenanthrene has only 14. This leaves only two other possibilities: Structure II, in which one sextet is delocalized over the carbon atoms of the inner ring, the outer rings possessing localized double bonds, or structure III, with two sextets on the outer rings and a localized double bond on the inner ring. The chemistry of phenanthrene suggests structure III since the molecule is known to readily undergo addition reactions to yield products similar to (IV). In this case, assigning a chemical structure was based on empirical chemistry. It is a goal of this chapter to predict a chemical structure consistent with observation using properties derived from the density and the pair density.

Howard and Krygowski (*Howard and Krygowski, 1997*) use the quantum theory of atoms in molecules (QT-AIM) (*Bader, 1990*) to analyze the density of some polycyclic aromatic hydrocarbons. Even though they do not examine electron delocalization, these authors report excellent correlations between topological properties of critical points in the density and empirical aromaticity indices based on bond lengths (*Krygowski, 1993; Krygowski et al., 1995*) and on absolute magnetic shielding at the geometric ring centers (NICS). (*Schleyer P.v-R. et al., 1996; Schleyer P.v-R. et al., 2001*) The study of Howard and Krygowski appears to be the only previous work discussing aromaticity in terms of the total density. Excellent reviews on other aspects of aromaticity have recently appeared, (*Schleyer P.v-R. (Guest Ed.), 2001*) however, a study of the relationship between aromaticity and the density and the pair density is missing. This chapter is a step in filling this gap.

8.3 Delocalization of Electrons

Electron delocalization has long been invoked to rationalize problems of structural stability and chemical reactivity, particularly in conjugated and aromatic molecules. These discussions were understandably couched initially in terms of the orbital model, (*Coulson, 1961; Salem, 1966*) an early example of prime importance being Coulson's use of molecular orbital theory to define a bond order. (*Coulson, 1939; Coulson and Longuet-Higgins, 1947; Coulson and Longuet-Higgins, 1948*) His definition of a mobile bond order and its application referred to delocalized π-electrons and involved a summation of products of coefficients of atom centred basis functions over the π-molecular orbitals. Such a definition that relates delocalization to a property determined by all of the occupied orbitals is to be contrasted with the subsequent introduction of localized orbitals and their use in discussions of electron localization. In this model, the spatial localization/delocalization of individual pairs of electrons is related to the corresponding property of an individual molecular orbital. In fact, all orbitals contribute to the pair density, the density that determines electron localization. The pair density, like the one-electron density, is invariant to any unitary transformation of the orbitals and no individual orbital can be identified as determining any observable feature of a molecule; the use of Koopman's theorem to approximate an ionization potential in terms of a canonical orbital energy, is a notable exception. With the advent of modern computational facilities, the need for the use of models in the description of electron delocalization is obviated and instead, one may use quantum mechanics to obtain a definition of electron delocalization and determine its effect on measurable properties of a system.

Salem (*Salem, 1966*) noted that Coulson's definition of a mobile bond order (*Coulson, 1961*) in the molecular orbital theory of conjugated systems serves as a link between the two notions of Fermi correlation and electron delocalization. In 1975 Bader and Stephens, (*Bader and Stephens, 1975*) showed that the spatial pairing of electrons and their localization is a consequence of the Fermi correlation determined by the electron pair density. They demonstrated that the extent to which

electrons are spatially localized or delocalized is determined by the corresponding spatial extent of the density of the Fermi hole, as measured by the exchange of same-spin electrons. It has since been proposed (*Bader et al., 1996*) that the spatial distribution of the Fermi correlation can be used to provide a common, quantitative basis for the concept of electron delocalization, as it is used throughout chemistry. This conclusion is supported by the present study wherein the Fermi correlation is used to account for the patterns of delocalization that explain the spectroscopic and reactive properties of the polycyclic aromatic hydrocarbons. Most importantly, it was demonstrated (*Bader et al., 1996*) that Fermi correlation can be considered to be the mechanism whereby distant atoms communicate with one another. This is also exemplified further in the present chapter by showing how the delocalization of the spin density between the basins of different hydrogen atoms provides a model for the understanding of measured long-range proton spin-spin coupling constants, constants that do not necessarily exhibit a simple fall-off with increasing internuclear separation.

8.4 Electron Delocalization and Fermi Correlation

The Fermi hole, the physical manifestation of Pauli's exclusion principle, has a simple physical interpretation: (*Bader and Stephens, 1975*) it provides a description of how the density of an electron of given spin, called the reference electron, is spread out from any given point, into the space of another same-spin electron, thereby excluding the presence of an identical amount of same-spin density. (*Bader et al., 1996*) It is a negative quantity, as it decreases the amount of same spin density throughout space by one electronic charge. If the density of the Fermi hole is maximally localized in the vicinity of the reference point, then all other same-spin electrons are excluded from its vicinity and the reference electron is localized. For a closed-shell molecule, the result is a localized pair. Correspondingly, the electron can go wherever its Fermi hole goes and if the Fermi hole of an electron when referenced to a given atom, is delocalized into the basin of a second atom, then the electron is shared between them.

While the ideas developed here apply to any level of theory, the discussion is given in terms of the Hartree-Fock (HF) model. Not only is Fermi correlation the sole source of electron correlation in the HF model of electronic structure, the HF description is acknowledged to provide an excellent approximation to this quantity. The Fermi hole is related to the pair density, which we will define first.

The pair density ($\rho(\mathbf{r}_1,\mathbf{r}_2)$) is a function of six spatial coordinates of a pair of electrons and is proportional to the conditional probability that an electron is in volume element $d\mathbf{r}_1$ centered at $\mathbf{r}_1$ when another electron is in volume element $d\mathbf{r}_2$ centered at $\mathbf{r}_2$:

$$\rho(\mathbf{r}_1,\mathbf{r}_2) = \frac{N(N-1)}{2}\sum_{spins}\int d\mathbf{r}_3\int d\mathbf{r}_4 \ldots \int d\mathbf{r}_N\ \Psi^*(\mathbf{x}_1,\mathbf{x}_2\ldots\mathbf{x}_N)\,\Psi(\mathbf{x}_1,\mathbf{x}_2\ldots\mathbf{x}_N)\,. \tag{8.1}$$

The factor $N(N\text{-}1)/2$ is the number of *distinct* electron pairs in the molecule. It is not generally appreciated that the pair density is an experimentally accessible quantity which can be obtained from inelastic x-ray and electron scattering. (*Bartel and Gavin, 1964; Stewart, 1977; Watanabe et al., 2001*)

The density of α-β pairs is uncorrelated in HF and in this case Eq. 8.1 reduces to the simple product of the corresponding spin densities:

$$\rho^{\alpha\beta}(\mathbf{r}_1,\mathbf{r}_2) = \rho^{\alpha}(\mathbf{r}_1)\rho^{\beta}(\mathbf{r}_2). \tag{8.2}$$

The pair density for same-spin electrons is however, mediated by the density of the Fermi hole $h^{\alpha}(\mathbf{r}_1,\mathbf{r}_2)$ and given by:

$$\rho^{\alpha\alpha}(\mathbf{r}_1,\mathbf{r}_2) = \rho^{\alpha}(\mathbf{r}_1)\left[\rho^{\alpha}(\mathbf{r}_2)+h^{\alpha}(\mathbf{r}_1,\mathbf{r}_2)\right], \tag{8.3}$$

where the density of the Fermi hole for electrons of α-spin is given by:

$$h^{\alpha}(\mathbf{r}_1,\mathbf{r}_2) = -\sum_i\sum_j \frac{\phi_i^*(\mathbf{r}_1)\,\phi_j(\mathbf{r}_1)\,\phi_j^*(\mathbf{r}_2)\,\phi_i(\mathbf{r}_2)}{\rho^{\alpha}(\mathbf{r}_1)} \tag{8.4}$$

and the sums run over the α-spin orbitals. This density exhibits the required properties of reducing to $-\rho^{\alpha}(\mathbf{r}_1)$ when $\mathbf{r}_2 = \mathbf{r}_1$, corresponding to the complete removal of same-spin density from the position of

the reference electron and yielding –1 when integrated over the space of the second electron, demonstrating that the Fermi hole associated with a given electron removes the equivalent of one electronic charge. The quantity in the square brackets in Eq. 8.3 is the conditional same-spin density, the probability of an α-electron being at $\mathbf{r}_2$ when another is at $\mathbf{r}_1$. Since $h^\alpha(\mathbf{r}_1,\mathbf{r}_2)< 0$, the role of the Fermi hole is to decrease the α-spin density at $\mathbf{r}_2$ by an amount determined by the extent of the delocalization of the Fermi hole away from $\mathbf{r}_1$.

One sees from Eq. 8.4 that the product $\rho^\alpha(\mathbf{r}_1)h^\alpha(\mathbf{r}_1,\mathbf{r}_2)$ equals the HF exchange density. When this product is integrated over the coordinates of both electrons it yields $-N^\alpha$, the negative of the total number of α-electrons, that is, the total Fermi correlation for electrons of α-spin. Thus the double integration of this product over some region A, a quantity denoted by $F^\alpha(A,A)$ in Eq. 8.5, yields the total Fermi correlation for the α-electrons in the region A: (*Bader and Stephens, 1975*)

$$F^\alpha(A,A) = \int_A d\mathbf{r}_1 \int_A d\mathbf{r}_2 \; \rho^\alpha(\mathbf{r}_1)\, h^\alpha(\mathbf{r}_1,\mathbf{r}_2). \tag{8.5}$$

Its limiting value is $-N^\alpha(A)$, the negative of the α-spin population of region A, a value which would correspond to the electrons in A being totally localized to this region since all remaining α-spin density would then be excluded from A. The limiting value implies that the electrons in A do not exchange with electrons outside of A, i.e., they would be totally localized within A. Thus a *localization index* ($\lambda(A)$) is defined:

$$\lambda(A,A) = \left|F^\alpha(A,A)\right| + \left|F^\beta(A,A)\right|. \tag{8.6}$$

The limit of total localization, while approached quite closely ($\geqslant$95%) in ionic systems, can never be attained and one finds that $|F^\alpha(A,A)| < N^\alpha(A)$, indicating that the electrons in region A exchange with electrons outside the boundaries of A, i.e., they are delocalized. The delocalization of the electrons from a region A into another region B is determined by the quantity $F^\alpha(A,B)$, Eq. 8.4: (*Bader and Stephens, 1975*)

$$F^\alpha(A,B) = \int_A d\mathbf{r}_1 \int_B d\mathbf{r}_2 \; \rho^\alpha(\mathbf{r}_1)\, h^\alpha(\mathbf{r}_1,\mathbf{r}_2). \tag{8.7}$$

One necessarily has $F^{\alpha}(A,B) = F^{\alpha}(B,A)$ and the exchange of electrons between the two regions – the delocalization of the electrons between the two regions - is given by the delocalization index ($\delta(A,B)$) defined by: (*Fradera et al., 1999*)

$$\delta(A,B) = 2\left|F^{\alpha}(A,B)\right| + 2\left|F^{\beta}(A,B)\right|. \tag{8.8}$$

The integrals $F^{\alpha}(A,A)$ and $F^{\alpha}(A,B)$ determining the localization and delocalization of electrons respectively, are obtained by the corresponding double integrations of the Hartree-Fock exchange density, as illustrated for $F^{\alpha}(A,B)$:

$$F^{\alpha}(A,B) = -\sum_i \sum_j \int_A d\mathbf{r}_1 \int_B d\mathbf{r}_2 \left[\phi_i^*(\mathbf{r}_1)\,\phi_j(\mathbf{r}_1)\,\phi_j^*(\mathbf{r}_2)\,\phi_i(\mathbf{r}_2)\right] = -\sum_i \sum_j S_{ij}(A)\,S_{ji}(B). \tag{8.9}$$

The second-order density matrix obtained from a CI calculation can also be expressed in terms of products of basis functions multiplied by the appropriate coefficients enabling one to express the integrated pair density in terms of overlap contributions (Eq. 1.41). Thus, terms similar to those in Eq. 8.9 multiplied by the appropriate coefficients appear in the CI expression for $F^{\alpha}(A,B)$ and delocalization is still described in terms of the exchange of electrons between molecular orbitals in a wave function that includes Coulomb as well as Fermi correlation. (*Fradera et al., 1999*)

The above ideas, when used in conjunction with QT-AIM enable one to determine the extent to which electrons are localized within the basin of a given atom or delocalized into the basin of a second atom. Because the Fermi correlation counts all of the electrons, the localization and delocalization indices sum to N and they provide a quantitative measure of how the N electrons in a molecule are localized within the individual atomic basins and delocalized between them, this is expressed:

$$N(A) = \lambda(A) + \frac{1}{2}\sum_{B \neq A} \delta(A,B). \tag{8.10}$$

The delocalization index for a pair of bonded atoms - atoms whose nuclei are linked by a bond path (*Bader, 1998b*) - exhibits the properties associated with a Lewis bond order. (*Bader and Matta, 2001a; Fradera et al., 1999*) For a pair of identical bonded atoms or one with small

interatomic charge transfer, its value equals the number of Lewis electron pairs shared between the two atomic basins: 1.0 for H_2, 3.0 for N_2, 0.99 and 0.97 for CC and CH in saturated hydrocarbon, 1.9 and 0.98 for the same atomic pairs in ethylene. The values for a polyatomic molecule are invariably somewhat less than the limiting integer values because of the delocalization of density into the basins of other atoms linked to the pair in question (Eq. 8.10). Theory provides a precise determination of bonding – two atoms being bonded if their nuclei are linked by a bond path. (*Bader, 1998b*) Hence the $\delta(A,B)$ will be referred to as a "bond order" only between atoms sharing a bond path and in absence of significant charge separation.

There is close connection between Coulson's mobile bond order and the Fermi correlation, as described in Salem's book on molecular orbital theory of conjugated systems. (*Salem, 1966*) Within this theory, atomic charges and bond orders are determined by the products of coefficients of the carbon atomic π-orbitals (AOs) obtained when a given MO is squared to obtain an expression for its density, the products of the AOs themselves being equated to unity. A product of coefficients is identified with an atomic charge when both AOs refer to the same atom and with a bond order when they are on different carbons. When these products are summed over all MOs, one obtains q_r, the atomic charge for atom r and p_{rs}, the mobile bond order for atoms r and s. Using this formalism, the same-spin pair density, Eq. 8.3, reduces to:

$$\rho^{\alpha\alpha}(\mathbf{r}_1,\mathbf{r}_2) = \frac{1}{4}\left[q_r q_s - p_{rs}^2\right], \tag{8.11}$$

which is to be compared with Eq. 8.3. To quote Salem "The square of the bond order between atoms r and s represents the extent of the *total* (Fermi) *correlation between the two electrons with parallel spin*, one at r, the other at s", that is, it equals the exchange of electrons between r and s. (*Salem, 1966*) This expression of the Fermi correlation has been used by McConnell to account for the long-range spin-spin coupling between protons. (*McConnell, 1956*)

8.5 Computational Methods

The Gaussian94 (*Frisch et al., 1995*) package was used to obtain RHF//6-31G**/6-31G** densities which were analyzed using AIMPAC. (*Biegler-König et al., 1982*) Delocalization indices were calculated using a program written by the author, AIMDELOC, and which is listed in appendix 8.1. (*Matta, 2001a*) Molecular graphs were plotted using AIM2000. (*Biegler-König et al., 2001*)

The total energies compared to the sum of atomic energies as well as other criteria for measuring the numerical accuracy of atomic integrations are presented in table 8.1. The total energies calculated in the present work and those reported in the literature at the same level of theory (*Howard and Krygowski, 1997*) are identical except that the values for phenanthrene and anthracene appear to be interchanged. At this level of theory, we found that phenanthrene is *more stable* than anthracene by 6.9 kcal/mole.

Table 8.1. Total energies and measures of numerical accuracy of the calculations

Molecule	SCF Energy[a]	$E_{SCF} - \Sigma E(\Omega)$[b]	N[c]	$\Sigma[(1/2)\delta+\lambda]$[d]	$\Sigma q(\Omega)$[e]
Ethane	-79.23824	-1.7	18	18.0181	-0.0091
Ethylene	-78.03884	-0.1	16	16.0018	-0.0009
Cyclohexane	-234.22625	0.7	48	47.9918	0.0041
Cyclohexene	-233.03544	1.5	46	45.9791	0.0104
1,3-Cyclohexadiene	-231.84522	1.5	44	43.9842	0.0079
Benzene	-230.71386	1.6	42	41.9859	0.0070
Naphthalene	-383.36936	1.7	68	67.9862	0.0069
Anthracene	-536.01669	2.1	94	93.9841	0.0079
Phenanthrene	-536.02765	0.6	94	93.9953	0.0023
Chrysene	-688.68239	0.0	120	120.0022	-0.0014
9,10-Dihydroanthracene	-537.18752	-1.6	96	95.9813	0.0093
9,10-Dihydrophenanthrene	-537.19150	-0.9	96	95.9860	0.0066
Tetralin[f]	-385.70577	-1.9	72	71.9809	0.0096

[a] SCF energy is the HF/6-31G** molecular energy in au

[b] $E_{SCF} - \Sigma E(\Omega)$ is the difference between the SCF energy and the sum of the integrated atomic energies in kcal/mol.

[c] N is the number of electrons.

[d] $\Sigma[(1/2)\delta+\lambda]$ is the sum of the localization plus one half of the delocalization indices in the molecule, which formally equals the total number of electrons N. Deviation of this sum from N is a measure of the accuracy of the atomic integrations of the overlap matrix elements.

[e] $\Sigma q(\Omega)$ is the sum of atomic charges, which should vanish for neutral molecules, providing another measure of the quality of atomic integrations, this time of the electron density.

[f] 1,2,3,4-tetrahydronaphthalene.

8.6 Results and Discussion

8.6.1 Aromaticity in Terms of the Electron Density and of the Pair Density

Figure 8.1 shows relationships among some bond properties (the bond length (BL), ρ_b , $\nabla^2\rho_b$, ϵ) and $\delta(A,B)$. The bond properties are defined in section 2.5.1. Figures 8.1(a)-(c) show that the C-C bond length is strongly correlated with ρ_b, $\nabla^2\rho_b$, and ϵ, properties that can be considered as measures of the bond order and which are all derived from the electron density. The bond length is also strongly correlated to $\delta(A,B)$ for bonded atoms, figure 8.1(d). The delocalization index between bonded atoms is also found to be highly correlated with ρ_b, $\nabla^2\rho_b$, or ϵ, figures 8.1(e)-(g). These results indicate how a property derived from the pair density, $\delta(A,B)$, provides the same bond order picture as to the one assigned on the basis of properties of the electron density. There is no a priori reason for properties derived from the electron density to be correlated to properties derived from the pair-density.

Going beyond the global aspects concerning the trends observed between the properties of the electron density among themselves and with the delocalization index, it is also useful to analyse the values of these properties for individual bonded pairs of atoms. The delocalization indices in ethane have the values δ(C,C') = 0.991 and δ(C,H) = 0.967, indicating that there is one electron pair shared between the two carbon atoms as well as between each carbon atom and its bonded hydrogen atom, consistent with the presence of only single bonds in this saturated molecule. For ethylene, while δ(C,H) = 0.982 still corresponds to one electron pair shared between the bonded carbon and hydrogen atoms, δ(C;C') = 1.890 indicates two electron pairs shared between the two carbon atoms and corresponds to a double bond. The values of δ(C,C') for 1,3-cyclohexadiene given in figure 8.2 clearly characterize two double bonds. Moreover, there is always one electron pair shared between a carbon atom and its bonded hydrogen atoms, as expected. The values of δ(C,H) and ρ_b of the hydrogen-to-carbon bonds are omitted from the discussion and from figure 8.2 as they are nearly constant across the

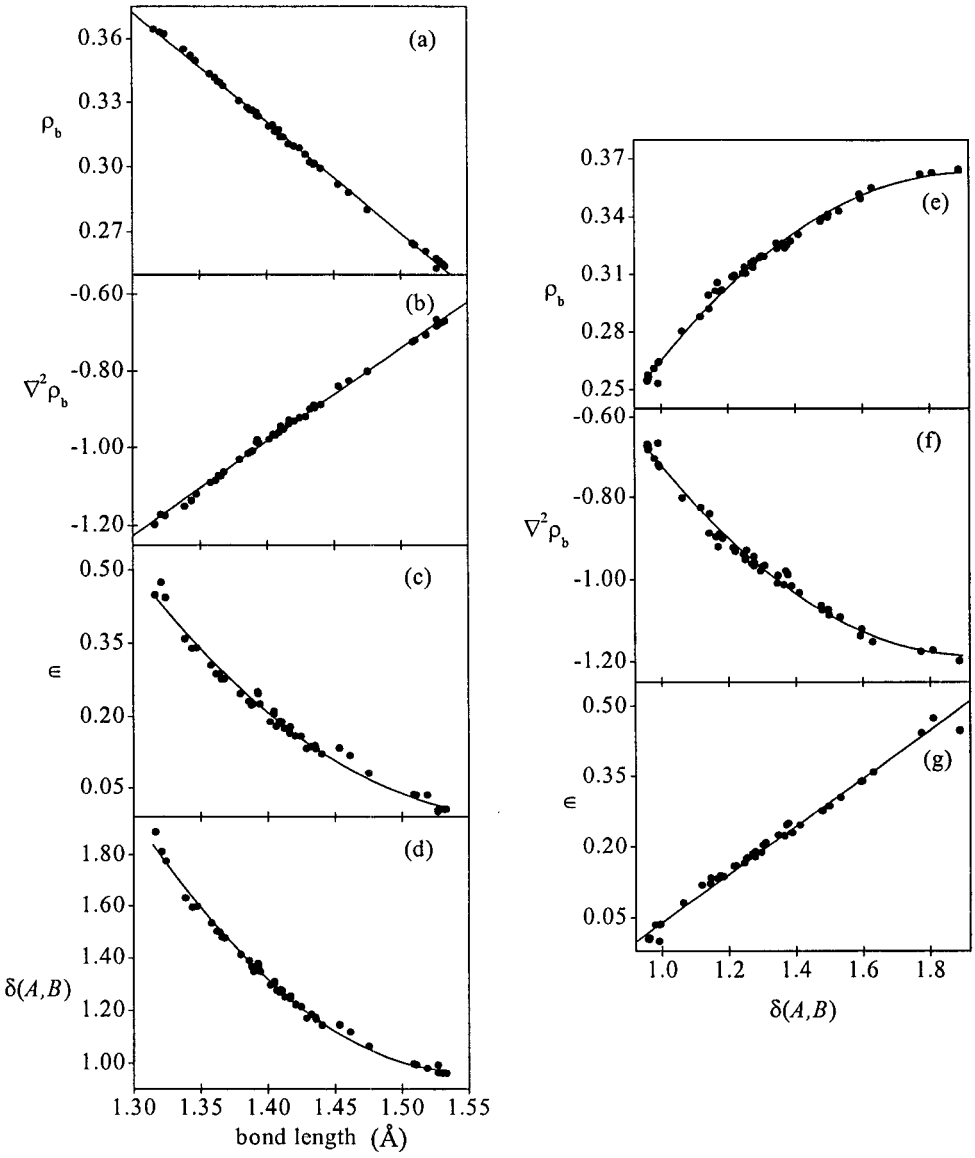
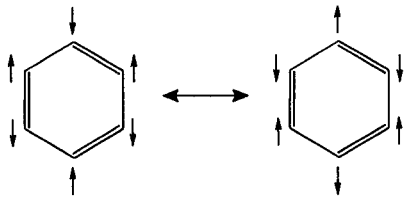


Figure 8.1 **(a)-(c)** Plots displaying the relationships between one-electron properties and bond length (BL); **(d)** between delocalization index and BL; and **(e)-(g)** between one-electron properties and delocalization index $\delta(A,B)$. The one-electron properties analyzed are the electron density (ρ_b), its Laplacian ($\nabla^2\rho_b$) and the ellipticity (ϵ); The regression equations (linear or quadratic), the correlation coefficients (r) and the standard deviations (σ) for each case are: **(a)** ρ_b = 1.0429 - 0.5157BL (r^2 = 0.998,σ = 0.0026); **(b)** $\nabla^2\rho_b$ = - 4.3852 + 2.4295BL (r^2 = 0.997,σ = 0.0086); **(c)** ϵ = 15.6520 - 19.7487BL + 6.2263BL^2 (r^2 = 0.986,σ = 0.0146); **(d)** $\delta(A,B)$ = 39.1643 - 49.3430BL + 15.9345BL^2 (r^2 = 0.994,σ = 0.0186); **(e)** $\delta(A,B)$ = 3.9645 - 25.2867ρ_b + 53.1159ρ_b^2 (r^2 = 0.992,σ = 0.0222); **(f)** $\delta(A,B)$ = 1.7125 + 2.6363$\nabla^2\rho_b$ + 2.2749$(\nabla^2\rho_b)^2$ (r^2 = 0.987,S = 0.0278); **(g)** $\delta(A,B)$ = 0.9268 + 1.9318ϵ (r^2 = 0.989,S = 0.0256). Included in the statistical analysis are 48 symmetry-unique carbon-to-carbon bonds in ethane, ethylene, cyclohexane, cyclohexene, cyclohexadiene, benzene, naphthalene, tetralene, anthracene, phenanthrene, and chrysene. ρ_b and $\nabla^2\rho_b$ in au, BL are in Å.

series of compounds considered in this study. Hence, the following focuses on carbon-carbon interactions.

Figure 8.2 displays the delocalization indices for bonded carbon atoms in several saturated and unsaturated molecules. The delocalization indices between neighbouring atoms in these molecules are in agreement with the expected Lewis structures. For benzene, the delocalization index for bonded C atoms is δ(C,C') = 1.388, a value that lies between those corresponding to ethane and ethylene, as expected. The magnitudes of δ(C,C') for non-bonded C atoms in the molecule are: C-C_{meta}=0.071, C-C_{para}=0.099. The greater value of $\delta(C,C_{para})$ than that of $\delta(C,C_{meta})$, despite the latter corresponding to a shorter distance, is a result of favorable spin pairing. The ground state of benzene has three π-electrons of α-spin and three of β-spin. Scheme 8.2 shows two equivalent resonance structures with acceptable spin pairing. Since the spins of π-electrons on carbon atoms that are *meta* with respect to each other are always the same, according to Pauli exclusion principle the densities of these same-spin electrons are expected to exclude each other with the result of a reduced *meta*-delocalization. On the other hand, since the spins of π-electrons on carbon atoms *para* to each other are opposite, the densities of these same-spin electrons are not expected to exclude one another with the result of an enhancement of the *para*-delocalization with respect to *meta*-delocalization.

Scheme 8.2



Based on its reactivity, benzene is the molecule that exhibits the greatest aromatic character in the series of polycyclic benzenoid aromatic hydrocarbons. (*Clar, 1972*) The molecules in the series lose some aromatic character (aromatic dilution), have more localized bonds, and a differential reactivity at different sites of a given molecule. The values of the delocalization indices for non-bonded carbon atoms (values not shown in figure 8.2), when taken together with those between bonded atoms, shed

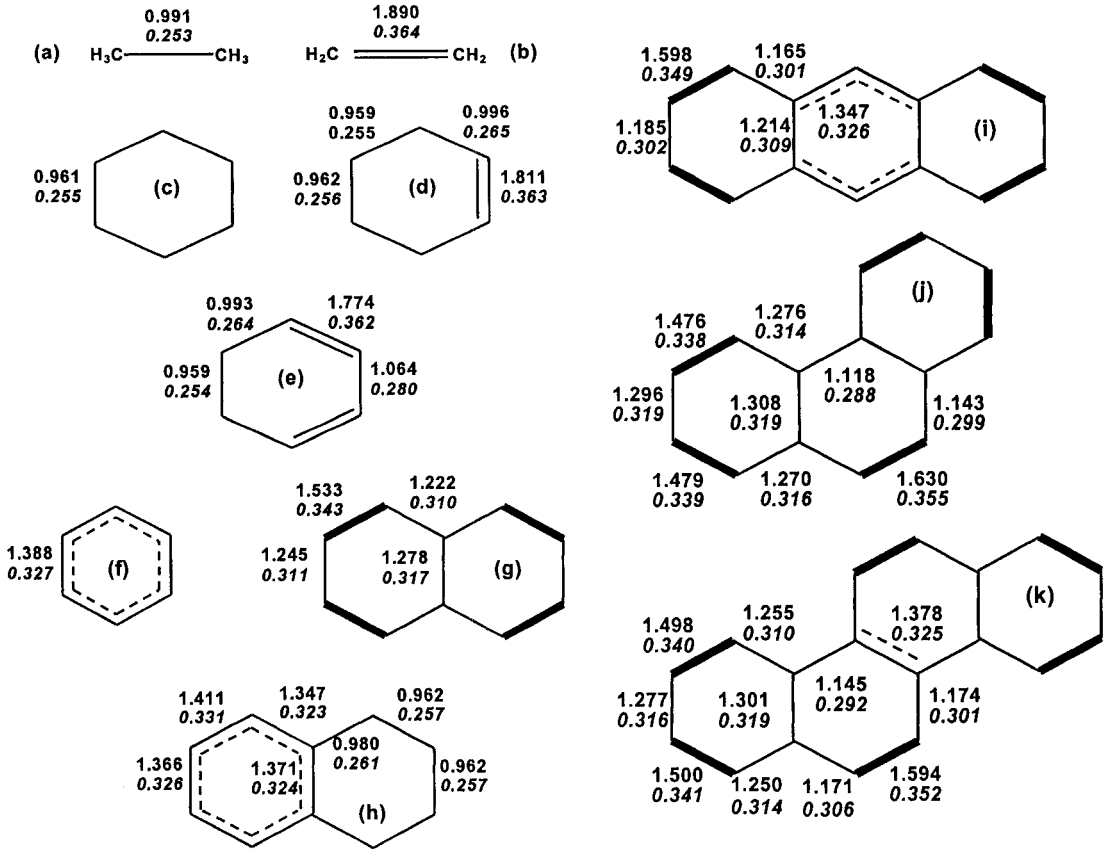


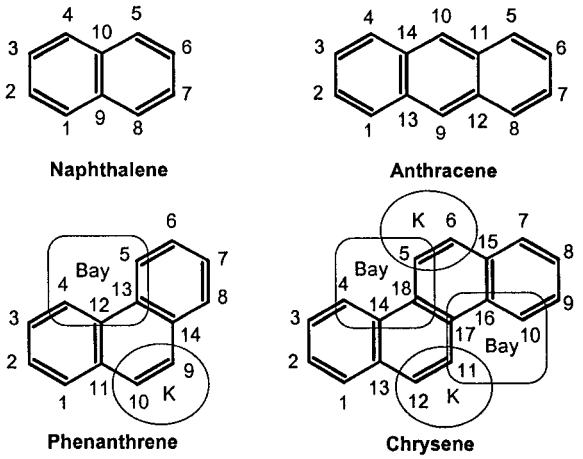
Figure 8.2 C-C Bond orders (delocalization indices δ(C,C') between bonded carbon atoms) - top number - and electron densities at the bond critical point ρ_b (bottom number in italic). **(a)** ethane, **(b)** ethylene, **(c)** cyclohexane, **(d)** cyclohexene, **(e)** 1,3-cyclohexadiene, **(f)** benzene, **(g)** naphthalene, **(h)** tetralin, **(i)** anthracene, **(j)** phenanthrene, and **(k)** chrysene. The following conventions are adopted: A simple single bond is drawn when $\rho_b < 0.338$ au and $\delta(A,B) < 1.476$; a double bond when $\rho_b > 0.362$ au and $\delta(A,B) > 1.774$; and a heavy (thickened) single bond when $0.338 \leqslant \rho_b \leqslant 0.355$ au and $1.476 \leqslant \delta(A,B) \leqslant 1.630$ au; and an aromatic bond is drawn when $0.325 \leqslant \rho_b \leqslant 0.327$ au and $1.347 \leqslant \delta(A,B) \leqslant 1.388$.

light on this differential reactivity, and their sum over a ring system is shown to provide a quantitative measure of the aromatic character of the ring.

The next molecule in the series is naphthalene. The C^1 and C^2 atoms in the molecule are commonly referred to as the α and β positions, respectively. (See scheme 8.3). Notice that there are four equivalent α and four equivalent β carbon atoms. Naphthalene exhibits preferential localization of electrons at C^{α}. In addition, the largest bond order in this molecule is found for the C^{α}-C^{β} bond

[$\delta(C^{\alpha},C^{\beta})$=1.533], the shortest bond (1.36 Å) and the one with the highest ρ_b (0.34 au) and ellipticity (0.31) in the molecule. (See definitions of bond properties in section 2.5.1). These values are intermediate between the corresponding ones in ethane and ethylene, but are closer to the values for ethylene.

Scheme 8.3



Numbering Scheme for Polycyclic Aromatic Hydrocarbons

The C^{α}-C^{β} bond in naphthalene involves the delocalization of *ca.* 0.15 pairs in excess of the number shared between two adjacent carbon atoms in benzene. This is mirrored in a higher ρ_b and shorter bond length than in benzene, all indicative of a higher bond order than in the latter molecule. These observations are in keeping with the known faster reaction rates of the molecule towards electrophilic substitution at the α position, compared to benzene, under conditions favouring kinetic control.

Using a similar analysis, unsaturation patterns can be assigned to all of the molecules in figure 8.2. The convention adopted in this figure is that a double bond is drawn between any two bonded carbon atoms when δ(A,B) and ρ_b are > 90% of the magnitude of the values for the double bond in ethylene. A dotted C⋯C bond is drawn for those bonds having $\delta(A,B)$ and ρ_b values that do not deviate more than 10% from the values in benzene, and single bonds are drawn when these values are < 10% of that in benzene. The bonds with $\delta(A,B)$ and ρ_b values falling approximately halfway between their values in benzene and those in ethylene are thickened. Although arbitrary, the lines

drawn in this way help to visualize the unsaturation patterns assigned on the basis of bond orders obtained from quantum mechanical observables.

The highest bond order in phenanthrene is found between C^9 and C^{10}. The value of this delocalization index, $\delta(C^9,C^{10}) = 1.630$, lies between $\delta(C,C')$ of benzene and ethylene. In addition, atoms C^9 and C^{10} also have the largest localization indices in the molecule. These observations are in agreement with the known olefinic character of the C^9-C^{10} bond manifested by the ability of the molecule to undergo addition as well as oxidation and reduction reactions at this bond under the proper experimental conditions (see structure III in scheme 8.1 for example). In the case of anthracene, notice the high degree of dienic character of a four carbon atom fragment in the central ring, for which the values $\delta(C^9,C^{12}) = \delta(C^{10},C^{11}) = 1.347$ and $\delta(C^{11},C^{12}) = 1.214$, are in agreement with the characteristic addition reactions of the molecule to yield 9,10-disubstituted dihydroanthracenes. Chrysene is similar to phenanthrene in its reactions because the molecule can be oxidised to the 5,6-quinone. This reactivity is also in line with the delocalization indices in figure 8.2 from which the largest value is 1.594 and corresponds to the C^5-C^6 pair.

The main features of the unsaturation pattern deduced on the basis of ρ_b and $\delta(A,B)$, depicted in figure 8.2, can also be obtained from the electron density isosurfaces. Two isodensity envelopes are displayed for each molecule in figure 8.3. The outer envelope is the 0.001 au isodensity envelope representing the van der Waals surface of the molecule in the gas phase. The inner one is that having the value of the density at the bond critical point in benzene ($\rho_{b(benzene)}$=0.327au). This envelope encloses without breaks any nuclei of bonded atoms having a ρ_b value higher than that between two carbon atoms in benzene, and encloses a larger volume the larger the value of ρ_b. On the other hand, when ρ_b is less than in benzene, the envelope will break into two separate closed regions each surrounding one nucleus of the two bonded atoms. The visual unsaturation patterns provided by the $\rho_{b(benzene)}$ isodensity envelopes in figure 8.3 reproduce those assigned in terms of properties of the electron density and of the pair desnity (figure 8.2).

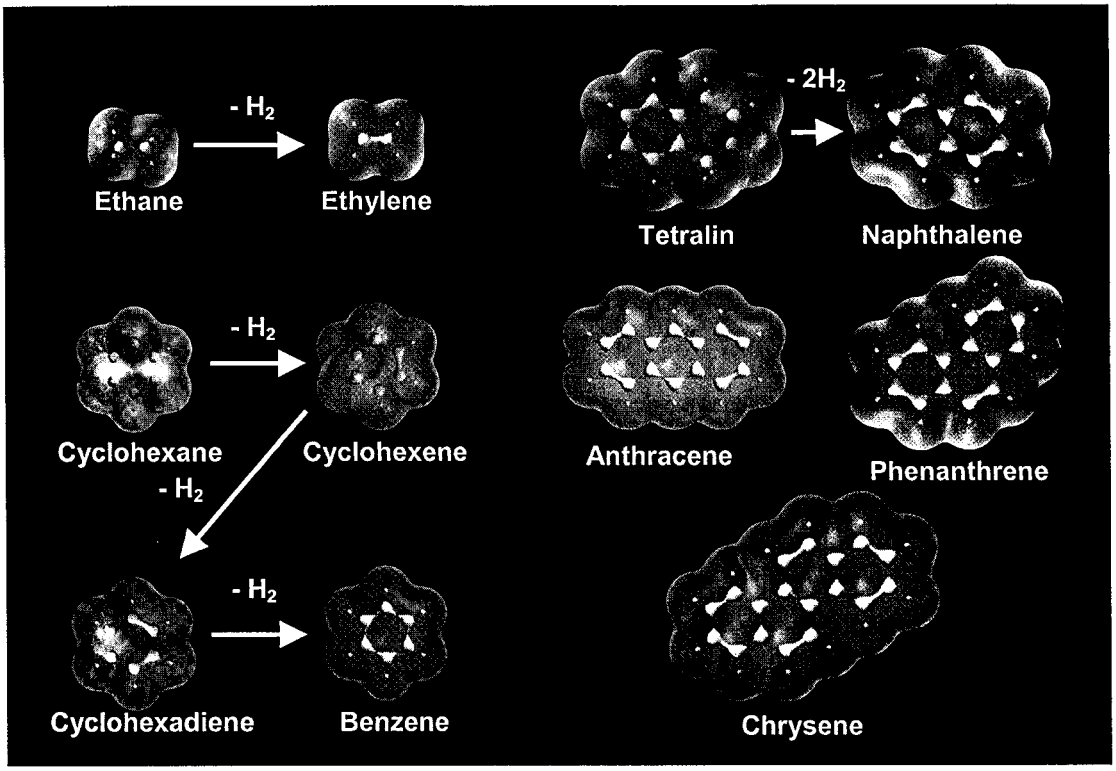


Figure 8.3 Three-dimensional electron isodensity envelopes of some saturated and unsaturated hydrocarbons. The outer (transparent) isodensity surface is the 0.001 au envelope which corresponds to the van der Waals envelope in gas phase and which typically contains more than 99% of the molecular electron population. The inner isodensity surface has the value of ρ_b for a carbon-carbon bond in benzene (0.3274 au).

8.6.2 Origin of the Aromatic Sextet

The previous analysis exploits to a great extent the information that the delocalization of the Fermi hole and the topological properties of the electron density provide on the chemistry of polycyclic aromatic hydrocarbons. The properties and reactivity of these compounds had been fully analyzed by Clar (*Clar, 1972*) in terms of a model that considers the π-electrons in the molecule. In this model, π-electrons are arranged so that the molecule has the maximum possible number of sextets, viewed in terms of Kékulé structures. In this way, there can only be one sextet in naphthalene or anthracene, located at any of the rings of each molecule, indicating that there can not be more than one aromatic

ring in the molecule at the same time. Phenanthrene can have either one or two sextets, as illustrated in scheme 8.1, from which structure III is the most stable one according to Clar's assumptions. The ability of Clar's model to explain a number of properties of a wide range of polycyclic aromatic hydrocarbons gives support to the usefulness of the model in the understanding of the chemistry of polycyclic aromatic hydrocarbons. (*Li and Jiang, 1995*) In this section, we seek to show that the delocalization indices account for, and provide a quantitative foundation for, Clar's model.

The sum of the delocalization indices between a particular carbon atom and the remaining five carbon atoms is 3.017 in benzene but only 2.010 in cyclohexane. These values correspond respectively to two σ- and one π-bonds in benzene and only two σ-bonds in cyclohexane. Thus, in benzene there are three electron pairs shared between a particular carbon atom and the remaining five carbon atoms in the ring, while in cyclohexane there are only two pairs shared between a carbon atom and the remaining five. Since every carbon atom participates equally in any electron pair in these rings, one carbon atom contributes three electrons to the benzene ring and only two to the cyclohexane ring. The same analysis for the other five carbon atoms in each molecule leads to the number of electron pairs shared in a ring and hence to the total number of electrons delocalized in the ring. If one subtracts the total ring contribution of the carbon atoms in cyclohexane (6 x 2.0) from the values in benzene (6 x 3.0) the result is 6.0 electrons delocalized over the benzene ring *in excess to cyclohexane.* In other words, six electrons (a sextet) are responsible for the aromatic character exhibited by benzene. Of course, one can obtain this result for benzene by taking the π-electron population only, since in planar molecules σ- and π-orbitals are orthogonal, a fact reflected in a block diagonal structure of the atomic overlap matrices. However, the procedure of subtracting the value of the total delocalization of the aromatic ring from the corresponding saturated one is a generalization that can apply for example in the case of polycyclic aromatic hydrocarbons departing from planarity as is the case in some helicenes and bucky balls. The calculation of the ring contributions above can be extended to analyse molecules in other cyclic systems. Thus, one can define the *aromatic electrons count* (θ) for any cyclic hydrocarbon as:

$$\theta = \left[\sum_{i=1}^{n_C}\sum_{\substack{j=1\\ j\neq i}}^{n_C}\delta(C_i,C_j)\right]_{\text{unsat}} - \left[\sum_{i=1}^{n_C}\sum_{\substack{j=1\\ j\neq i}}^{n_C}\delta(C_i,C_j)\right]_{\text{sat}} . \qquad (8.12)$$

In this expression n_C is the number of carbon atoms in the ring, "unsat" refers to the ring in the unsaturated molecule and "sat" to the saturated one. Since the polycyclic aromatic hydrocarbons analysed here are all benzenoid (having a 6-carbon atom skeleton), the saturated non-aromatic prototype molecule is cyclohexane. At the level of theory of this work, the contribution of cyclohexane (second term in Eq. 8.12) is - 6 x 2.010 = -12.060 electrons. θ defined in Eq. 8.12 will be used to explain Clar's structures. Table 8.2 shows θ for the rings in a number of cyclic molecules. In the case of cyclohexane, θ = 0 by definition, while θ = 1.946 and 3.935 for cyclohexene and 1,3-cyclohexadiene, respectively, indicating one and two electron pairs, in agreement with the presence of one and two double bonds.

The aromatic electron count for benzene (θ = 6.045) correctly predicts the delocalization of one sextet of electrons in the ring. In the case of naphthalene, θ = 5.214 is less than the value for benzene, an indication of aromatic dilution which in turn is a consequence of the impossibility of having the two equivalent rings aromatic at the same time. For phenanthrene, Fermi delocalization predicts the outer rings to be more aromatic than the central one. This result and the olefinic character of the C^9-C^{10} bond indicated by the delocalization index of this individual pair of atoms, clearly agrees with Clar's structure for the molecule (Structure III of scheme 8.1).

According to Clar's model, from a set of isomers the most stable molecule is the one having the maximum number of sextets. The delocalization of the Fermi hole also predicts this behaviour. Even though the outer rings of anthracene are more aromatic than the central one, θ for these rings is considerably less than the corresponding value in phenanthrene, a consequence of having a greater number of sextets in the latter molecule than in the former. The results in table 8.2 also show that the larger the number of fused rings in the acene series, the lesser the benzene-like character of the rings in the molecules, as illustrated by the series benzene, naphthalene, and anthracene, whose largest θ values

are 6.045, 5.214 and 4.877, respectively. This result can be correlated to the shift of the β-absorption bands in the electronic spectra, from λ_β = 1852 through 2210 to 2515 Å, a consequence of the linear annelation effect for this series of molecules.

Deviations of θ values of a particular ring from the ideal aromatic value of 6 is a measure of the loss of aromatic character. Whereas the rings in naphthalene are not fully aromatic, hydrogenation to yield tetralin (1,2,3,4-tetrahydronaphthalene) leads to a molecule where one of the rings has become fully (97%) aromatic, as indicated by the change of θ from 5.214 to 5.863. From a similar comparison one finds that hydrogenation of the inner ring of anthracene leads to a 16% increase in the aromatic character of the outer rings. A similar yet less pronounced trend occurs in phenanthrene, the outer rings of which gain 6% in aromatic character upon hydrogenation of the inner ring.

It is also possible to calculate the delocalization between the two external entire rings in both anthracene and phenanthrene, and in the corresponding dihydrogenated compounds. This is achieved using Eq. 8.7 considering the six carbon atoms forming each ring as a single functional group (*A* or *B*). Upon hydrogenation of the inner ring of anthracene, the delocalization between the two external rings reduces from 0.421 to 0.173 while in the case of phenanthrene the inter-ring delocalization decreases from 1.510 to 1.378. The trends in these values are as expected; the effect of hydrogenation of the inner ring is to "break the communication" between the outer rings which in turn become fully (over 96%) aromatic, due to the elimination of aromatic dilution.

It can be concluded that the empirical aromatic sextet rule and the chemistry of polycyclic aromatic hydrocarbons can be readily understood in terms of the delocalization of the Fermi hole between the atoms in the molecule.

Table 8.2. Aromatic electron count (θ) of polycyclic aromatic hydrocarbons and some non-aromatic cyclic molecules

Molecule	Outer Ring	%[a]	Inner Ring	%[a]
Cyclohexane	0.0000	0	-	-
Cyclohexene	1.9458	32	-	-
1,3-Cyclohexadiene	3.9348	65	-	-
Benzene	6.0450	100	-	-
Naphthalene	5.2140	86	-	-
Anthracene	4.8770	81	4.7146	78
Phenanthrene	5.4369	90	4.1778	69
Chrysene	5.3635	89	4.4807	74
9,10-Dihydroanthracene	5.8634	97	1.8048[b]	30
9,10-Dihydrophenanthrene	5.7749	96	1.8100[b]	30
Tetralin[c]	5.8630	97	0.9422[b]	16

[a] Percent aromatic character defined as $(\theta_{ring}/\theta_{benzene})x100$.
[b] Value for the hydrogenated ring.
[c] 1,2,3,4-Tetrahydronaphthalene.

8.6.3 Correlation Between the Total Aromatic Count (θ_t) and Resonance Energies

The *total* number of aromatic electrons of the carbon atom framework, θ_t, of the polycyclic aromatic hydrocarbons is highly correlated with their experimental delocalization energies. The total aromatic count is calculated by subtracting the contribution of a carbon atom in cyclohexane (2.010 e) from the contribution of every carbon atom in the polycyclic aromatic molecule to the other C atoms, followed by a sum over all contributions in the molecule. A set of θ_t values and experimental resonance energies obtained from heats of combustion are given in table 8.3 for a number of molecules. Figure 8.4 shows an empirical correlation between these two quantities. The ability of θ_t to account for these experimental values is reflected in the magnitude of the linear regression correlation coefficient (r^2=0.994). Notice that the delocalization energy of naphthalene is not twice that of benzene, nor the delocalization energy of anthracene or phenanthrene is three times that of benzene, a consequence of these molecules not having all their rings aromatic at the same time. In the case of the hydrogenated derivatives, the delocalization energy of 9,10-dihydroanthracene is twice that of benzene. The values of θ_t of these molecules also display the same trends as the delocalization energies: while θ_t of

phenanthrene and anthracene are not three times that of benzene, θ_t of 9,10-dihydroanthracene is 2.12 times that of benzene.

Table 8.3. Total aromatic count (θ_t) and resonance energy (kcal/mol) of polycyclic aromatic hydrocarbons and some partially saturated derivatives

Molecule	θ_t	Resonance energy[a]
Benzene	6.0450	36.0
1,2,3,4-Tetrahydronaphthalene	6.3821	-
Naphthalene	10.4541	61.0
9,10-Dihydroanthracene	12.8026	72.2
9,10-Dihydrophenanthrene	14.5704	-
Anthracene	14.8580	83.5
Phenanthrene	14.8856	91.3
Chrysene	19.3271	116.5

[a] Taken from (*Wheland, 1955*)

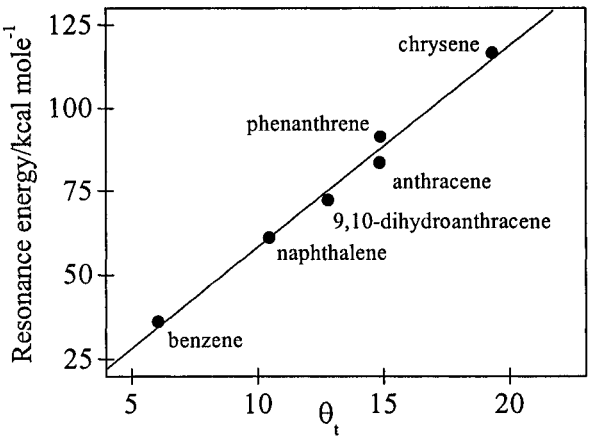


Figure 8.4 Statistical correlation between experimental resonance energies and total aromatic count (θ_t). The correlation coefficient of the fitting is 0.994.

8.6.4 The Isomerization of Anthracene to Phenanthrene and the Effect of the Presence of a Dihydrogen Bond on their Relative Stability

The existence of a bond path is invariably accompanied by the existence of another line in space of maximally negative potential energy density (i.e. a line of maximal stability), called the virial path, linking the same bonded nuclei. (*Keith et al., 1996*) The stabilization due to a particular bonding

interaction can be traced back to individual atomic contributions since one can define atomic energies ($E(\Omega)$) in QT-AIM (Eq. 2.67). These ideas are used here to show that the presence of a dihydrogen bond is a source of stability in angular polycyclic aromatic hydrocarbons. The analysis presented here is in sharp contrast with the one presented in the paper where this interaction was first discovered in polycyclic aromatic hydrocarbons, and in which it is referred to erroneously as a "nonbonding hydrogen-hydrogen interaction" despite the presence of a bond path and its associated interatomic surface. (*Cioslowski and Mixon, 1992*) Contrary to the main arguments in Cioslowski and Mixon's paper, in this section, evidence is provided for the stabilizing nature of this *bonding interaction*. The dihydrogen bond was subsequently found in several other systems, (*Alkorta et al., 1998; Bodige et al., 1999; Garbowski, 2000; Garbowski, 2001; Matta and Bader, 2000; Popelier, 1998*) and is currently the subject of an ongoing investigation in the Bader research group.

Phenanthrene is more stable than anthracene by 6.9 kcal/mole. The atomic basis of this energy difference can be identified by comparing the total energies of the set of carbon atoms and of the set of hydrogen atoms separately in the two molecules. One finds that the total energy of all the hydrogen atoms is 15.9 kcal/mol lower in phenanthrene than in anthracene, while the total energy of all the carbon atoms is 10.5 kcal/mol higher in phenanthrene than in anthracene. The net effect, 5.4 kcal/mole, accounts for the difference between these two molecules (the deviation from 6.9 kcal/mol is due to numerical integration errors). In phenanthrene, the atomic energy of H^4 or H^5 (the hydrogen atoms bonded to C^4 and C^5, respectively) is 6.4 kcal/mole lower than the average energy of the hydrogen atoms in anthracene. On the other hand, the average energy of the eight remaining hydrogen atoms in phenanthrene (not including H^4 and H^5) is only 0.4 kcal/mole lower than the average energy of the hydrogen atoms in anthracene. This means that out of the 15.9 kcal/mole of excess stability provided by the hydrogen atoms in phenanthrene over anthracene, 12.8 kcal/mole (over 80%) is provided by H^4 and H^5 alone and the remainder is provided by the rest of the hydrogen atoms (8X0.4=3.2 kcal/mole). The isomerization of anthracene into phenanthrene leads to the presence of one K region and one bay region (scheme 8.3). The bay region is accompanied by a new topological

feature in the electron density: a hydrogen-to-hydrogen (dihydrogen or H-H) bond path forming a new six-membered ring consisting of C^{12}-C^{4}-H^{4}-H^{5}-C^{5}-C^{13}. The dihydrogen bond path and the accompanying interatomic surface in phenanthrene are displayed in figure 8.5. It is also clear from the figure that this topological feature is missing in its isomer anthracene. The existence of this bond imparts most of the stability of phenanthrene over anthracene by lowering the atomic energies of the two hydrogen atoms involved, 6.4 kcal/mol each.

Chrysene is another molecule where K and bay regions are identified and as in the case of phenanthrene, the bay regions are also accompanied by the appearance of a dihydrogen bond and a corresponding ring structure. Moreover, for both phenathrene and chrysene, the properties at the dihydrogen bond critical point and the delocalization between the corresponding H-H atoms have the same values to the number of decimals shown: ρ_b=0.01 au, $\nabla^2\rho_b$=0.05 au, ϵ=0.4, and $\delta(H^4,H^5)$=0.03. The existence of H-H bonding with very similar bond properties has also been found in the hydrocarbon part of the amino acid leucine. (*Matta and Bader, 2000*) The dihydrogen bond occurs not only in the planar acene molecules, but it persists even for significant departures from planarity as in 9,10-dihydrophenanthrene (where the C^{4}-C^{12}-C^{13}-C^{5} dihedral angle = -22.8°). See figure 8.6. Thus, the H-H bond appears to be ubiquitous to the angular acene series imparting a stability of *ca.* 12kcal/mol per H-H bond, an effect absent in the corresponding linear acenes.

8.6.5 Proton Spin-Spin Coupling and Electron Exchange

The spin-spin coupling of protons is dominated by the product of the Fermi contact terms for the two nuclei. (*Ramsey, 1953; Stephen, 1957*) The spin magnetic moment of a proton polarizes the spin density of the electrons in its immediate vicinity, resulting in a small excess of oppositely polarized electron spin density. This information is transmitted to the vicinity of a second proton via the mechanism of electron exchange. If one interprets the pair density as a pair probability distribution and denotes the position coordinates of two protons by **n** and **n'**, then $\rho^{\alpha\beta}(\mathbf{n},\mathbf{n}')$, the probability of an α-electron being at **n** and a β-electron at **n'** is given by the (uncorrelated) product $\rho^{\alpha}(\mathbf{n})\,\rho^{\beta}(\mathbf{n}')$. However,

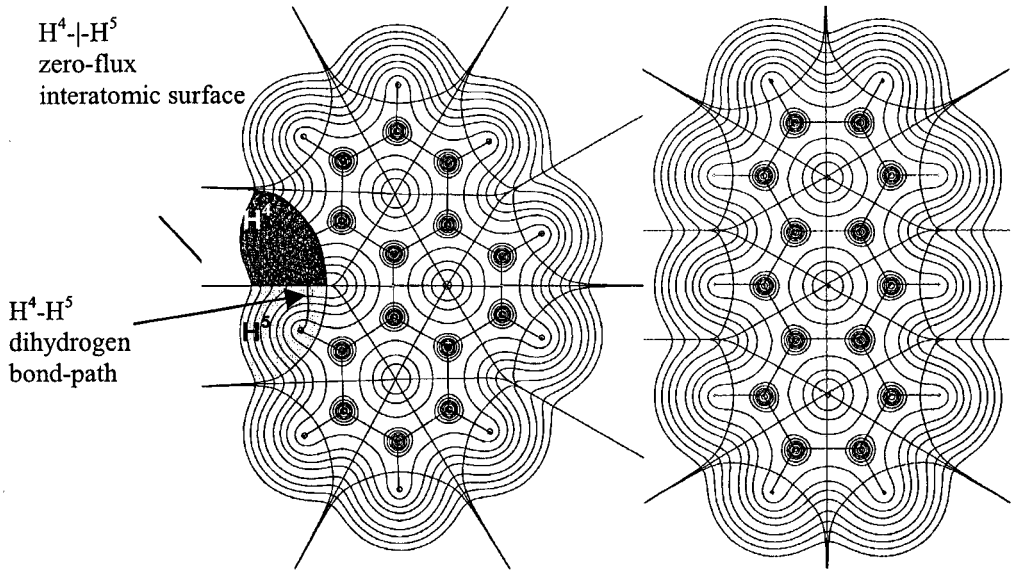


Figure 8.5 Contours of $\rho(\mathbf{r})$ in the molecular planes of phenanthrene (left) and of anthracene (right). The interatomic surfaces and bond paths are also shown for both isomers. The bounded atomic regions of H^4 and H^5 in phenanthrene are shadowed in different shades of grey.

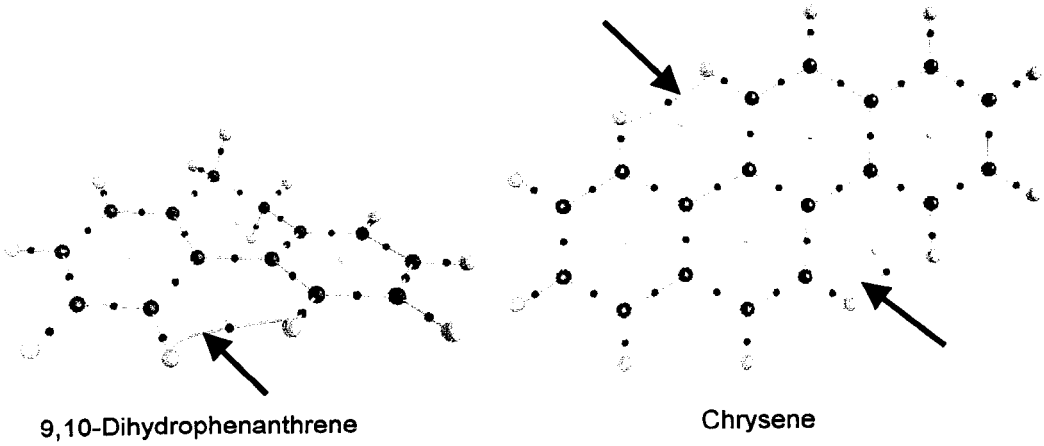


Figure 8.6 Molecular graphs of 9,10-dihydrophenanthrene and of chrysene. The lines linking the nuclei are the bond paths, the red dots on the bond paths are the bond critical points, and the yellow dots inside rings are the ring critical points. The dihydrogen bond paths are indicated by the arrows.

$\rho^{\alpha\alpha}(\mathbf{n},\mathbf{n}')$, the probability of an α-electron being at **n** when another α-electron is at **n'** is mediated by the density of the Fermi hole and is given by $\rho^{\alpha}(\mathbf{n})\ \rho^{\alpha}(\mathbf{n}') + \rho^{\alpha}(\mathbf{n})\ h^{\alpha}(\mathbf{n},\mathbf{n}')$, Eq. 8.3. For a closed shell system, $\rho^{\alpha}(\mathbf{r}) = \rho^{\beta}(\mathbf{r})$ and the difference $\rho^{\alpha\beta}(\mathbf{n},\mathbf{n}') - \rho^{\alpha\alpha}(\mathbf{n},\mathbf{n}')$ reduces to $-\rho^{\alpha}(\mathbf{n})\ h^{\alpha}(\mathbf{n},\mathbf{n}')$, that is, to minus the exchange density. Thus the probability of β-electrons being in excess over α electrons at the position of nucleus n', given that there is an α-electron at nucleus n, is proportional to the negative of the exchange density. The density of α-spin electrons at nucleus n' is lessened to the extent that the density of the Fermi hole extends from the reference nucleus n to n' and this is the mechanism whereby the spin perturbation caused by magnetic interaction of an electron with nucleus n is transmitted to n'. (*Bader et al., 1996*)

McConnell showed that the same conclusion regarding the proportionality between the excess of β over α spins and exchange can be obtained from the second-order perturbation expression for the coupling of nuclear spins via the electron spin. (*McConnell, 1956*) There are three contributions to the Hamiltonian describing the perturbations arising from a magnetic nucleus interacting with the spin magnetic moment of an electron: (*Pople et al., 1959*) one involving the motion of the electrons in the magnetic field of the nuclei, another representing the dipole-dipole interaction between the nuclear and electronic and magnetic moments and thirdly, that arising from the coupling of the nuclear spins via the spin of the electrons. It is this latter term that dominates the spin-spin coupling of protons. It is determined by the product of matrix elements for the operator $\mathbf{S}_k \cdot \mathbf{I}_n$, the operators for the electronic and nuclear spin, a contribution that reduces to the product of the Fermi contact terms for the two nuclei in question. The matrix elements mix the singlet ground state with triplet excited states in the second-order expression for the energy of interaction. If one makes the approximation of replacing all of the triplet excitation energies by a mean value ΔE, then the perturbation expression reduces to the averaging of terms involving the operators $\mathbf{S}_j \cdot \mathbf{S}_k\ \delta(\mathbf{r}_{jn})\ \delta(\mathbf{r}_{kn'})$, products of the electronic spin operators mediated by the term giving the product of electron densities at the positions of the two nuclei n and

n', averaged over the ground state wave function. (The symbol $\delta(\mathbf{r}_{jn})$ used by McConnell, is an abbreviation of the Dirac delta operator $\delta(\mathbf{r}_j\text{-}\mathbf{X}_n)$, where $\mathbf{X}_n$ is the coordinate of nucleus n).

It is the averaging of this operator over the coordinates of all j,k pairs of electrons for the ground state wave function that yields an expression for the spin-spin coupling constant $J_{nn'}$ that is proportional to the negative of the exchange density. Specifically, the expression for the coupling constant in standard notation becomes: (*Pople et al., 1959*)

$$J_{nn'} = -\frac{2}{3h}\left(\frac{16\pi\beta\hbar}{3}\right)^2 \gamma_n \gamma_{n'} \frac{1}{\Delta E} \sum_i \sum_j \left\langle \phi_i(\mathbf{r}_1)\ \delta(\mathbf{r}_{1n})\ \phi_j(\mathbf{r}_1)\right\rangle \left\langle \phi_j(\mathbf{r}_2)\delta(\mathbf{r}_{2n'})\phi_i(\mathbf{r}_2)\right\rangle. \quad (8.13)$$

Eq 8.13 states that $J_{nn'}$ is proportional to the product of densities at the two nuclei that arise from the exchange of same-spin electrons. This general expression was first obtained by McConnell (*McConnell, 1956*) who pointed out that if one considers each hydrogen atom to be described by a 1*s* orbital, then the exchange term appearing in Eq. 8.13 reduces, as shown in Eq. 8.11, to a term proportional to Coulson's mobile bond order (*Coulson and Longuet-Higgins, 1947; Coulson and Longuet-Higgins, 1948*) between nonbonded hydrogen atoms multiplied by the product of spin densities at the two protons.

The dominance of the Fermi contact contribution to $J_{HH'}$ for nonbonded hydrogens is primarily a consequence of 1*s*-like functions dominating the basis set for the hydrogen atoms. It is therefore reasonable to assume that the density at a proton should parallel the increase in the density within the basin of a hydrogen atom, enabling one to make the identification:

$$\left\langle \phi_i(\mathbf{r}_1)\delta(\mathbf{r}_{1n})\phi_j(\mathbf{r}_1)\right\rangle \left\langle \phi_j(\mathbf{r}_2)\delta(\mathbf{r}_{2n'})\phi_i(\mathbf{r}_2)\right\rangle \rightarrow \left\langle \phi_i(\mathbf{r}_1)\phi_j(\mathbf{r}_1)\right\rangle_H \left\langle \phi_j(\mathbf{r}_2)\phi_i(\mathbf{r}_2)\right\rangle_{H'} = S_{ji}(H)S_{ij}(H'), \quad (8.14)$$

wherein one replaces the exchange density at the positions of the protons n and n' with its integration over the respective atomic basins H and H′. Hence, the double sum appearing in the expression for the coupling constant in Eq. 8.13 reduces, with a change in sign, to Eq. 8.9, the expression for $|F^{\alpha}(\mathrm{H,H'})|+|F^{\beta}(\mathrm{H,H'})| = \delta(\mathrm{H,H'})/2$, one-half the delocalization index that measures the exchange of electrons between the two atomic basins. *Thus one is led to propose that the coupling constant*

between the protons of nonbonded hydrogen atoms H and H′, be proportional to the delocalization index δ(H,H′). This proportionality is not expected to apply to atoms other than hydrogen because the presence of basis functions with non-zero angular momentum in a primary rather than a polarizing role for non-hydrogenic atoms increases the contributions of the other two coupling contributions relative to the Fermi contact term in the determination of $J_{nn'}$. The higher angular momentum basis functions also destroy the assumed proportionality between the atomic integration of the exchange density and its value at the nucleus of the atom.

One also anticipates that the model will be restricted to what McConnell termed the "long-range coupling of nuclear spins." (*McConnell, 1956*) One of the most striking features of proton spin-spin coupling is the relatively small values found for protons bonded to the same carbon atom, a result not recovered by the delocalization indices. The small values are attributed to the nodal properties of a pair of C-H bonding orbitals on one carbon exhibiting planes in the proximities of the protons bonded to the same carbon, (*Pople et al., 1959*) thus negating the proportionality between the integrated atomic exchange density and its value at a proton. This inter-orbital nodal effect is not found for hydrogens bonded to adjacent or more widely separated carbons.

8.6.6 Correlation of δ(HH′) with $J_{HH'}$

It was previously demonstrated (*Bader et al., 1996*) that the delocalization index for the vicinal protons in ethane yields an excellent correlation with their coupling constants as a function of the torsion angle about the C-C axis predicted by the Karplus equation. (*Karplus, 1959*) In the present work the correlation of δ(H,H′) was compared with measured coupling constants for hydrogen atoms bonded to different carbon atoms in alkanes, alkenes, their cyclic congeners and polycyclic hydrocarbons. An interesting feature of the measured constants $J_{HH'}$ is that they can be largest for protons that are separated by the greatest distance, a feature faithfully recovered by the corresponding values of δ(H,H′).

As seen in previous sections of this chapter, the fusion of one or more benzenoid rings results in the delocalisation of electrons increasing between certain pairs of bonded carbon atoms and decreasing between others, relative to that found in benzene. For example, the value of 1.39 for δ(C,C′) in benzene, increases to 1.63 for the unique 9-10 bond in phenanthrene (compared to a value of 1.89 for the CC bond in ethylene) while the values for the neighbouring CC bonds decrease to 1.14. Thus the hydrogens in these molecules are bonded to a carbon framework that exhibits bond orders lying between those for single and double bonds.

It is found that the values of δ(H,H′) yield a linear statistical correlation with the measured $J_{HH'}$ coupling constants expressed in cycles/sec given in the following regression equation:

$$J_{HH'} = -0.1369 + 1274.9 \times \delta(H,H'), \tag{8.15}$$

with $r^2 = 0.989$. The experimental and calculated values for $J_{HH'}$ using this equation are given in table 8.4 and the correlation between $J_{HH'}$ and δ(H,H′) is displayed in figure 8.7. The experimental values are grouped into two sets: those around 600, for protons separated by three bonds and those around 100 cycles/sec for protons with greater inter-bond separations, with the latter set extending down to and past zero. A negative coupling constant implies that the two protons that are coupled have the same spin. While three negative values of $J_{HH'}$ are obtained from the correlation compared to the single experimental example, all three are for couplings that are close to zero.

One notes that both the experimental and theoretical values for the 1-5 coupling in naphthalene are greater than for 1-4, even though the 1-5 internuclear separation exceeds that for 1-4. Similarly, the 1-5 coupling is greater than the 1-6 coupling, although the former pair of protons are separated by seven bonds, the latter pair by six.

An even more striking example of this behaviour is found for the *cis* and *trans* proton couplings in ethylene where the *trans* coupling is nearly twice that of the *cis*. Experimentally, the values for variously substituted ethylenic molecules fall in the ranges of 17-18 and 8-11 cycles/sec for the *trans* and *cis* coupling, respectively, (*Pople et al., 1959*) while Eq. 8.15 yields corresponding values of 16.2 and 9.7 cycles/sec, values that are bracketed by the experimental ones. Experimental

$J_{HH'}$ constants for the vicinal protons linked by a carbon-carbon double bond are known for both cyclohexene, 8.8 cycles/sec, and 1,3-cyclohexadiene, 9.4 cycles/sec, while those predicted using Eq. 8.15 are 10.5 and 9.3 cycles/sec, respectively.

The possibility of the π-system contributing to the coupling of aromatic protons via the mechanism of configuration interaction is considered to be relatively small. (*Pople et al., 1959*) McConnell (*McConnell, 1957*) has estimated the π-contribution to the coupling between the *ortho* protons in benzene to be 0.80 cycles/sec. The ability to linearly relate the coupling constants to the δ(H,H′) values further speaks to the relative non-importance of the π-contribution to the coupling, since there is no exchange between the σ- and π-electrons and the π-orbitals make no direct contribution to the hydrogen delocalization indices.

The proton spin-spin coupling constants for vicinal hydrogens in saturated hydrocarbons correlate equally well with the delocalization index, although they require a separate regression equation:

$$J_{HH'} = -0.7520 + 1209 \times \delta(H,H') . \qquad (8.16)$$

Eq. 8.16 was fitted to $J_{HH'}$ values for the vicinal hydrogens listed in table 8.5 to yield an r^2 value of 0.966. The magnitude of $J_{HH'}$ in saturated hydrocarbons can also exhibit the property of increasing with increasing internuclear separation, as exemplified by the coupling between the *trans* protons in staggered ethane predicted to be 15.5 cycles/sec compared to 5.3 cycles/sec for the *gauche* pair, a trend also predicted by the Karplus equation. (*Karplus, 1959*) Similarly the axial-axial coupling of neighbouring protons in cyclohexane which are *trans* related, is greater than for the axial-equatorial or equatorial-equatorial related pairs whose internuclear separations decrease in the same order. These variations are faithfully recovered by the relative values of the corresponding δ(H,H′) indices (table 8.5).

The success of the correlations obtained with Eqs. 8.15 and 8.16 speak not only to the physical soundness of the underlying model that relates the Fermi contact terms to electron

delocalisation, but also to the ability of the quantum definition of an atom to isolate the atomic or functional group contributions to measurable molecular properties.

Table 8.4. $J_{HH'}$ coupling constants in cycle/s in some aromatic molecules*

molecule	H	H'	δ(H,H') x10^5	$J_{HH'}$ Calc.	$J_{HH'(av)}$ Exptl.	σ	n	Reference
benzene	1	2	585	7.32	7.51	0.03	5	(*Memory and Wilson, 1982*)
benzene	1	3	139	1.64	1.34	0.03	5	(*Memory and Wilson, 1982*)
benzene	1	4	30	0.25	0.66	0.03	5	(*Memory and Wilson, 1982*)
naphthalene	1	2	650	8.14	8.28	0.19	4	(*Memory and Wilson, 1982*)
naphthalene	1	3	136	1.60	1.24	0.11	4	(*Memory and Wilson, 1982*)
naphthalene	1	4	30	0.24	0.00		1	(*Memory and Wilson, 1982*)
naphthalene	1	5	38	0.35	0.83		1	(*Memory and Wilson, 1982*)
naphthalene	1	6	4	-0.09	-0.16		1	(*Memory and Wilson, 1982*)
naphthalene	1	7	8	-0.03	0.21		1	(*Memory and Wilson, 1982*)
naphthalene	2	3	526	6.57	6.71	0.19	4	(*Memory and Wilson, 1982*)
naphthalene	2	4	136	1.60	1.24	0.11	4	(*Memory and Wilson, 1982*)
naphthalene	2	6	5	-0.07	0.28		1	(*Memory and Wilson, 1982*)
anthracene	1	2	674	8.46	8.30		1	(*Memory and Wilson, 1982*)
anthracene	1	3	135	1.59	1.20		1	(*Memory and Wilson, 1982*)
anthracene	2	3	501	6.25	6.50		1	(*Memory and Wilson, 1982*)
phenanthrene	1	2	623	7.81	8.42	0.17	3	(*Memory and Wilson, 1982*)
phenanthrene	1	3	141	1.66	1.21	0.28	3	(*Memory and Wilson, 1982*)
phenanthrene	1	4	28	0.22	0.58	0.06	3	(*Memory and Wilson, 1982*)
phenanthrene	2	3	539	6.74	7.13	0.21	3	(*Memory and Wilson, 1982*)
phenanthrene	2	4	127	1.48	1.50	0.09	3	(*Memory and Wilson, 1982*)
phenanthrene	3	4	670	8.40	8.14	0.23	3	(*Memory and Wilson, 1982*)
phenanthrene	4	10	36	0.32	0.40		1	(*Bartle et al., 1969*)
Chrysene	1	2	686	8.61	8.00		1	(*Memory et al., 1966*)
Chrysene	1	3	125	1.45	1.20		1	(*Memory et al., 1966*)
Chrysene	1	4	29	0.23	0.50		1	(*Memory et al., 1966*)
Chrysene	2	3	531	6.64	7.30		1	(*Memory et al., 1966*)
Chrysene	2	4	142	1.67	1.00		1	(*Memory et al., 1966*)
Chrysene	3	4	627	7.86	7.50		1	(*Memory et al., 1966*)
Chrysene	5	6	726	9.12	9.00		1	(*Memory et al., 1966*)

* *Explanation of the table*: The columns labeled H and H' list the standard numbering of the carbon atom to which the hydrogen atom in question is bonded. The column labeled δ(H,H') lists the value of the delocalisation indices. The column labeled $J_{HH'}$ lists the calculated spin-spin coupling constants obtained from the regression equation [Eq. 8.15]. The column labeled $J_{HH'(av)}$ provides the experimental value of the coupling between the two hydrogen atoms. Entries in the $J_{HH'(av)}$ are average values reported by n authors and collected in (*Memory and Wilson, 1982*). The standard deviation of each set of n reported values is given in the column labeled σ.

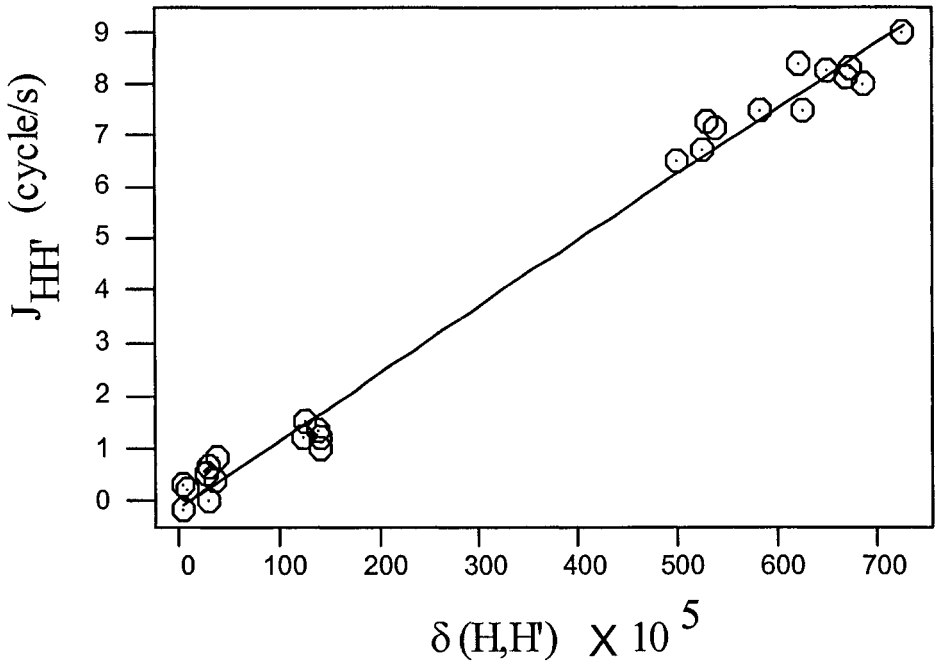


Figure 8.7 Linear correlation between δ(H,H') and $J_{HH'}$ for the cyclic hydrocarbons listed in table 8.4 and the fitted line according to the regression equation Eq. 8.15.

Table 8.5. $J_{HH'}$ coupling constants in cycle/s for some saturated molecules

	Atom label	δ(H,H') $\times 10^5$	$J_{HH'}$ Calc.	$J_{HH'}$ Exptl.	Reference
ethane (staggered)	av[(a)]	657	9.7	8[(b)]	(*Abraham et al., 1998*)
cyclohexane	a-a	1173	13.4	13.2	(*Abraham et al., 1998*)
cyclohexane	a-e	446	4.6	3.6	(*Abraham et al., 1998*)
cyclohexane	e-e	360	3.6	3.1	(*Abraham et al., 1998*)
cyclohexene	av[(c)]	304	2.9	3.1	(*Smith and Kriloff, 1963*)
1,3-cyclohexadiene	[(d)]	419	4.3	5.1	(*Gunther, 1980*)

(a) Average value taken over the two gauche vicinal hydrogen atoms (±60°) and the trans vicinal hydrogen (180°).
(b) Datum not used to obtain the regression equation.
(c) Average values for J(H,H') and for δ(H,H') taken between the hydrogen atom bonded to carbon C2 (IUPAC numbering) and the two non-equivalent vicinal hydrogen atoms bonded to the saturated carbon C3. Symmetry equivalents: C1 and C4 respectively.
(d) Values for vicinal hydrogen atoms separated by the single bond within the conjugated part of the molecule (i.e. C2 and C3, IUPAC numbering).

The correlation of the delocalization indices for nonbonded hydrogen atoms with their proton spin-spin coupling constants is an example of how the Fermi exchange density provides the vehicle for the transmission of information between the basins of nonbonded atoms. The transmitted effects have important consequences for reactivity as well, effects that provide further examples of exchange overriding internuclear separations. For example, there is a significantly greater delocalisation of the π-density of a carbon atom in benzene onto the *para* rather than the *meta* carbon (see scheme 8.2). The same pattern of delocalization is found for the corresponding atoms relative to the carbon bearing an electron releasing substituent such as NH_2 and for the same atoms relative to the nitrogen atom in pyridine. Thus the delocalization pattern of the Fermi correlation parallels the reactivity patterns of substituted benzene and related aromatic molecules. (*Bader et al., 1996; Matta and Hernández-Trujillo, 2002*)

The same pattern of alternating behaviour is displayed by the Laplacian of the electron density in substituted benzenes where the greatest charge concentrations defined by its topology are located *para* to the carbon bearing an electron releasing substituent and *meta* for an electron withdrawing one. (*Bader and Chang, 1989*) This association is not unexpected, as the topology of the Laplacian of the electron density provides a mapping of the essential pairing information from six- to three-dimensional space, as determined by the properties of the Fermi density. This statement follows from the form of the conditional same-spin pair density,[$\rho^{\alpha}(\mathbf{r}_2) + h^{\alpha}(\mathbf{r}_1,\mathbf{r}_2)$]: when the density of the Fermi hole is localized in the vicinity of $\mathbf{r}_1$, then the conditional pair density reduces to the spin density at points $\mathbf{r}_2$ removed from $\mathbf{r}_1$. Under these conditions, the Laplacian of the conditional pair density for a closed shell system, whose concentrations indicate where the density of the remaining electron pairs are to be found for a fixed position of a reference pair, will reduce to the Laplacian of the total density. Indeed, the two fields have been demonstrated to be structurally homeomorphic[1]. (*Bader and Heard,*

---

[1] A continuous map f: $\nabla^2\rho(\mathbf{r}_1,\mathbf{r}_2)\rightarrow\nabla^2\rho(\mathbf{r})$ is a homeomorphism if there exists another continuous map g: $\nabla^2\rho(\mathbf{r})\rightarrow\nabla^2\rho(\mathbf{r}_1,\mathbf{r}_2)$ such that $g(\nabla^2\rho(\mathbf{r}))=\nabla^2\rho(\mathbf{r}_1,\mathbf{r}_2)$ for all $\mathbf{r}_1$ and $\mathbf{r}_2\in\nabla^2\rho(\mathbf{r}_1,\mathbf{r}_2)$ and $f(\nabla^2\rho(\mathbf{r}_1,\mathbf{r}_2))=\nabla^2\rho(\mathbf{r}))$ for all $\mathbf{r}\in\nabla^2\rho(\mathbf{r})$. (*Nelson D. (Ed.), 1998*)

1999) Thus the local maxima in the field $L(\mathbf{r}) = -\nabla^2\rho(\mathbf{r})$, the local charge concentrations (CCs) of the Laplacian, denote regions of partial pair condensation, regions with greater than average probability of occupation by a single pair of electrons. The number, relative size and orientation of the CCs of $L(\mathbf{r})$ provide a remarkably faithful mapping of the localized bonded and nonbonded Lewis pairs assumed in the VSEPR model of molecular geometry. (*Gillespie, 1972*)

The CCs of the Laplacian denote regions of localization of the Fermi correlation, concentrations that correlate with the localization index $\lambda(A)$. Just as the total Fermi correlation sums to $-N$, it sums separately to $-N(A)$, the population of each atom given by Eq. 8.10, an expression detailing how the population of atom A that is not localized on A is spread out into the basins of each of the remaining atoms in the molecule. Thus both the localisation of electrons determined by the CCs of $L(\mathbf{r})$ and their delocalisation determined by the index $\delta(A,B)$ are manifestations of the density of the Fermi hole and grounded in the physics of the pair density.

8.7 Conclusions

Aromaticity has been shown to be closely related to the delocalization of electrons between the carbon atoms in a ring. A physical description of interatomic delocalization emerges when the Fermi hole density, a consequence of the Pauli exclusion principle, is integrated over bounded regions of space defined topologically according to QT-AIM. The aromatic sextet of Clar's empirical model has been recovered quantitatively as the difference between the total electron delocalization within each individual ring in the polycyclic aromatic molecule and the total electron delocalization within the ring of the corresponding saturated monocyclic molecule (in this case cyclohexane). Electron delocalization measured by the delocalization indices has been shown to correctly account for a host of experimental phenomena related to aromaticity, an agreement testifying to the physical foundations of this new theoretical treatment. Thus the description of aromaticity presented in this chapter has been shown to account for: (i) Experimental resonance enthalpies; (ii) Unsaturation patterns assigned on the

basis of the empirical Clar's sextet theory and which explain the chemistry and reactivity of polycyclic aromatic hydrocarbons; (iii) It has been shown that the higher stability of phenanthrene with respect to its isomer anthracene is caused by the appearance of a dihydrogen bond in phenanthrene, imparting enough stabilizing energy which exceeds the aromatic de-stabilization manifested in the carbon skeleton. This topological feature of the electron density appears to occur universally in the bay region of the angular isomers of the polycyclic aromatic hydrocarbons and warrants a separate future in-depth study; (iv) It has been shown how one can use the delocalization indices to accurately predict the behaviour of the proton-proton spin coupling constants, which do not fall monotonically with distance, and which are dominated by the Fermi contact term for the two nuclei. For nonbonded hydrogens, the Fermi contact term is primarily a consequence of 1*s*-like functions which dominate the basis set for the hydrogen atoms. As a result of the spherical symmetry of these functions, the changes in the density at the position of the proton are proportional to the changes in the density within the entire basin of a hydrogen atom. This proportionality enables one to make the identification expressed in Eq. 8.14, and which summarizes the physics behind the regression equations Eqs. 8.15 and 8.16.

APPENDIX 8.1: Listing of the AIMDELOC Program (QCPE 0802)

The following program, written by the author, computes the localization and delocalization (Eqs. 8.5 and 8.7) given the atomic overlap matrices. The program was used to generate these indices in the present thesis. The program consists of *ca.* 200 lines of code written as a UNIX bash shell script. (*Matta, 2001a*)

```
#!/bin/bash
clear
echo "      WARNING!"
echo "  File names: *tmp*, *temp*, *table*, *delete*, *trunc*, *.*.joint  will be lost in this directory."
echo "  If you wish to exit use 'Cntrl+c' key combination"
echo "       The integration files were generated using:"
echo "       ----------------------------------------"
echo "         [1]  PROAIMV - Version 94 - Revision B "
echo "         [2]  PAIM96 - Version 96, Revision C"
echo "       Choose ONE option  1  or  2 :"
read  option
# Select the version of proaim (different overlap matrix identifying text)
            case $option
              in
                1) string="Restricted Closed-Shell Wavefunction";;
                2) string="Restricted, SCF, closed-shell orbital wavefunction";;
                *) echo You entered an invalid number, please enter 1 or 2
                  exit 1;;
              esac
echo Ovelap matrix identifying string is:
echo "        $string"
echo
# Cleanup work place
  rm -f *table* *deleteme* *temp* *tmp*   *trunc* *.*.joint *fileslist* *j*   *prd* *fileinfo*   *fileslist*   MatrixIndeces
DelocIndeces
clear
echo  "     AIMDELOC  RUNNING  "
echo  "      Please wait"
echo
# Loop extracting the overlap matrix from the integration files of the atoms in the molecule (the output files of
# PROAIM). The end-result of this loop is a one-column sequential file listing the elements of the lower triangle of the
# overlap matrix.  One file is produced for each integration file. The output of this loop has the form atom.int.table.
            for i in $(ls *.int)
            do
                        # Get the overlap matrix of each atom from its integration file into a separate file (i.e. extract
                        # the overlap matrix portion from the .int files by removing all lines before and after it).
                            grep -A 20000 "$string" < $i > $i.truncated
                            wc $i.truncated > truncinfo
                            read x1 x2 x3 x4 < truncinfo
                            y=`expr $x1 - 2`
                            tail -$y $i.truncated > $i.temp
                            z=`expr $x1 - 14`
                            head -$z $i.temp > $i.temp2
                            rm -f $i.truncated
                            rm -f $i.temp
                            rm -f truncinfo
                        # Transform the lines of the overlap matrix into one column of numbers each (lower trianglular
                        # overlap matrix is converted into a one-colum list of numbers (atom.int.table)
                            mv $i.temp2 $i.OvrlpMtrx
                            cp $i.OvrlpMtrx $i.OM
                            wc $i.OvrlpMtrx > truncinfo
```

```
                read xx1 xx2 xx3 xx4 < truncinfo
              # xx1 = number of LINES (NOT ROWS) in the overlap matrix, the maximum number of entries
              # per line is 8 numbers: a1 ... a8, which are transposed in the following loop.
                xxx1=`expr $xx1 + 1`
                        j=1
                        until [ $j -eq $xxx1 ]
                          do
                                  echo "Append/transpose row $j of the overlap matrix of $i to the column"
                                   read a1 a2 a3 a4 a5 a6 a7 a8 < $i.OM
                                   echo  $a1 >> $i.table
                                   echo  $a2 >> $i.table
                                   echo  $a3 >> $i.table
                                   echo  $a4 >> $i.table
                                   echo  $a5 >> $i.table
                                   echo  $a6 >> $i.table
                                   echo  $a7 >> $i.table
                                   echo  $a8 >> $i.table
                                   tail -$(($xx1-$j))  $i.OM  > $i.temp.$j
                                   rm -f $i.OM
                                   mv $i.temp.$j $i.OM
                                   j=$(($j+1))
                             done
              # Remove empty lines from atom.int.table
                   sed -e '/^$/d'  < $i.table >  $i.temp
                   rm -f $i.table
                   mv $i.temp $i.table
                   echo
                   echo "     Column rep. of the matrix $i done!"
                   echo
               done
          echo "    First part done!"
          echo "    Overlap matrices converted into single column representation."
          echo "     AIMDELOC RUNNING"
          echo "      Please wait"
# Count the number of lines in the tabular form of the atomic overlap matrix
           wc *.table > deleteme
           read MatrixSize y z s < deleteme
           rm -f deleteme
# Create a file with serial numbering and matrix elements labels for the tabular form of the atomic overlap matrix.
 k=1
 i=1
           while test $i -gt 0
             do
                     j=0
                     ii=`expr $i - 1 `
                               while test $j -le $ii
                                 do
                                         if test $k -eq $(($MatrixSize+1))
                                           then break 2
                                         fi
                                         j=$((j+1))
                                         echo $k $i $j  >> MatrixIndeces
                                         k=$((k+1))
                                 done
                      i=$((i+1))
            done

          echo "    Overlap matrix indeces done!"
          echo "     AIMDELOC RUNNING"
          echo "      Please wait"
# number the lines of atom.int.table
           for k in $(ls *.int.table)
           do
                     cp $k $k.temp
```

```
                rm -f $k
                cat -n $k.temp > $k
        done
        rm -f *.int.temp*
        rm -f *.int.O*
# Prepare a list of atom.int.table files & counting them:  X1= the number of .int files
        ls *.int.table > fileslist
        wc fileslist > fileslist.info
        read X1 X2 X3 X4 < fileslist.info
        rm -f fileslist.info
        cp fileslist fileslist.tmp1
        cp fileslist fileslist.tmp11
# Pepare all possible permutations of atom.int.table files and multiplying the atomic overlap matrix elements of two
# atoms (diagonal elements are taken once and off diagonal elements are taken twice to account for the missing
# upper triangular part of the matrix, as proaim gives the lower triangle of this symmetric matrix).
I=0
while test $X1 -gt $I
  do
        tail -$(($X1-$I))  fileslist.tmp11  > fileslist.tmp22
        rm -f fileslist.tmp11
        mv fileslist.tmp22 fileslist.tmp11
        wc fileslist.tmp11 > fileslist.tmp11.info
        read XX1 XX2 XX3 XX4 < fileslist.tmp11.info
        rm -f fileslist.tmp11.info
        I=$(($I+1))
                k=0
                while test $XX1 -gt $k
                        do
                                read filenameA < fileslist.tmp11
                                tail -$(($XX1-$k))  fileslist.tmp1  > fileslist.tmp2
                                read filenameB < fileslist.tmp2
                                echo  $filenameA $filenameB >> jointname.$k.$I
                                echo "joining  $filenameA  and  $filenameB, cycle $I,$k"
                                sed -e 's/.int.table//g' <jointname.$k.$I> $I.$k.jointname
                                rm -f jointname.$k.$I
                                sed -e 's/ /__/g' <$I.$k.jointname> jointname.$k.$I
                                read jointfilename < jointname.$k.$I
                                rm -f jointname.$k.$I
                                join $filenameA $filenameB >>  $jointfilename.tmp
                                join MatrixIndeces $jointfilename.tmp > $jointfilename.jnt
                                rm -f *.tmpm  joint.$k.$I
                # Read the joint-file elements and testing if they are lower triangular matrix elements (in which
                # case they are multiplied by 2 to compensate for the missing upper triangular matrix elements)
                # or if they are diagonal elements in which case their product is outputed as such.
                                I=0
                                J=0
                                awk '{ I=I+1
                                    J=J+1
                                      if ( $2 == $3 ) w[I,J] = $4*$5
                                      else          w[I,J] = 2*$4*$5
                                    ++num
                                printf("%d %d %d %11.7f %11.7f %11.8f\n", $1, $2, $3, $4, $5, w[I,J])
                                    }'  $jointfilename.jnt  >  $jointfilename.prd
                                rm -f *.jnt
                # Summing the column of the products of the corresponding matrix elements belonging to two
                # atoms. (summing loop).  The sum is multiplied by two to account for two electrons (closed
                # shell), then by 2 to include F(A,B) AND F(B,A).  The delocalization index is given by:
                # delta(A,B) = 2|F(A,B)| + 2|F(B,A)|, where F(A,B) = F(B,A).
                                K=0
                                awk '    {
                                        K=K+1
                                        sum += 2*2*$6
                                        ++num
                                                }
```

```
                                                    END  {
                                                   printf("%s %11.8f\n", "The bond order is: ",  sum)
                                                  } ' $jointfilename.prd >> $jointfilename.prd
                                            k=$(($k+1))
                        done
    done
            grep "The bond order is: " *.prd > DelocIndeces.tmp
            sed -e 's/.prd//g' <DelocIndeces.tmp> DelocIndeces.tmp2
            sed -e 's/:/__/g' <DelocIndeces.tmp2> DelocIndeces.tmp3
                        awk ' BEGIN{  FS = "__" }
                        {
                          if ( $1 == $2 )
                            { printf ("%s %s %s %11.8f\n", $1, $2, "delta(A,A) = |F(A,A)|        = ", 0.5*$4 ) }
                            else printf ("%s %s %s %11.8f\n", $1, $2, "delta(A,B) = |F(A,B)+|F(B,A)| = ", $4 )
                         }'  DelocIndeces.tmp3   > DelocIndeces
# summing the deltas and lambdas over all the molecule, the sum = the total number of electrons in the molecules.
            K=0
            sum=0
                    awk  ' BEGIN {  FS = "=" }
                              {
                                sum += $3
                                   ++num
                                }
                                END  {
                                    printf("%s\n", " ------------------------------------")
                                     printf("%s %11.8f\n", "  SUM [ | F(AB) | ]   =   ", sum)
                                     printf("%s\n", " ------------------------------------")
                                     printf("%s\n", " Normal termination of AIMDELOC.")
                                             } ' DelocIndeces  >> DelocIndeces
 echo
 cat DelocIndeces
 echo
 rm -f *table*  *deleteme* *temp* *tmp* *trunc* *.*.joint  *fileslist*  *j*  *fileinfo* *fileslist*  MatrixIndeces
 mv DelocIndeces deloc.dlc
echo "      All atom-atom permutations done and bond orders computed"
echo "      YOUR RESULTS ARE SUMMARIZED IN THE FILE:  deloc.dlc "
echo "       AIMDELOC DONE"
echo "Normal termination of AIMDELOC on:"
date
```

Chapter 9

Long Bonds or Short Non-Bonds: A Case Study of the Bonding to Titanium

"Plus que je change, plus que je suis la même chose"
(The motto of a French speaking fluxional molecule)
F. Albert Cotton (*Cotton, 2002*)

9.1 Statement of the problem

The recent synthesis of a crystalline compound containing Ti bonded to cyclopentadienyl and a substituted dienyl fragment prompted the question as to whether Ti-C contacts that were found to be shorter than other such bonded contacts in the same molecule should be considered as short nonbonded contacts or 'nonclassical metal-to-saturated-carbon atom interactions,' fitting the description of agostic interactions. This question is one that has a unique answer within the framework of the quantum theory of atoms in molecules (QT-AIM). All of the classical bonding descriptors, when recast in terms of the topologies of the electron density and the pair density, are faithfully recovered when QT-AIM is applied to the hydrocarbon framework of the Ti complex, thereby justifying its application to the analysis of the Ti-C interactions. No bond paths are found to link the Ti to the carbons exhibiting the 'short contacts' and the topology of the density gives no indication of an incipient change in structure that would result in bond formation. This chapter is based on (*Bader and Matta, 2001a*).

9.2 Introduction: Assigning a Bonded Structure

At the Pacifichem 2000 meeting in Honolulu, a session on 'The boundary between long bonds and short non-bonds' considered whether or not a "nonbonded" contact found between a metal atom and a saturated carbon atom that is shorter than bonded carbon-metal separations in the same transition metal molecule, should be described as a bonded interaction. (*Tomaszewski et al., 1998*) If indeed bonded, the interaction would qualify as the first experimental example of an 'agostic' bond between a transition metal atom and a saturated carbon, a name previously applied only to metal-hydrogen interactions. (*Brookhart and Green M. L. H., 1983*) This chapter presents arguments against such a finding in this molecule, based upon the quantum theory of atoms in molecules (QT-AIM). The chapter illustrates how theory enables one to assign a molecular structure, defined by the network of bond paths, to any system for which the electron density is known from either experiment or from a theoretical calculation.

The presence of a bond path provides a universal indicator of bonding between the atoms that it links, one that is applicable to all types of atomic interactions. (*Bader, 1998b*) In particular, the presence of a bond path linking a titanium atom to a methyl hydrogen in an agostic interaction has been demonstrated in both experimental (*Scherer et al., 1998*) and theoretical charge distributions. (*Popelier and Logothetis, 1998*) The network of bond paths has been shown to recover the structures familiar to chemists assigned on the basis of the Lewis electron pair model applied to molecules composed of main group atoms. The Lewis model is however, not of universal applicability. In particular, it cannot be used to predict whether or not a short 'nonbonded' interaction is indicative of the presence of a bond. It is, for example, unable to account for the bonding found in crystals composed of neutral closed-shell molecules, that results from so-called "nonbonded interactions", interactions that are readily identified in terms of bond paths. This is exemplified by the bond paths present in both the experimental and theoretical charge distributions of solid chlorine that account for the short - relative to the sum of the van der Waals radii - directed intermolecular interactions

responsible for the layered structure exhibited by this crystal, a structure not anticipated when the 'nonbonded interactions' are described using a non-directional van der Waals potential. (*Tsirelson et al., 1995*)

The Lewis model provides the basis for the localized bonded and nonbonded electron pairs assumed in the VSEPR model of molecular geometry, (*Gillespie and Hargittai, 1991*) a model that can fail in its application to transition metal molecules, an indication that the Lewis model can also fail for such molecules. Thus to account for the bonding in a transition metal molecule and in particular to decide on the possible presence of an agostic interaction, it is necessary to understand the limitations of the Lewis model by appealing to the underlying physics.

9.3 The Lewis Model and the Spatial Pairing of Electrons

The spatial localization of an electron pair to a given atom and its delocalization over a pair of atoms are the physical embodiments of the Lewis concepts of a nonbonded and bonded pair respectively. The pairing of electrons is a consequence of the Pauli exclusion principle and the spatial localization of the pairing is determined by the corresponding property of the density of the Fermi hole. (*Bader and Stephens, 1975*) The use of the Fermi hole density in conjunction with QT-AIM, which leads to the definition of the localization ($\lambda(A)$) and delocalization indices ($\delta(A,B)$), is discussed in detail in the previous chapter, notably in section 8.4. We re-emphasize that because the Fermi correlation counts all of the electrons, the localization and delocalization indices sum to N and they provide a quantitative measure of how the N electrons in a molecule are localized within the individual atomic basins and delocalized between them. Eq. 8.9 determines how the electron population of atom A is delocalized over the remaining atoms in the molecule. (*Fradera et al., 1999*) Like the Lewis model, the indices provide a book-keeping of the electrons, but unlike the Lewis model, the tabulation is applicable to all systems at any level of theory. At the Hartree-Fock (HF) level the delocalization index equals unity for a single pair of electrons that is equally shared between two identical atoms A and A'. The electron

pairing predicted by the HF model of the pair density is found to be remarkably successful in recovering the Lewis model as discussed in section 8.4. The delocalization indices for homopolar interactions decrease with the addition of Coulomb correlation, as it disrupts the pairing of electrons between the atoms. For example, with the addition of Coulomb correlation, the value of δ(H,H') in the hydrogen molecule decreases from 1.00 to 0.85, in the nitrogen molecule δ(N,N') decreases from 3.0 to 2.2, and δ(C,C') in ethane decrease from 0.99 to 0.83. (*Fradera et al., 1999*) The values reported here are obtained at the HF level and they will represent upper bounds to the number of Lewis electron pairs shared between equivalent atoms.

One can argue that comparisons of the Lewis model with the pairing of electrons determined by theory should be restricted to the single determinant (HF) model of the wave function. It is Fermi correlation and only Fermi correlation that determines the spatial pairing of electrons and this is the sole source of electron correlation at the Hartree-Fock level. In a real sense, the Hartree-Fock model retrieves the Lewis model from any, more general description of a molecular system.

The Lewis model is evident in the charge concentrations (CCs) displayed by the Laplacian of the electron density in the valence shell of charge concentration (VSCC). (*Bader, 1990; Bader et al., 1984*) The density is concentrated in regions where the Laplacian is negative and thus the CCs are local maxima in the function $L(\mathbf{r}) = -\nabla^2\rho(\mathbf{r})$. These CCs have been shown to yield a faithful mapping of the localized electron domains that are assumed to be present in the valence shell of a central atom in the VSEPR model of molecular geometry. (*Bader et al., 1988; Bader et al., 1984*) There is agreement not only in the number of CCs found in the VSCC of the central atom with the number of bonded and nonbonded electron pairs assumed in the model, but also in their angular orientation and relative sizes.

The maxima in $L(\mathbf{r})$ are a reflection of the spatial localization of the electrons, as shown through a study of the properties of the conditional pair density. (*Bader and Heard, 1999*) This density distribution shows where α,β pairs of electrons are most likely to be found for a given position of a reference pair. The conditional pair density exhibits a very important property: *when the Fermi*

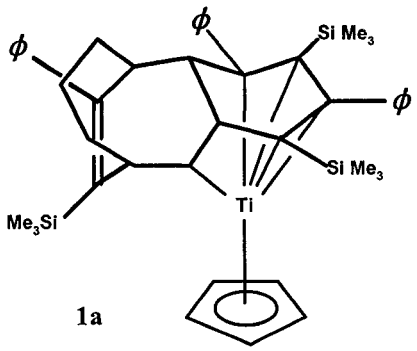
hole for the reference electron pair is maximally localized to a given region of space, the conditional pair density reduces to the electron density outside of the region of localization. Under this condition, the Laplacian of the conditional pair density reduces to the Laplacian of the electron density, and the CCs of $L(\mathbf{r})$ thus coincide with the local charge concentrations defined through the conditional pair density. The latter CCs indicate the positions where the density of the remaining electron pairs are most likely to be found for a fixed position of the reference pair. Thus the CCs displayed in $L(\mathbf{r})$ signify the regions of partial pair condensation, regions with greater than average probabilities of occupation by a single pair of electrons and $L(\mathbf{r})$ provides a mapping of the essential aspects of electron pairing determined in six-dimensional space, onto the real space of the density. (*Bader and Heard, 1999*)

9.4 Bonding to the Titanium Atom

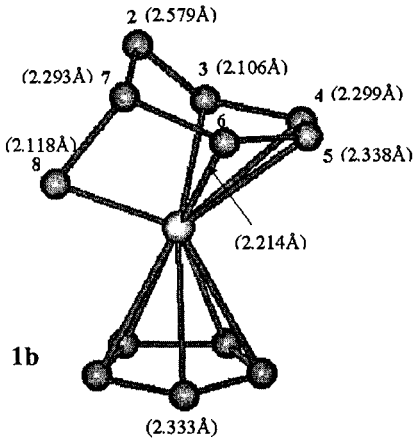
The x-ray crystallographic structure of the complex with the short 'nonbonded' Ti-C separations is reported in (*Tomaszewski et al., 1998*) and corresponds to the structure shown as **1a**. The ligands distant to the titanium atom are omitted in the model structure **1b**, that is used by the authors as well as in the present work for the purpose of discussing the bonding to the titanium atom. (See scheme 9.1) The model retains the cyclopentadienyl and cyclohexadienyl fragments, the latter having an attached methylenic carbon, that constitute the carbon framework containing the atoms that are or possibly can be bonded to the titanium atom. Their structure yields "an apparent 14-electron count" by indicating the presence of bonds from Ti to atoms C3, C4, C5 and C6 of the cyclohexadiene ring (the 6-MR), to the methylene carbon C8 and to the carbons of the cyclopendienyl ring (the 5-MR). The lengths of the assumed Ti-C bonds are indicated in **1b** and they are to be compared with the "remarkably short Ti-C(sp^3) contacts" Ti-C2 of 2.579 Å and Ti-C7 of 2.293 Å, the latter being less than the separations between Ti and C4 or C5 or any of the carbons of the 5-MR. If these short contacts with C2 and C7 are indicative of Ti-C bonding, that is, of agostic interactions, then the complex "could provide

important insight into the pathways by which C-C bonds may be selectively activated and functionalized by transition-metal centers." (*Tomaszewski et al., 1998*)

Scheme 9.1



1a



1b

9.5 Computational Methods

Wavefunctions for **1b** were obtained at the restricted Hartree-Fock (RHF) and at the BLYP-DFT levels of theory with the experimental geometry of **1a** being used for the atoms in common with those in **1b** followed by an energy optimization of the carbon-hydrogen separations using Gaussian94. (*Frisch et al., 1995*) The 6-31+G* basis set was used for the carbon and hydrogen atoms and a triple zeta valence basis set consisting of (14*s*11*p*6*d*) contracted to [10*s*8*p*3*d*] obtained from the GAMMESS program (*Schmidt et al., 1993*) was used for titanium. Both sets contain diffuse functions whose presence is required for the description of possible weak interactions between carbon and titanium. The topological structure and the atomic properties were determined using AIMPAC. (*Biegler-König et al., 1982*) Delocalization and localization indices were computed using AIMDELOC (*Matta, 2001a*), listed in appendix 8.1.

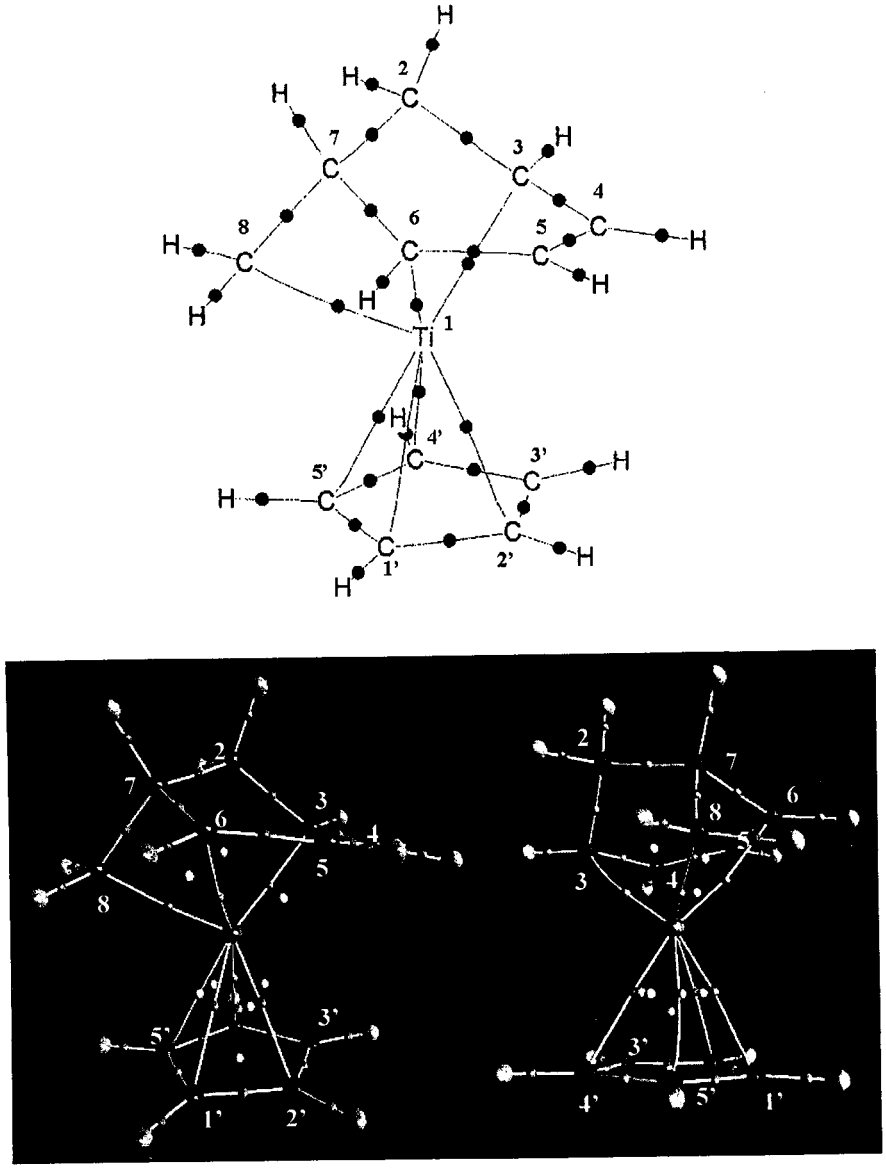


Figure 9.1 Three views of the molecular graph for **1b** determined by the topology of the DFT/BLYP electron density, together with the numbering scheme. A bond critical point (CP) is denoted by a dot in the top diagram and a red dot in the bottom plots. It denotes the intersection of the bond path with the interatomic surface. Note the curvature of the Ti-C bond paths at their termini with the carbons of the 5-MR. In the bottom plots, the positions of the ring CPs are denoted by yellow dots, and the cage CP is denoted by a green dot. The Ti-C bond and associated ring CPs of the 5-MR are linked by an almost constant ring of density signifying the presence of a bonded cone of density enclosing a cage critical point denoted by a green dot. The bond CP of the Ti-C bond path in the center foreground of the bottom-right plot is nearly superimposed on a neighbouring ring CP and the bond path is on the verge of annihilation. Elements colour code: H (light grey), C (dark grey), Ti (violet-blue).

9.6 Molecular Graph of the Titanium Complex

The molecular graph obtained from the BLYP-DFT density is shown in figure 9.1. An identical structure was obtained from the RHF density with the exception of one less bond path to the carbons of the 5-MR. Bond paths were found to reproduce the C-C and C-H interactions depicted in structure **1b** but unlike this structure, the Ti is bonded only to the methylenic carbon C8 and to the two terminal carbons of the diene fragment, C3 and C6, as well as to four of the carbons of the 5-MR. There were no bond paths linking Ti to the two central carbons of the diene, carbons C4 and C5, as pictured in **1b** and none linking Ti to the two saturated carbon atoms C2 and C7 that exhibit the short contacts. Thus the topology of the density does not support the presence of agostic interactions between Ti and C2 or C7. A total of 33 bond, 8 ring and a cage critical point were found, as required to satisfy the Poincaré-Hopf relationship (Eq. 2.5) for the molecule which contains 27 nuclei and there were no missing critical points.

9.7 The Hydrocarbon Framework Interactions

The values of the BLYP critical point (cp) properties are given in figure 9.2(a) and the RHF delocalization indices for the C-C and C-H interactions in figure 9.2(b). Electron correlation in general decreases the value of ρ_b, the density at a bond cp, (*Gatti et al., 1988*) and the RHF values of ρ_b thus exceed those obtained from BLYP by 2 to 3% for the C-C and C-H interactions and by less than 1% for the Ti-C interactions. The δ(C,H) values indicate that approximately one Lewis pair is shared between the carbon atoms and their bonded hydrogens. The δ(C,C) values between carbons where one or both are saturated also indicate an interaction resulting from the sharing of a single pair of electrons, while the values for the C-C interactions in the dienyl fragment and in the 5-MR indicate a sharing in excess of one electron pair. In an isolated *cis*-butadiene, the delocalization index between a terminal atom and its bonded neighbour equals 1.83, while that between the two interior atoms is slightly

greater than unity, 1.07, indicating only slight delocalization of the π- electrons across the 'single' bond. In the complex, the bonding of the terminal atoms C3 and C6 to Ti reduces the delocalization indices δ(C3,C4) and δ(C5,C6) from 1.83 in the free diene to 1.2 and 1.3 respectively, a result of C3 and C6 sharing density with the Ti. The two interior atoms C4 and C5 are not bonded to the Ti, as proposed in (*Tomaszewski et al., 1998*) and consequently δ(C4,C5) increases from 1.07 to 1.5 and they retain their unsaturated nature. The change in the delocalization indices for the dienyl carbons from the values found in the free diene are suggestive of a Diels-Alder type addition of the Ti, with the metal bonding restricted to the two terminal atoms and π- density being transferred to the two central carbons. The indices for non-neigbouring atoms indicate that the sharing of electrons between such atoms is negligibly small when one or both of the atoms are saturated carbons, but they are equally indicative of significant delocalization of electrons over the conjugated atoms of the 6- and 5-MRs.

The values of the bond CP properties are typical of nonpolar shared C-C and C-H interactions: ρ_b ranges from 0.21 to 0.28 au for the C-C interactions in the 6-MR and including C8, the values increase in parallel with the increase in δ(C,C), the number of shared electron pairs. The largest values, 0.31 au, are found for the 5-MR. The values of $\nabla^2\rho_b$ are all negative, the magnitude increasing with the value of δ(C,C). The local expression for the virial theorem for a molecule in an equilibrium geometry (in which $\mathcal{V} = V$) equates $(1/4)\nabla^2\rho_b$ to the sum of $2G_b$, the kinetic energy density, and V_b, the potential energy density (Eq. 2.60). Since $G_b > 0$ and $V_b < 0$, the finding that $\nabla^2\rho_b < 0$ demonstrates that the interactions are stabilized by the decrease in the potential energy that results from the contraction of the electron density towards the bond path and to its accumulation in the binding region; both effects increase with the value of δ(C,C). The energy density H_b (*Cremer and Kraka, 1984*) is given by the sum $G_b + V_b$ and its magnitude parallels the value of ρ_b for the 6-MR and 5-MR rings. The largest magnitudes of $\nabla^2\rho_b$ and H_b and of the potential energy density V_b, are found for the delocalized charge distribution of the 5-MR, a system associated with 'resonance' stabilization. These same atoms are more stable than C(2), the least stable carbon atom of the 6-MR, by an amount in

excess of 120 kcal mole^{-1}. The bond paths and their properties together with the electron delocalization data demonstrate that theory is able to both recover and make quantitative the classical descriptors of bonding for the hydrocarbon framework. The same theory is next applied to the interactions with the Ti atom where the classical models can fail.

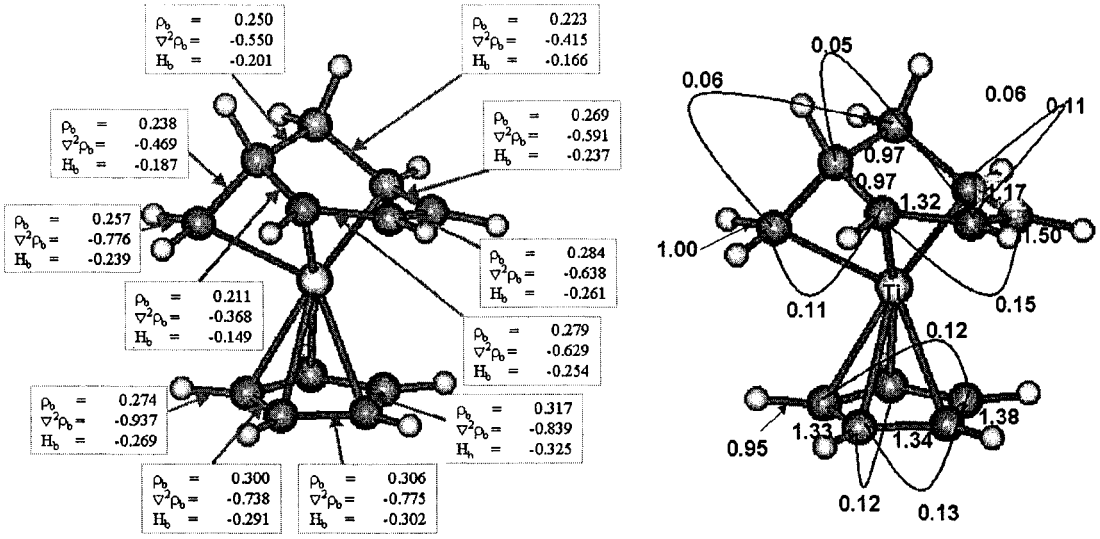


Figure 9.2 (a) A depiction of the molecular graph wherein the nuclei are denoted by spheres and the bond paths by lines to display the DFT/BLYP values of properties at the C-C and two typical C-H bond CPs of the hydrocarbon framework: the electron density ρ_b, the Laplacian of the density $\nabla^2\rho_b$ and the energy density H_b. The magnitudes of all three bond descriptors increase with the assumed classical bond order, the values for the central bond of the diene fragment, C4-C5, exceeding those for the two terminal bonds. **(b)** The Hartree-Fock values of the delocalization index δ(C,C') for neighbouring (bonded) and next nearest (NN) neighbouring carbon atoms, the NN pair of atoms being linked by a curved line. Note that there are significant NN neighbour delocalizations only between pairs of unsaturated atoms with exception of the value of 0.11 between the methylenic carbon C8 and C6 of the diene fragment.

9.8 Titanium-Carbon Interactions

The data summarizing the Ti-C interactions are given in figure 9.3. We first comment on the bonding between Ti and the carbons of the 5-MR. The bond paths are characterized by relatively small values of $\rho_b \sim 0.05$ au, values considerably less than those for the C-C and C-H interactions. The Ti-C bond paths to bonded pairs of carbon atoms, to C1-C2, to C4-C5 and to C5-C1 form 3-MRs and a ring CP is present in the face of each ring. Another ring is formed by the closed loop of interactions Ti-C2-C3-

C4-Ti. Each of these rings possesses a CP in the associated ring surface where the electron density attains its minimum value, a value denoted by ρ_r. There is also a ring critical point in the face of the 5-MR and together, these ring faces define a cage enclosing a cage critical point where the density attains a minimum value, figure 9.1(a) and (b). What is special about the rings formed with the Ti is that the values of ρ_r are only slightly less than the values of ρ_b for the Ti-C bond paths, the lowest values for the density in the perimeter of each ring. In addition, all of the bond and ring CPs are approximately equidistant from the Ti atom, 2.07 ± 0.02 Å, and as a result a ring of nearly constant density, ρ = 0.050 ± 0.003 au, encircles the bonded cage, as indicated in figure 9.1(b). Thus the bonding of the Ti to the carbons of the 5-MR is not well represented in terms of individual atomic interactions, but rather by a *bonded cone* of density, with the individual bond paths having values for the density only in slight excess of the density in the face of each ring that links one Ti-C interaction with the neighbouring one. This situation is pictured in figure 9.4 which shows the cone-shaped density envelope for a value equal to the average density of the bond and ring CPs lying in the bonded cone. Note that the cone extends out to all of the atoms of the cyclopentadienyl ring since the ring cps, including that of the 4-MR, are all approximately equidistant from the axis of the ring of carbon atoms. Clearly the interaction of a metal atom with the carbons of a cyclopentadienyl ring is best viewed as involving the delocalized density of the entire ring perimeter, a picture that is conceptually similar to that used to denote a metal-ring interaction in a conventional structure diagram, **1a**. The topological description provides a real space representation of Moffitt's view of the bonding in cyclopentadienyl molecules in terms of 'cup-shaped lobes' formed from d orbitals on the metal overlapping with the ring of π-orbitals on the 5-MRs. (*Moffitt, 1954*) The result is an enhanced binding over what one would anticipate based on the individual ρ_b values. This view of the bonding to the 5-MR is consistent

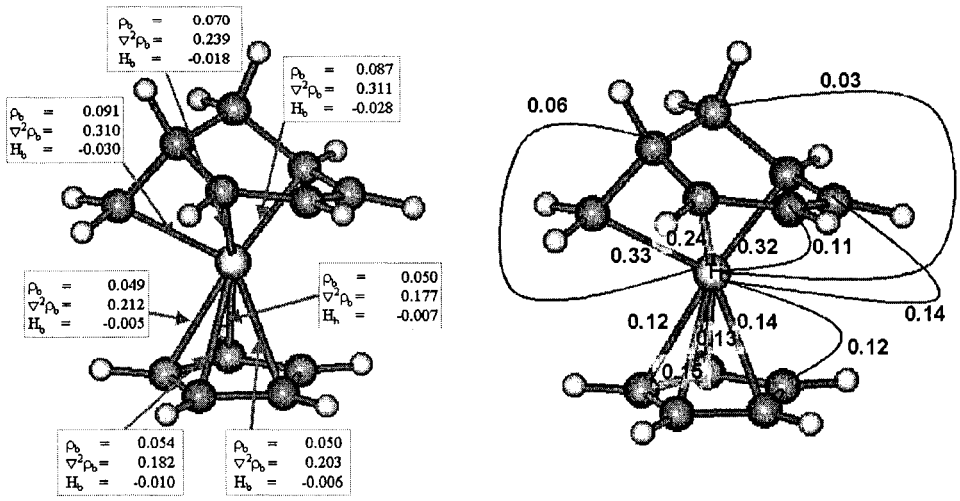


Figure 9.3 (a) DFT/BLYP values of the Ti-C bond critical point data. Note the parallel increase in the magnitude of the energy density H_b with ρ_b. **(b)** Hartree-Fock delocalization indices for the Ti-C bonded interactions and for the delocalization between Ti and the carbons to which it is not bonded. The values of δ(Ti-C) for the 'close contact' carbons C2 and C7 are no larger than those found for the delocalizations between next nearest carbon atoms where one or both are saturated (figure 9.2(b)). The electron density is nearly uniformly delocalized between Ti and the carbons of the 5-MR that are linked by the bonded cone of density.

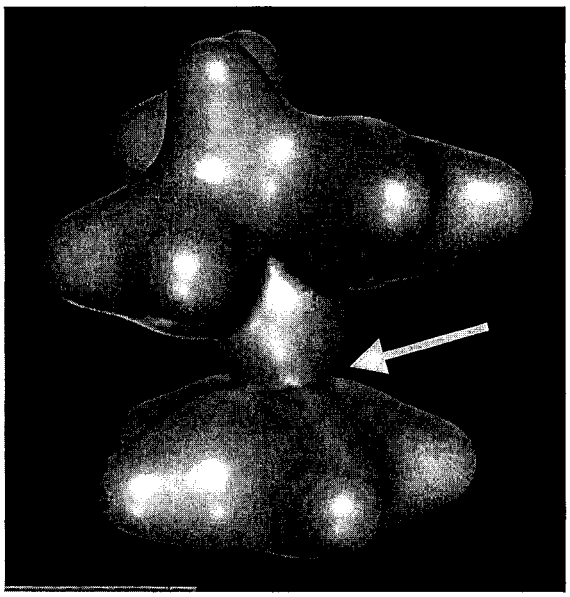


Figure 9.4 A display of an envelope of the electron density for a value equal to the average of the bond and ring CP values that form the ring of density in the surface of the bonded cone. These CPs lie in the interatomic surfaces that separate the Ti atom from the carbon atoms of the 5-MR ring and the ring of density they define, denoted by an arrow in the figure, lies at the union of the Ti atom density with that of the 5-MR. The ring represents the minimum value in the cone of density that links the Ti nucleus to the carbon nuclei of the 5-MR (figure 9.1).

with the observation that the barrier to free rotation of a cyclopentadienyl ring in metal complexes is quite low. (See review by (*Cotton, 2002*)). Moffitt shows that a single electron pair is involved in the bonding of a metal atom to a cyclopentadienyl (Cp) ring and the interaction is formally electron deficient. Thus it is unsurprising that the bonding topology found in Ti-Cp is the same as that found for the boranes and carboranes. (*Bader and Legare, 1992*) The stability of these electron deficient molecules has been shown to be a consequence of the delocalization of the density over the surfaces of 3- and 4-MRs of bonded boron and/or carbon atoms, a picture in agreement with the description of these systems as globally delocalized by King and Rouvray. (*King and Rouvray, 1977*)

Since the density is almost flat along a line linking a ring CP to the two neighbouring Ti-C bond CPs, the associated curvatures of the density at all three CPs approach zero and the Ti-C bond CPs exhibit large ellipticities. That is, the negative curvature of ρ at a bond CP tangent to the ring of CPs is much smaller in magnitude than the one normal to it and to the bond path. Only a small amount of vibrational energy is required to perturb the form of the density in the region where its curvatures are vanishingly small. (*Bader, 1990; Bader and Laidig, 1992*) In the present case, the perturbation causes a ring CP to coalesce with one of the neighbouring bond CPs, resulting in the formation of a singularity in the density, a CP where both the first and second derivatives of the density vanish along the line of approach of the two CPs. Such a CP is unstable and any further nuclear motion results in its annihilation and in the rupture of the associated Ti-C bond path. It is clear from figure 9.1(b) that the central Ti-C bond path of the cage is on the verge of annihilation, as its bond CP and the neighbouring ring CP are nearly superimposed, being separated by only 0.054 au. In turn, their densities differ by only 1×10^{-6} au and their respective curvatures are –0.001 and + 0.002 au. It is this bond path that is missing in the Hartree-Fock density. The density in the face of the 4-MR in the bonded cone is also flat and a motion of the nuclei can result in the creation of a new singularity that bifurcates into a new bond and a new ring CP. The magnitudes of the associated curvatures of the bond and ring CPs are 0.008 au, compared to a value of 0.26 au for the curvature along the Ti-C bond path. These two structural changes are not independent and low energy ring puckering motions will constantly cause

the annihilation and creation of Ti-C bond paths, leaving the bonded cone of density intact at all times. In $Ti(C_5H_5)^+$, Ti is symmetrically placed with respect to the carbons of the 5-MR and is linked by bond paths with large ellipticities to all five carbon atoms. The Ti-C bond and the 3-MR ring CPs yield a ring of density that varies by only 0.0006 au between its maximum and minimum values. The bond CPs exhibit large ellipticities, with a curvature of small magnitude tangent to the ring of CPs. The curvatures of the density at the bond and ring CPs alternate between –0.006 au at the bond CPs and +0.006 au at the ring CPs, the topological requirements for the continuous making and breaking of Ti-C bond paths by low amplitude ring puckering motions. Large ellipticites are also found for the agostic interactions of Ti with H in $C_2H_5TiCl_3$(dmpe), (*Scherer et al., 1998*) and in $C_3H_7TiCl_2^+$, (*Popelier and Logothetis, 1998*) indicative of their topological instability. The formation of the agostic bond results in the formation of a ring structure and it is broken by the coalescence of the bond CP with the associated ring CP that occurs with a rotation about a C-C axis.

While the three bonded interactions of the Ti with carbons of the remaining framework exhibit larger values for ρ_b than found for the 5-MR, they all exhibit similar characteristics: $\rho_b < 0.1$ au, $\nabla^2\rho_b > 0$ and $H_b < 0$. The CP data are similar to those obtained in theoretical calculations of Ti-C interactions in $CH_3TiCl_2^+$, (*Popelier and Logothetis, 1998*) and $Ti(CH_3)_2Cl_2$. (*Bader et al., 1998*) The Ti-C bond lengths in these molecules, 1.97 to 2.03 Å respectively, are shorter than those in the present molecule where the shortest distance is 2.12 Å and the ρ_b values of ~ 0.12 au are somewhat larger. The experimental value of $\rho_b = 0.082$ au determined for Ti-C in $TiCl_3$(dmpe) with the Ti-C separation equal to 2.15 Å (*Scherer et al., 1998*) is in good accord with the value reported here for Ti-C8 with $\rho_b = 0.091$ au for a bond length of 2.12 Å. The longer bond lengths and smaller ρ_b values found for the molecules where the Ti is bonded to a large complex organic ligand indicate weaker Ti-C interactions than are predicted for bonds to the carbons of a methyl group.

The longer Ti-C bond lengths found in **1b** compared to a Ti bonded to a carbon of a methyl group, result in values of $\nabla^2\rho_b$ that are approximately ten times larger, ranging from 0.2 to 0.3 au.

Small values of ρ_b and positive values of $\nabla^2\rho_b$ for bonds between main group atoms are usually found for closed-shell interactions, which includes ionic interactions between ions that approach closed-shell electronic structures. In these cases the kinetic energy density at the bond CP so outweighs the potential energy density, that the total energy density H_b, as well as $\nabla^2\rho_b$ are both positive. In the case of interactions between transition metal atoms or between a transition metal atom and a ligand however, H_b appears to be invariably negative. Thus bonding to a transition metal defines a new set of bond CP characteristics that represent a mix of the closed-shell and shared parameters, with ρ_b small and $\nabla^2\rho_b > 0$, as found for closed-shell interaction, but with $H_b < 0$, as found for a shared interaction. (*Frenking and Pidun, 1997; Macchi et al., 1998*) The magnitude of H_b parallels the increase in the values of ρ_b and $\nabla^2\rho_b$ (figure 9.3). The positive values of $\nabla^2\rho_b$ are discussed further below.

There is considerable charge transfer from Ti to the ligands, as indicated by the atomic charges displayed in figure 9.5(a), with Ti losing approximately three electronic charges. The charge on the (C_5H_5) group is –0.8 e. Hydrogen transfers electronic charge to a unsaturated carbon and the net charge on the cyclopentadienyl carbons is –1.1 e. The increasing amount of electronic charge transferred to each of the three other carbons bonded to Ti, C3, C6 and C8, parallels the increase in their associated ρ_b value. The other carbons bearing significant negative charges are the remaining members of the dienyl goup, C4 and C5 which also withdraw electronic charge from their bonded hydrogens. The electronegativity of a carbon atom decreases with decreasing *s* character and carbons in saturated hydrocarbons bear small positive charges. In 1b, the two saturated carbons C2 and C7 bear small negative charges as a result of charge transfer from Ti.

Figure 9.5(b) gives the energies of both sets of carbon atoms relative to the least stable atom in each set. In each case this is a carbon not bonded to Ti. The most stable atoms are those bonded to the Ti and the degree of stabilization increases in parallel with the value of ρ_b and with the magnitude of the negative charge on carbon within each set.

The delocalization indices δ(Ti,C) for the bonded interactions also increase along with the values of ρ_b (figure 9.3(b)), but all are considerably smaller than the values of δ(C,C) found for the hydrocarbon framework. The value of the delocalization index equals the number of shared Lewis pairs only for homopolar bonding. In the presence of charge transfer, as found for the Ti-C interactions, the electron pair is not equally shared, but partly localized on the more electronegative atom, that is, λ(C) increases at the expense of δ(Ti,C). Quite independent of the Lewis model, δ(Ti,C) determines the extent to which electrons are exchanged between Ti and a carbon atom, and this value is greatest for the atoms bonded to Ti within each set. It is least for atoms C2 and C7, their values of δ(Ti,C) = 0.03 and 0.06, being no greater than those determining the delocalization of their electrons into the basins of carbon atoms to which they are not bonded, values of 0.06 or smaller. These are to be compared to values of δ(Ti,C) ranging from 0.2 to 0.3, for the carbon atoms bonded to Ti within the same set. The only saturated carbon with a significant delocalization into the basin of an atom it is not bonded to is C8 and its delocalization is into the basin of C6, both of which are bonded to Ti.

The atomic population of Ti equals 19 electrons (and hence bears a charge of +3e) of which approximately 18 are localized within its own basin meaning that approximately one electronic charge is delocalized into the basins of other atoms. The number of localized electrons (λ(Ti) = 17.9 for both RHF and BLYP) corresponds to a nearly complete localization of the density of the inner quantum shells (2,8,8), and the net charge of +3 to the loss of all but one of its valence electrons. Of the one electronic charge that is delocalized, 0.7 are shared with the carbons to which it is bonded, 0.4 to C8 and C3 and C4 and 0.3 with the carbons of the 5-MR. Another 0.1 are shared with the unsaturated atoms, C4 and C5, and the remaining 0.2 are delocalized over all of the other atoms in the molecule.

9.9 The Laplacian and Bonding to Titanium

A carbon atom exhibits two quantum shells in $L(\mathbf{r}) = -\nabla^2\rho$, as shown in the contour maps of $L(\mathbf{r})$ in figure 9.6. In accordance with the Lewis and VSEPR models, the VSCC of each saturated carbon

atom exhibits four bonded charge concentrations in a tetrahedral arrangement and the VSCC of each of the unsaturated carbons exhibits three such concentrations in a trigonal arrangement. These local maxima in the VSCC of an atom are located and identified, along with the other possible CPs in $L(\mathbf{r})$,

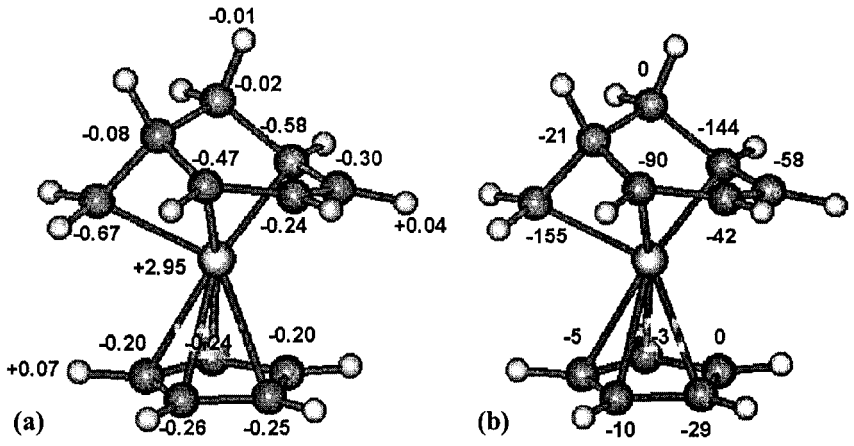


Figure 9.5 (a) The DFT/BLYP atomic charges for **1b**. **(b)** The DFT/BLYP energies in kcal/mol of the carbon atoms in the cyclohexadienyl fragment relative to the energy of C2 and of the carbons in the 5-MR relative to the energy of the carbon not bonded to Ti. The Hartree-Fock values are nearly identical and yield the same trends. The carbons of the 5-MR are more stable by 123 kcal/mol and the most stable carbon atoms in each set are those that are bonded to the Ti atom..

using AIMPAC. (*Biegler-König et al., 1982*) Each carbon bonded to Ti, that is, linked to it by a bond path, has a CC directed at the Ti core. Spatial displays of the geometric patterns of CCs are illustrated in figure 9.6 and by the zero-envelope of the Laplacian of the density in figure 9.7. The zero-envelope defines the outer boundary of the VSCC on the carbons and hydrogens and the outer core of Ti. Included in the group of saturated carbons with four bonded CCs are C3 and C6, the terminal carbons of the dienyl system that are bonded to Ti. The inner atoms of the dienyl group, C4 and C5, exhibit only three CCs as found for a sp^2 carbon. The bonded CC on each of the atoms C3, C6 and C8, that is directed at the Ti core is clearly evident in figures 9.6 and 9.7. There is no CC on C2 nor on C7 that is directed at Ti. Neither is there a suggestion of the formation of a fifth CC on C2 or C7 directed at Ti. The carbons of aromatic ring systems such as benzene and cyclopentadienyl display a set of secondary

CCs on each carbon, on each side of the ring system. (*Bader and Chang, 1989*) These are evident in figure 9.6 and in the zero-envelope, figure 9.7, where they appear directed at the Ti atom.

The Laplacian of a transition metal atom does not possess a valence shell charge concentration, whether free or bound. This is not an artifact of the shell structure defined by the Laplacian, but is determined by the conditional pair density. (*Bader and Heard, 1999*) The Laplacian of the conditional pair density, which is homeomorphic with the Laplacian of the electron density for a localized reference pair, demonstrates that there is no separate localization of electrons beyond the shell of charge concentration found for the (n-1) shell of the core for a transition metal atom in row n of the periodic table. This is evident in the contour plots of the Laplacian of the density given in figure 9.6. In addition to the spike-like charge concentration at the position of the Ti nucleus, there are two essentially spherical shells of charge concentration, indicating the presence of three quantum shells for the Ti atom, the zero-envelope surface of the third shell appearing in figure 9.7. Because of the lack of a VSCC, the interatomic surfaces of the bonded neighbours all border on the outer shell of charge depletion of a transition metal atom, a feature that accounts for the ubiquitous appearance of positive $\nabla^2\rho_b$ values for bonding to a transition metal atom.

The Laplacian of a transition metal differs from that of a main group atom in a second fundamental way: no bonded charge concentrations are found within its outer shell of charge concentration. Instead the maxima associated with the ligands that are present, are opposed to, rather than adjacent to, the position of the ligands and are labeled ligand opposed CCs (LOCCs). This important topological feature of $L(\mathbf{r})$ is found both theoretically (*Bader et al., 1996; Gillespie et al., 1998; Gillespie et al., 1996*) and experimentally (see for example the review (*Spackman, 1998*)). It is important to note that the LOCCs are formed from the remaining *ns* and (n-1)*d* electrons on the transition metal atom whose density is concentrated in the same shell of charge concentration as are the (n–1)*s* and *p* electrons of the core. (*Bader et al., 1998*) In the present case with only one valence electron remaining on Ti and that strongly delocalized onto the ligands, the local maxima in L(**r**) representing the LOCCs are nearly indistinguishable from the neighbouring minima and the outer shell

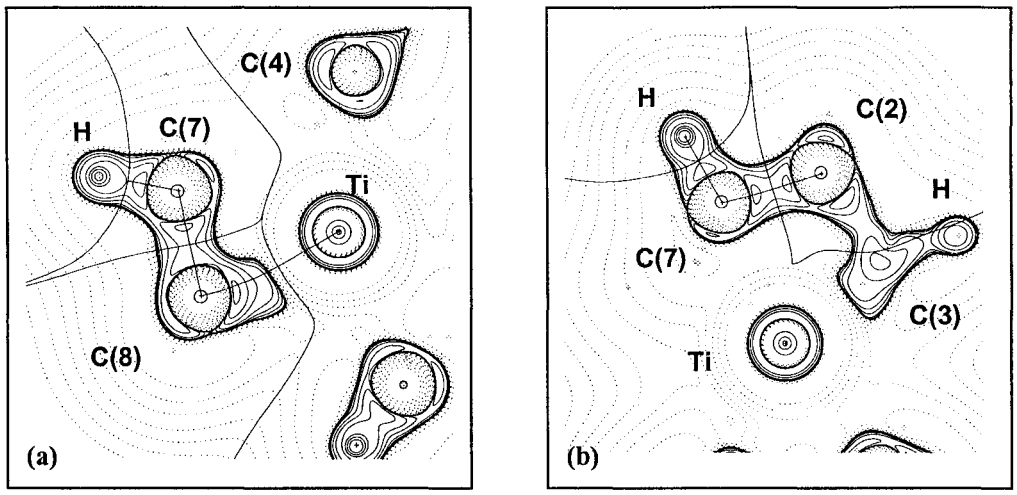


Figure 9.6 Contour maps of the Laplacian of the electron density with solid contours denoting regions of charge concentration where $\nabla^2\rho < 0$. Bond paths in the plane are indicated as are projections of interatomic surfaces onto the plane. **(a)** Plane containing the Ti, C8 and C7 nuclei and slightly above the nucleus of a carbon of the 5-MR. C8 exhibits four CCs characteristic of a saturated carbon: two bonded CCs directed at hydrogens above and below this plane and evident in figure 9.7, a bonded CC to C7 and a very pronounced bonded CC directed at the Ti atom. **(b)** Plane containing the Ti, C7 and C2 nuclei and showing as well, the pronounced bonded CC from C3 directed at the Ti atom. Both C7 and C2 exhibit four bonded CCs. C2 for example, possesses a bonded CC to C7, two to out-of-plane hydrogens and another to C3. Neither C2 nor C7 exhibit a bonded CC directed at the Ti atom. The contours in the nonbonded regions of the carbon atoms denote (3,-1) CPs that link the bonded CCs.

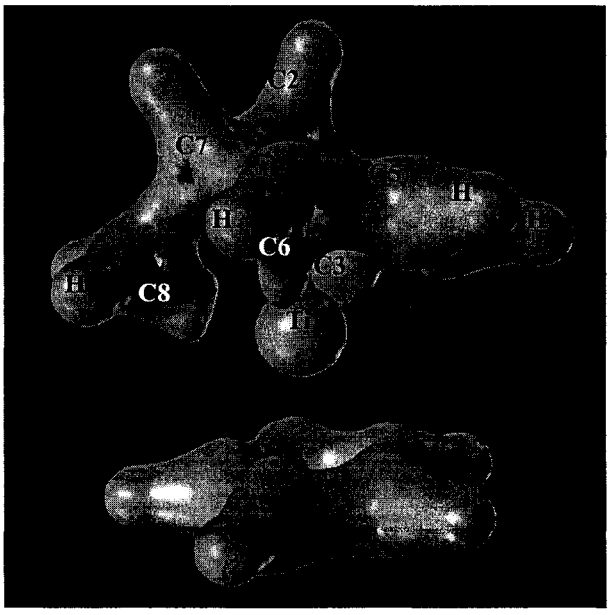


Figure 9.7 The zero envelope map of the Laplacian of the electron density that encloses the regions of charge concentration. The extension protruding from C8 is the bonded CC directed at the Ti core . Similar extensions are found protruding from C6 and C3. The secondary charge concentrations directed at the Ti from the carbons of the 5-MR are also evident.

of Ti appears spherical (figure 9.7). Ti does display LOCCs when the charge on Ti is +2 or less, corresponding to the presence of two or more valence electrons. (*Bader et al., 1998*) Thus CCs opposed to the carbons of the 5-MR are displayed within the outer core of Ti in $Ti(C_5H_5)^+$ where the charge on Ti is approximately + 2 e. While the Lewis model of bonding is recovered in the pattern of bonded and nonbonded CCs of $L(r)$ for main group atoms, this is clearly not the case for the topology of $L(r)$ displayed by a bound transition metal atom and the classical models of bonding must be abandoned. As illustrated in this chapter, one can use model-independent properties determined by the electron density and the pair density that have been shown to successfully account for the bonding in main group elements, to aid in the carry-over of the traditional models to obtain an understanding of transition metal bonding.

9.10 Conclusions

QT-AIM gives physical expression to the chemical concepts of bonding, structure and electron delocalization and it recovers the classical descriptors of bonding within the hydrocarbon framework of the complex. The use of this theory in the description of the bonding between this framework and the Ti atom is thus justified. The electron density, the Laplacian of the density and the conditional pair density demonstrate that Ti is bonded to the methylenic carbon C8, to the terminal atoms C3 and C6 of the dienyl fragment in the 6-MR and to the carbons of the 5-MR in a fluxional manner. No bonded interactions are present between Ti and the inner atoms C4 and C5 of the dienyl fragment and the interaction of Ti with the diene fragment corresponds to a Diels-Alder addition. Nor are there bonded interactions of Ti with atoms C2 and C7 that exhibit 'short nonbonded' Ti-C separations and they retain the characteristics of tetrahedrally bonded saturated carbon atoms. The atomic charges and energies also clearly differentiate between the carbons that are bonded to Ti and those that are not.

The electron density distribution does not give any indication of incipient structural change resulting from the formation of a singularity in ρ in the region between Ti and C2 or C7 or between Ti

and C4 or C5. Instead, the density exhibits steep troughs between the basins of these pairs of atoms with correspondingly large curvatures along the line linking their nuclei. For example, ρ attains a minimum value of 0.047 au, 2.0 au from Ti and 2.3 au from C7 on the line linking their nuclei. Diagonalization of the Hessian of the density at this point yields one negative and two positive curvatures and the largest of the two positive curvatures, equal to 0.268 au, is directed along the Ti-C7 axis to within 1.5°. The value of this curvature is hundreds of times larger than those recorded above as portending structural changes of the Ti with the carbons of the 5-MR and its value precludes any possibility of structural change for motions confined to the bonded wells on the potential energy surface. A DFT optimization of the geometry in which only the dihedral angles were held fixed at the experimental values resulted in both of the 'close contacts' to C2 and C7 increasing by 0.04 Å. The molecule is not on the verge of forming bond paths from Ti to either of these atoms. A bond path, again with a significant ellipticity, denoting an agostic interaction of Ti with the terminal saturated carbon atom is found in $C_3H_7TiCl_2^+$ at HF, BLYP and MP2 levels of theory, (*Popelier and Logothetis, 1998*) indicating that such an interaction is possible and is identified by theory when present.

Tomaszewski *et al* (*Tomaszewski et al., 1998*) cite "remarkably low" ^{13}C coupling constants for pairs of atoms involving C2 or C7 as evidence of an agostic interaction between Ti and these saturated carbons. The coupling constant for C2-C7 itself is normal and only those couplings between pairs of atoms that involve one atom linked to the Ti by a bond path exhibit low values. The values increase in the order C6-C7, C3-C2 and C8-C7, an order that parallels the decrease in bond length, including that for C2-C7 which possesses the shortest bond length and largest coupling constant. The principal contribution to a coupling constant is from the Fermi contact term resulting from the electron spin density induced at one nucleus by the nuclear moment of the other. Such spin information is transmitted by the delocalization of the spin density as described by the individual spin contributions to the delocalization index. (*Bader et al., 1996*) Only those coupling constants that involve an atom that has significant delocalization onto the Ti atom, that is, those linked to Ti by a bond path, have 'low' values.

Tomaszewski *et al* (*Tomaszewski et al., 1998*) employ natural atomic and bond orbital analyses to describe the changes that result from the addition of $Ti(C_5H_5)^+$ to a 5-methylene-1,3-cyclohexadiene anion. The changes are described in terms of orbital overlaps, in the occupancy of C-H and C-C bond orbitals and in terms of overlap weighted bond orders. Without criticizing the natural orbital analysis in particular, it is a feature of the orbital model that any chosen representation is but one of many obtainable by a unitary transformation of some initial set, all of which leave the physics and chemistry of the system unchanged. Relating an observed effect to the properties of individual orbitals leaves open the possibility of compensating changes occurring in other members of the set. The use of orbital contributions to interpret magnetic properties for example, is without physical meaning, as the orbital contributions to the total induced current do not individually satisfy the condition of being divergence free and they thus correspond to the local creation or destruction of electronic charge. Unlike atoms defined as quantum open systems, orbitals or orbital-derived quantities, however they are defined, are not confined to the atoms or bonds they are associated with when used to define a population or a bond order.

Consider for example, the 'significant overlap weighted NAO bond order' of 0.168 they find for Ti-C7. (*Tomaszewski et al., 1998*) The delocalization of electrons, a property underlying the notion of bond order, is uniquely determined by the pair density and is independent of the orbital representation used to determine it. (*Bader and Stephens, 1975*) Indeed, all pairs of orbital products contribute to both $\lambda(\Omega)$ and $\delta(\Omega,\Omega')$ at all levels of theory. Using the physical definition of 'bond order' one finds the delocalization of electrons between Ti and C7 to be minimal, equal to that found between next nearest neighbour saturated carbon atoms (figure 9.3(b)). (*Fradera et al., 1999*)

PART V:

GENERAL CONCLUSIONS AND REFERENCES

Chapter 10

General Conclusions and Closing Remarks

> " There things must be left. As has been said at various stages of this account, scientific interpretations express the unknown in terms of the known, and the known into which the translation is made becomes progressively more sophisticated. There is no particular reason for supposing that this process is in any way completed or that the last word has been said on any of the fundamental questions. Nor can scientific enquiry predict its own future: that is perhaps the greatest fascination of its adventure. "
>
> C. N. Hinshelwood (*Hinshelwood, 1951*), p.471.

It was impossible to have predicted the exact path of the research and anticipate the results presented in this thesis at the time the author embarked on this endeavor. It is equally impossible to predict the directions or developments that can take place in the future from where we stand now. This unpredictability is in fact the point of the exercise. All that can be offered to the reader is, perhaps, an assessment of what has been achieved in this thesis as it compares to what the author would like to achieve from his vantage point at the present. This assessment is by its nature subjective and it would be very hard, if not impossible, to free it from the biases and interests of the author.

A central idea in this thesis is that a molecule can be understood in terms of the atoms and functional groups that it is composed of. Atoms and functional groups, when determined as proper open quantum systems, exhibit two important properties: additivity (the sum of the atomic properties equals the corresponding molecular properties) and maximal transferability (the atom or group retains characteristic properties on going from one molecule to another). The study was undertaken within the framework of a coherent theory that accounts for chemical structure, reactivity, and measurable

physical properties of molecules: the quantum theory of atoms in molecules (QT-AIM). In doing so, new insight has been shed on a number of important chemical and biochemical problems, which were organized into the three parts of this thesis reviewing the author's research (parts II-IV).

Effort was made to relate theory to experiment troughout the work. On some occasions a direct causal connection between experimentally accessible quantites and theoretically calculated quantities could be established as in the case of the NMR spin-spin coupling constants and the delocalization indices. On other occasions, physical reasoning and intuition were used to develop empirical statistical models between experiment and theory as in the case of the partial molar volumes at infinite dilution and the calculated van der Waals volume and the charge separation indices. In both situations, the usefulness the theory as a predictor of experimental outcome has been made evident.

Part II established the high degree of transferability of geometric, bond, and atomic properties of an amino acid side-chain between conformations, an essential property if one is to use the electric multipoles - for example - in modeling of the force field. The transferability of the geometrical parameters of the side chains in the free amino acids to ones in tripeptides provides further evidence of the usefulness and validity of this transferability in the amino acid family. In the last chapter of this part, the transferable group properties were shown to provide a basis for the construction of excellent quantitative structure-to-activity relationship (QSAR) models, and the calculated group contributions were found to parallel the corresponding experimental group additivity schemes. Even a classification of the genetic code based on the calculated polarity measures of the amino acid side-chain has been shown to be possible. The encouraging success with which QT-AIM atomic properties have been used here in simple QSAR models (univariate regression models) capable of predicting physicochemical properties accurately leads one to optimism for their successful use in drug design. (*Bader et al.,Alber and Carloni, 2002*) The pharmacological properties of a series of molecules, though still dependent on their electron densities, are generally more difficult to predict than physicochemical properties. Predicting pharmacological properties often requires robust multivariate statistical regression models, (*Mager, 1984*) models that can be built using QT-AIM properties.

The transferability and additivity of atomic properties has been exploited in chapter 6 to obtain the properties of large computationally inaccessible molecules by assembling them from smaller readily accessible molecules. The chapter demonstrates that accurate (*ab initio* quality) properties of any organic molecule, no matter how large, and no matter how complex (not only simple repeating units) can be obtained by the buffered fragments method. The results reported in chapter 6 may have opened the door for the calculations of the properties of a multitude of large biologically important molecules which were out of reach to this point. The buffered fragment approach presented here is an initial effort which leaves much room for improvement. For example, a systematic way to break the large molecule into fragment molecules would be needed if the procedure is to be automated.

The properties of a crystal or an extended system have also been shown to be understandable in terms of the properties of the atoms or unit cells composing it. A particular example was chosen to emphasize how this idea of "additivity" applies to the dipolar polarization of a crystal or an extended system. This real-space approach is in contrast to what was previously thought: that macroscopic polarization can only be understood by invoking a quantum Berry phase in parameter space. While lacking the abstract elegance of the Berry phase approach, the real-space description of dipolar polarizations acquires an elegance of its own: the elegance of simplicity. The assembly of atomic dipolar (and higher multipolar) polarizations in buffered fragments to obtain the molecular multipoles of an assembled molecule should be straightforward and can be another extension for the buffered fragments method.

Up to this point in the thesis, the discussions were based primarily on the density (the one-electron density) and properties derived from it. In the last part of the work (part IV) it has been shown that one does not need to be confined to the one-electron density-derived properties to understand a system in terms of its parts. For instance, aromaticity is intimately linked to electron delocalization, which in turn is described in the language of the pair density. As emphasized in chapter 8, when the Fermi hole density is used in conjunction with the QT-AIM partitioning of space on the basis of the topology of the density one obtains a description of electron localization and delocalization

in and between atoms in a molecule. The application of these ideas to aromatic molecules has been shown to yield a novel and powerful description of aromaticity. The new approach predicts chemical structures of the polycyclic aromatic hydrocarbons which are identical to those assigned based on empirical chemical reactivity (Clar's sextet theory), predicts experimental resonance energies, NMR proton-proton spin coupling constants, and explains why angular (less aromatic) isomers are more stable than linear isomers of polycyclic aromatic hydrocarbons. In this study (chapter 8) as well as in chapter 3, the occurence of a type of the newly discovered dihydrogen bonding (*Alkorta et al., 1998; Bodige et al., 1999; Cioslowski and Mixon, 1992; Garbowski, 2000; Garbowski, 2001; Popelier, 1998*) is examined in detail especially in relation to its stabilizing effect on the molecule.

The last problem studied in the thesis relates to the criteria of bonding especially in borderline cases. Thus, chapter 9 seeks to answer two related questions: (i) when are two atoms bonded by a "long bond" as opposed to being close in space and non-bonded, and (ii) what are the consequences of the presence of a bonded interaction? The case study of an organometallic titanium compound has shown how the presence of a bond path is accompanied by a host of effects, not the least of which is a significantly greater electron delocalization between bonded atoms, regardless of the distance of their separation. The dangers of assigning a chemical bonded structure solely on the basis of distance have been discussed in detail, and a new type of fluxional bonding between a transition metal and an aromatic ring (bonded cone of density) has been characterized. The practice of assigning a bonded structure on the basis of interatomic distances in x-ray crystallographic determinations has to be abandoned and replaced by a practice of assigning bonding whenever the density itself requires it to be so, i.e. whenever a bond path links the nuclei in question.

In conclusion, it is asserted that there is no observable property of any system, (*Bader, 1990*) including one induced by an external field (*Bader and Keith, 1993; Bader et al., 1992; Keith and Bader, 1996*) or by the absorption of light, (*Bader et al., 2000*) whether measured directly or obtained through a combination of measured values, that cannot be predicted using the physics of an open system, the physics of an atom in a molecule.

References

Abraham RJ, Fosher J, Loftus F (1998). *Introduction to NMR Spectroscopy*. John Wiley & Sons.

Alkorta I, Rozas I, Elguero J (1998). "Non-conventional hydrogen bonds". *Chem. Soc. Rev.*; **27**: 163-170.

Ángyán JG, Jansen G., Loos M, Hättig C, Heß BA (1994). "Distributed polarizabilities using the topological theory of atoms in molecules". *Chem. Phys. Lett.*; **219**: 267-273.

Archer MC (1987). "Chemical carcinogenesis". In: Tannock IF, Hill RP, (Eds.). *The Basic Science of Oncology*. New York: Pergamon Press, p.89-105.

Armit JW, Robinson R (1925). "Polynuclear heterocyclic aromatic types. Part II. Some anhydronium bases". *J. Chem. Soc.*; 1604-1618.

Ashcroft NW, Mermin ND (1976). *Solid State Physics*. New York: Saunders College Publishing: Harcourt Brace & Co.

Aubry A, Birlirakis N, Sakarellos-Daitsiotis M, Sakarellos C, Marraud M (1989). "A crystal molecular conformation of Leu-enkephalin related to the morphine molecule". *Biopolymers*; **28**: 27-40.

Bader RFW (1990). *Atoms in Molecules: A Quantum Theory*. Oxford, U.K.: Oxford University Press.

Bader RFW (1994). "Principle of stationary action and the definition of a proper open system". *Phys. Rev. B*; **49**: 13348-13356.

Bader RFW (1995). "Chemistry and the near-sighted nature of the one-electron density matrix". *Int. J. Quantum Chem.*; **56**: 409-419.

Bader RFW (1998a). "1997 Polanyi Award Lecture: Why are there atoms in chemistry?". *Can. J. Chem.*; **76** : 973-988.

Bader RFW (1998b). "A bond path: A universal indicator of bonded interactions". *J. Phys. Chem. A*; **102**: 7314-7323.

Bader, R. F. W. (2002).*Private communication.*

Bader RFW, Bayles D, Heard GL (2000). "Properties of atoms in molecules: transition probabilities". *J. Chem. Phys.*; **112**: 10095-10105.

Bader RFW, Becker P (1988). "Transferability of atomic properties and the theorem of Hohenberg and Kohn". *Chem. Phys. Lett.*; **148**: 452-458.

Bader RFW, Beddall PM (1972). "Virial field relationship for molecular charge distributions and the spatial partitioning of molecular properties". *J. Chem. Phys.*; **56**: 3320-3328.

Bader RFW, Chang C (1989). "Properties of atoms in molecules: electrophilic aromatic substitution". *J. Phys. Chem.*; **93**: 2946-2956.

Bader RFW, Gillespie RJ, MacDougall PJ (1988). "A physical basis for the VSEPR model of molecular geometry". *J. Am. Chem. Soc.*; **110**: 7329-7336.

Bader RFW, Gillespie RJ, Martín F (1998). "Core distortions in metal atoms and the use of effective core potentials". *Chem. Phys. Lett.*; **290**: 488-494.

Bader RFW, Johnson S, Tang T-H, Popelier PLA (1996). "The electron pair". *J. Phys. Chem.*; **100**: 15398-15415.

Bader RFW, Keith TA (1991). "The prediction and calculation of properties of atoms in molecules". *J. Mol. Struct.(Theochem)*; **234**: 75-94.

Bader RFW, Keith TA (1993). "Properties of atoms in molecules: magnetic susceptibilities". *J. Chem. Phys.*; **99**: 3683-3693.

Bader RFW, Keith TA, Gough KM, Laidig KE (1992). "Properties of atoms in molecules: additivity and transferability of group polarizabilities". *Mol. Phys.*; **75**: 1167-1189.

Bader RFW, Laidig KE (1992). "The definition of a chemical bond, of molecular structure and of its change as exemplified by the structure diagram for $C_4H_7^+$". *J. Mol. Struct. (Theochem)*; **261**: 1-20.

Bader RFW, Legare DA (1992). "Properties of atoms in molecules: structure and reactivities of boranes and carboranes". *Can. J. Chem.*; **70**: 657-676.

Bader RFW, MacDougall PJ, Lau CDH (1984). "Bonded and nonbonded charge concentrations and their relations to molecular geometry and reactivity". *J. Am. Chem. Soc.*; **106**: 1594-1605.

Bader RFW, Martín FJ (1998). "Interdeterminancy of basin and surface properties of an open system". *Can. J. Chem.*; **76**: 284-291.

Bader RFW, Matta CF (2001a). "Bonding to titanium". *Inorg. Chem.*; **40**: 5603-5611.

Bader RFW, Matta CF (2001b). "Properties of atoms in crystals: Dielectric polarization". *Int. J. Quantum Chem.*; **85**: 592-607.

Bader RFW, Matta CF, Martín FJ (2002). "Atoms in medicinal chemistry". In: Alber F, Carloni P, (Eds.). *Medicinal Quantum Chemistry*. Weinheim: Wiley-VCH. p.[In press].

Bader RFW, Popelier PLA, Chang C (1992). "Similarity and complementarity in chemistry". *J. Mol. Struct. (Theochem)*; **255**: 145-171.

Bader RFW, Popelier PLA, Keith TA (1994). "Theoretical definition of a functional group and the molecular orbital paradigm". *Angew. Chem. Int. Ed. Engl.*; **33**: 620-631.

Bader RFW, Stephens ME (1975). "Spatial localization of the electronic pair and number distributions in molecules". *J. Am. Chem. Soc.*; **97**: 7391-7399.

Bader RFW, Streitwieser A, Neuhaus A, Laidig KE, Speers P (1996). "Electron delocalization and the Fermi hole". *J. Am. Chem. Soc.*; **118**: 4959-4965.

Bader RWF, Heard GL (1999). "The mapping of the conditional pair density onto the electron density". *J. Chem. Phys.*; **111**: 8789-8797.

Badger GM (1969). *Aromatic Character and Aromaticity*. Cambridge: Cambridge University Press.

Bartel LS, Gavin RMJr (1964). "Effect of electron correlation in x-ray and electron diffraction". *J. Am. Chem. Soc.*; **86**: 3493-3498.

Bartle KD, Jones DW, Matthews (1969). "Correlation of long-range H^1-H^1 spin-spin coupling in polycyclic hydrocarbons with bond-order calculation". *Rev. Pure and Appl. Chem.*; **19**: 191-203.

Benabicha F, Pichon-Pesme V, Jelsch C, Lecomte C, Khmou A (2000). "Experimental charge density and electrostatic potential of glycyl-L-threonine dihydrate". *Acta Cryst.*; **B56**: 155-165.

Bennett BI, Maradudin AA (1972). "Pseudopotential calculation of transverse effective charges for III-V and II-VI compounds of zinc-blende structure". *Phys. Rev. B*; **5**: 4146-4154.

Berry MV (1984). "Quantal phase factors accompanying adiabatic changes". *Proc. Roy. Soc. Lond. A*; **392** : 45-57.

Biegler-König FW, Bader RFW, Tang T-H (1982). "Calculation of the average properties of atoms in molecules. II". *J. Comput. Chem.*; **13**: 317-328.

Biegler-König FW, Schönbohm J, Bayles D (2001). "AIM2000 - A program to analyze and visualize atoms in molecules". *J. Comput. Chem.*; **22**: 545-559.

Blinder SM (1965). "Basic concepts of self-consistent-field theory". *Am. J. Phys.*; **33**: 431-443.

Bodige SG, Sun D, Marchand AP, Namboothiri NN, Shukla R, Watson WH (1999). "Short H..H distances in norbornene derivatives". *J. Chem. Cryst.*; **29**: 523-530.

Born M (1926a). "Quantenmechanik der stoßuorgänge". *Z. Physik*; **38**: 803-827.

Born M (1926b). "Zur quantenmechanik der stoßuorgänge". *Z. Physik*; **37** : 863-867.

Boys SF, Bernardi F (1970). "The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors". *Mol. Phys.*; **19**: 553-566.

Bradbury AF, Smyth DG, Snell CR (1976). "Biothynthetic origin and receptor conformation of methionine enkephalin". *Nature*; **260**: 165-166.

Breneman CM, Rhem M (1997). "QSPR Analysis of HPLC column capacity factors for a set of high-energy materials using electrnic van der Waals surface property descriptors computed by transferable atom equivalent method". *J. Comput. Chem.*; **18**: 182-197.

Breneman CM, Thompson TR, Rhem M, Dung M (1995). "Electron density modeling of large systems using the transferable atom equivalent method". *Comput. Chem.*; **19**: 161-179.

Breneman CM, Weber LW (1991). "Transferable atom equivalents. Molecular electrostatic potentials from the electric multipoles of *PROAIMS* atomic basins". In: Jeffrey GA, Piniella JFE, (Eds.). *NATO ASI Series: The application of charge density research to chemistry and drug design*. New York: Plenum Press, p.357-358.

Brillouin L (1934). *Les Champs 'self-consistents` de Hartree et de Fock.* Paris: Hermann et Cie. p. 19.

Brookhart M, Green M. L. H. (1983). "Carbon-hydrogen-transition metal bonds". *J. Organomet. Chem.*; **250**: 395-408.

Cao M, Newton SQ, Pranata J, Schäfer L (1995). "Ab initio conformational analysis of alanine". *J. Mol. Struct. (Theochem)*; **332**: 251-267.

Carbó-Dorca R, Robert D, Amat Ll, Gironé X, Besalú E (2000). *Molecular Quantum Similarity in QSAR and Drug Design.* Berlin: Springer.

Carbó R, Leyda L, Arnau M (1980). "How similar is a molecule to another? An electron density measure of similarity between two molecular structures". *Int. J. Quantum Chem.*; **17**: 1185-1189.

Carroll MT, Bader RFW (1988). "An analysis of the hydrogen bond in BASE-HF complexes using the theory of atoms in molecules". *Mol. Phys.*; **65**: 695-722.

Carven BM, Weber HP, He XM (1987). *The POP Least-Squares Refinement Procedure, Technical Report, Department of Crystallography, University of Pittsburgh, USA.*

Chang C, Bader RFW (1992). "Theoretical construction of a polypeptide". *J. Phys. Chem.*; **96**: 1654-1662.

Cheeseman JR, Carroll MT, Bader RFW (1988). "The mechanics of hydrogen bond formation in conjugated systems". *Chem. Phys. Lett.*; **143**: 450-458.

Chelli R, Gervasio FL, Procacci P, Schettino V (2002). "Stacking and T-shape competition in aromatic-aromatic amino acid interactions". *J. Am. Chem. Soc.*; **124**: 6133-6143.

Chen C-S, Parthasarathy R (1977). "Structure and conformation of amino acids containing sulfur. V. N-Formyl-L-methionione". *Acta Cryst.*; **B33**: 3332-3336.

Chipot C, Ángyán JG, Maigret B, Scheraga HA (1993a). "Modeling amino acid side chains. 2. Determination of point charges from electrostatic properties: toward transferable point charge models". *J. Phys. Chem.*; **97**: 9788-9796.

Chipot C, Ángyán JG, Maigret B, Scheraga HA (1993b). "Modeling amino acid side chains. 3. Influence of intra- and intermolecular environment on point charges". *J. Phys. Chem.*; **97**: 9797-9807.

Chipot C, Maigret B, Rivail JL, Scheraga HA (1992). "Modelling amino acid side chains. 1. Determination of net atomic charges form ab initio self-consistent-field molecular electrostatic properties". *J. Phys. Chem.*; **96**: 10276-10284.
(Correction: *J. Phys. Chem.* 1993: **97**; 3452).

Cioslowski J, Mixon ST (1992). "Universality among topological properties of electron density associated with the hydrogen-hydrogen nonbonding interactions". *Can. J. Chem.*; **70**: 443-449.

Clar E (1972). *The Aromatic Sextet.* London: John Wiley and Sons Ltd.

Clementi E, Cavallone F, Scordamaglia R (1977). "Analytical potential from "ab initio" computations for the interaction between biomolecules. 1. Water with amino acids". *J. Am. Chem. Soc.*; **99**: 5531-5545.

Clinton WL, Frishberg CA, Massa LJ, Oldfield PA (1973). "Methods for obtaining an electron-density matrix from x-ray data". *Int. J. Quantum Chem.*; **7**: 505-514.

Clinton WL, Galli AJ, Massa LJ (1969). "Direct determination of pure-state density matrices. II. Construction of constrained idempotent one-body densities". *Phys. Rev.*; **177**: 7-12.

Clinton WL, Massa LJ (1972). "Determination of the electron density matrix from x-ray diffraction data". *Phys. Rev. Lett.*; **29**: 1363-1366.

Cody V (1985). "X-ray crystal structures of amino acids and selected derivatives". In: Barrett GC, (Ed.). *Chemistry and Biochemistry of the Amino Acids*. London: Chapman and Hall, p.625-653.

Coleman AJ, Yukalov VI (2000). *Reduced Density Matrices: Coulson's Challenge*. Berlin: Springer.

Collantes ER, Dunn WJI (1995). "Amino acid side chain descriptors for quantitative structure-activity relationship studies of peptide analogues". *J. Med. Chem.*; **38**: 2705-2713.

Cooper DL, Stutchbury NCJ (1985). "Distributed multipole analysis from charge partitioning by zero-flux surfaces: The structure of HF complexes". *Chem. Phys. Lett.*; **120**: 167-172.

Coppens P (1997). *X-ray Charge Densities and Chemical Bonding*. New York: Oxford University Press, Inc.

Coppens P, Abramov Y, Carducci M, Korjov B, Novozhilova I, Alhambra C, Pressprich MR (1999). "Experimental charge densities and intermolecular interactions: Electrostatic and topological analysis of DL-Histidine". *J. Am. Chem. Soc.*; **121**: 2585-2593.

Cotton FA (2002). "A half-century of nonclassical organometallic chemistry: a personal perspective". *Inorg. Chem.*; **41**: 643-658.

Coulson CA (1939). "The electronic structure of some polyenes and aromatic molecules. VII. Bonds of fractional order by the molecular orbital method". *Proc. Roy. Soc. (London) A*; **169**: 413-428.

Coulson CA (1961). *Valence, (Second Edition)*. New York: Oxford University Press.

Coulson CA, Longuet-Higgins HC (1947). "The electronic structure of conjugated systems. I. General theory". *Proc. Roy. Soc. (London) A*; **191**: 39-60.

Coulson CA, Longuet-Higgins HC (1948). "The electronic structure of conjugated systems. Parts III and IV. III. Bond orders in unsaturated molecules. IV. Force constants and interaction constants in unsaturated hydrocarbons". *Proc. Roy. Soc. (London) A*; **193** : 447-464.

Creighton TE (1983). *Proteins: Structures and Molecular Principles*. New York: W. H. Freeman and Co. p.249.

Cremer D, Kraka E (1984). "Chemical bonds without bonding electron density - Does the difference electron-density analysis suffice for a description of the chemical bond?". *Angew. Chem. Int. Ed. Engl.*; **23**: 627-628.

Dahaoui S, Pichon-Pesme V, Howard JAK, Lecomte C (1999). "CCD Charge density study on crystals with large unit cell parameters: The case of hexagonal L-cystine ". *J. Phys. Chem. A*; **103**: 6240-6250.

Dalhus B, Görbitz CH (1996). "Triclinic form of DL-valine". *Acta Cryst.*; **C52**: 1759-1761.

Davidson ER (1976). *Reduced Density Matrices in Quantum Chemistry*. New York: Academic Press, Inc.

Desiraju GR, Steiner T (1999). *The Weak Hydrogen Bond in Structural Chemistry and Biology*. Oxford: Oxford University Press, Inc.

Desnoyers JE, Verrall RE, Conway BE (1965). "Electrostriction in aqueous solutions of electrolytes". *J. Chem. Phys.*; **43**: 243-250.

Destro R, Bianchi R, Gatti C, Merati F (1991). "Total electronic charge density of L-alanine from x-ray diffraction at 23 K". *Chem. Phys. Lett.*; **186**: 47-52.

Destro R, Marsh RE (1988). "A low temperature (23 K) study of L-alanine". *J. Phys. Chem.*; **92**: 966-973.

Destro R, Roversi P, Barzaghi M, Marsh RE (2000). "Experimental charge density of α–glycine at 23 K". *J. Phys. Chem. A*; **104**: 1047-1054.

Dirac PAM (1958). *The Principles of Quantum Mechanics, Third Edition*. Oxford: Oxford University Press.

Doucet J-P, Weber J (1996). *Computer-Aided Molecular Design: Theory and Applications*. London: Academic Press, ltd.

Dovesi, R, Saunders, VR, Roetti, C (1992). *CRYSTAL92-An Ab Initio Hartree-Fock LCAO Program for Periodic Systems.*

Duspara PA, Matta CF, Jenkins SI, Harrison PHM (2001). "Twisted amides: synthesis and structure of 1,6-dipivaloyl-3,4,7,8-tetramethyl-2,5-dithioglycoluril". *Org. Lett.*; **3**: 495-498.

Engelman DM, Steitz TA (1984). "On the folding and insertion of globular membrane proteins". In: Wetlaufer D, (Ed.). *The Protein Folding Problem*. New York: American Association for the Advancement of Science (AAAS).

Espinosa E, Lecomte C, Molins E, Veintemillas S, Cousson A, Paulus W (1996). "Electron density study of a new non-linear optical material: L-arginine phosphate monohydrate (LAP). Comparison between *X-X* and *X-(X+N)* refinements". *Acta Cryst.*; **B52**: 519-534.

Feinberg AP, Creese I, Snyder S (1976). "The opiate receptor: A model explaining structure-activity relationships of opiate agonist and antagonists". *Proc. Natl. Acad. Sci. USA*; **73**: 4215-4219.

Fermi E (1928). "Eine statistische methode zur bestimmung einiger eigenschaften des atoms und ihre anwendung auf die theorie des periodischen systems der elemente". *Z. Phys.*; **48**: 73-79.

Feynman RP, Leighton RB, Sands M (1964). *The Feynman Lectures on Physics. (Volume II)*. Reading, MA: Addison-Wesley Publishing Co., Inc.

Figgis BN, Reynolds PA, Williams GA (1980a). "Spin density and bonding in the $CoCl_4^{2-}$ ion in Cs_3CoCl_5. Part 2. Valence electron distribution in the $CoCl_4^{2-}$ ion". *J. Chem. Soc., Dalton Trans.*; 2339-2347.

Figgis BN, Reynolds PA, Williams GA, Mason R, Smith ARP, Varghese JN (1980b). "Spin density and bonding in the $CoCl_4^{2-}$ ion in Cs_3CoCl_5 . Part 1. Magnetic structure factors from polarised neutron diffraction". *J. Chem. Soc., Dalton Trans.*; 2333-2338.

Flaig R, Koritsanszky D, Zobel D, Luger P (1998). "Topological analysis of the experimental electron densities of amino acids. 1. D,L-Aspartic acid at 20 K". *J. Am. Chem. Soc.*; **120**: 2227-2238.

Fock V (1930). "Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems". *Z. Physik*; **61**: 126-148.

Fradera X, Austen MA, Bader RFW (1999). "The Lewis model and beyond". *J. Phys. Chem. A*; **103**: 304-314.

Frenking G, Pidun U (1997). "*Ab initio* studies of transtition-metal compounds: the nature of the chemical bond to a transition metal". *J. Chem. Soc., Dalton Trans.*; 1653-1662.

Frisch, MJ, Trucks, GW, Schlegel, HB, Gill, PMW, Johnson, BG, Robb, MA, Cheeseman, JR, Keith, T, Petersson, GA, Montgomery, JA, Raghavachari, K, Al-Laham, MA, Zakrzewski, VG, Ortiz, JV, Foresman, JB, Peng, CY, Ayala, PY, Chen, W, Wong, MW, Andres, JL, Replogle, ES, Gomperts, R, Martin, RL, Fox, DJ, Binkley, JS, Defrees, DJ, Baker, J, Stewart, JP, Head-Gordon, M, Gonzalez, C, Pople, JA (1995). *Gaussian94, Revision B.3.* Gaussian Inc.: Pittsburgh PA.

Frishberg C, Massa LJ (1981). "Idempotent density matrices for correlated systems from x-ray-diffraction structure factors". *Phys. Rev. B*; **24**: 7018-7024.

Fulmor W, Lancaster JE, Morton GO, Brown JJ, Howell CF, Nora CT, Hardy RAJr (1967). "Nuclear magnetic resonance studies in the 6,14-*endo*-Ethenotetrahydrothebaine series". *J. Am. Chem. Soc.*; **89**: 3322-3330.

Garbowski SJ (2000). "High-level ab initio calculations of dihydrogen-bonded complexes". *J. Phys. Chem. A*; **104**: 5551-5557.

Garbowski SJ (2001). "Ab initio calculations on conventional and unconventional hydrogen bonds - study of the hydrogen bond strength". *J. Phys. Chem. A*; **105**: 10739-10746.

Gatti, C (1998). *TOPOND98.* CNR-CSRSRC: Milano.

Gatti C, MacDougall PJ, Bader RFW (1988). "Effect of electron correlation on the topological properties of molecular charge distributions". *J. Chem. Phys.*; **88**: 3792-3804.

Gatti C, Saunders VR, Roetti C (1994). "Crystal field effects on the topological properties of the electron density in molecular crystals: the case of urea". *J. Chem. Phys.*; **101**: 10686-10696.

Gatti C, Silvi B, Colonna F (1995). "Dipole moment of the water molecule in the condensed phase: a periodic Hartree-Fock estimate". *Chem. Phys. Lett.*; **247**: 135-141.

Gillespie RJ (1972). *Molecular Geometry.* London: Van Nostrand Reinhold.

Gillespie RJ, Bayles D, Platts J, Heard GL, Bader RFW (1998). "The Lennard-Jones function: a quantitative description of the spatial pairing of electrons as determined by the exclusion principle". *J. Phys. Chem. A*; **102**: 3407-3414.

Gillespie RJ, Bytheway I, Tang T-H, Bader RFW (1996). "Geometry of the fluorides, oxofluorides, hydrides and methylides of vanadium (V), chromium (VI) and molybdenum (VI): understanding the geometry of some non VSEPR molecules in terms of core distortion". *Inorg. Chem.*; **35**: 3954-3963.

Gillespie RJ, Hargittai I (1991). *The VSEPR Model of Molecular Geometry*. Boston: Allyn and Bacon.

Gillespie RJ, Popelier PLA (2001). *Molecular Geometry and Chemical Bonding: From Lewis to Electron Densities*. New York: Oxford University Press.

Gopalan R. S., Kumaradhas P, Kulkarni GU, Rao CNR (2000). "An experimental charge density study of aliphatic dicarboxylic acids". *J. Mol. Struct.*; **521**: 97-106.

Görbitz CH, Dalhus B (1996a). "L-Cysteine, monoclinic form, redetermination at 120 K". *Acta Cryst.*; **C52**: 1756-1759.

Görbitz CH, Dalhus B (1996b). "Redetermination of L-leucine at 120 K". *Acta Cryst.*; **C52**: 1754-1756.

Gray CG, Gubbins KE (1984). *Theory of Molecular Fluids, (Vol.1)*. Oxford: Clarendon Press.

Gronert S, O'Hair RAJ (1995). "Ab initio studies of amino acid conformations. 1. The conformers of alanine, serine, and cysteine". *J. Am. Chem. Soc.*; **117**: 2071-2081.

Gunther H (1980). *NMR Spectroscopy: An Introduction*. John Wiley & Sons.

Hagler AT, Huler E, Lifson S (1977). "Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals". *J. Am. Chem. Soc.*; **96**: 5319-5327.

Hahn T (1957). "Structural characteristics of amino acids and carboxylic acids". *Z. Kristallogr.* **109** : 438-466.

Hansen NK, Coppens P (1978). "Testing aspherical atom refinement on small molecules data sets". *Acta Cryst.*; **A34**: 909-921.

Hartree DR (1928a). "The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods". *Proc. Cambridge Phil. Soc.*; **24**: 89-110.

Hartree DR (1928b). "The wave mechanics of an atom with a non-Coulomb central field. Part II. Some results and discussion". *Proc. Cambridge Phil. Soc.*; **24**: 111-132.

Hättig C, Jansen G, Heß BA, Ángyán JG (1997). "Intermolecular interaction energies by topologically partitioned electric properties II. Dispersion energies in one-centre and multicentre multipole expansions". *Mol. Phys.*; **91**: 145-160.

Head-Gordon T, Head-Gordon M, Frisch MJ, Brooks III C, Pople JA (1991). "Theoretical studies of blocked glycine and alanine peptide analogues". *J. Am. Chem. Soc.*; **113**: 5989-5997.

Head-Gordon T, Sorenson JM, Pertsemlidis A, Glaeser RM (1997). "Difference in hydration structure near hydrophobic and hydrophilic amino acids". *Biophys. J.*; **73**: 2106-2115.

Hendrickson JB, Cram DJ, Hammond GS (1970). *Organic Chemistry, (Third Edition)*. New York: McGraw-Hill Book Co.

Hernández-Trujillo J, Bader RFW (2001). "Properties of atoms in molecules: construction of one-density matrix from functional group densities". *J. Chem. Phys.*; **115**: 10595-10607.

Hinshelwood CN (1951). *The Structure of Physical Chemistry*. London: Oxford University Press.

Hinz H-JEd (1986). *Thermodynamic Data for Biochemistry and Biotechnology*. Berlin: Springer-Verlag.

Hohenberg P, Kohn W (1964). "Inhomogeneous electron gas". *Phys. Rev. B*; **136**: 864-871.

Housset D, Benabicha F, Pichon-Pesme V, Jelsch C, Maierhofer A, David S, Fontecilla-Camps JC, Lecomte C (2000). "Towards the charge-density study of proteins: a room-temperature scorpion-toxin structure at 0.96Å resolution as a first test case". *Acta Cryst.*; **D56**: 151-160.

Howard ST, Krygowski TM (1997). "Benzenoid hydrocarbon aromaticity in terms of charge density descriptors". *Can. J. Chem.*; **75**: 1174-1181.

Huang L, Massa L, Karle J (1996). "Kernel projector matrices for Leu^1-zervamicin". *Int. J. Quantum. Chem: Quantum Chem. Symp.*; **30**: 479-488.

Ide M, Maeda Y, Kitano H (1997). "Effect of hydrophobicity of amino acids on the structure of water". *J. Phys. Chem. B*; **101**: 7022-7026.

IUPAC-IUB Commission on Biochemical Nomenclature (1970a). "Abbreviations and symbols for the description of the conformation of polypeptide chains ". *J. Biol. Chem.*; **245**: 6489-6497.

IUPAC-IUB Commission on Biochemical Nomenclature (1970b). "Abbreviations and symbols for the description of the conformation of polypeptide chains ". *Int. J. Mol. Biol.*; **52**: 1-17.

Jelsch C, Pichon-Pesme V, Lecomte C, Aubry A (1998). "Transferability of multipole charge-density parameters: application to very high resolution oligopeptide and protein structures". *Acta Cryst.*; **D54**: 1306-1318.

Jelsch C, Teeter MM, Lamzin V, Pichon-Pesme V, Blessing RH, Lecomte C (2000). "Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin". *Proc. Natl. Acad. Sci. USA*; **97**: 3171-3176.

Karle IL, Karle J (1983). "[Leu^5]enkephalin: four cocrystallizing conformers with extended backbones that form an antiparallel β-sheet". *Acta Cryst. B*; **39**: 625-637.

Karplus M (1959). "Contact electron-spin coupling of nuclear magnetic moments". *J. Chem. Phys.*; **30**: 11-15.

Kaufman JJ, Semo NM, Koski WS (1975). "Microelectronic titration measurement of the pK_a's and partition and drug distribution coefficients of narcotics and narcotic antagonists and thie pH and temperature dependence". *J. Med. Chem.*; **18** : 647-655.

Kebarle P (1976). "Solvent effects on acidity and basicity from gas phase ion equilibria measurements". *Jerusalem Symp. Quantum Chem. Biochem. (1975)*; **8**: 81-94.

Keith TA, Bader RFW (1992). "Calculation of magnetic response properties using atoms in molecules". *Chem. Phys. Lett.*; **194**: 1-8.

Keith TA, Bader RFW (1993). "Calculation of magnetic response properties using a continuous set of gauge transformations". *Chem. Phys. Lett.*; **210**: 223-231.

Keith TA, Bader RFW (1996). "Use of electron charge and current distributions in the determination of atomic contributions to magnetic properties". *Int. J. Quantum Chem.*; **60**: 373-379.

Keith TA, Bader RFW, Aray Y (1996). "Structural homeomorphism between the electron density and the virial field". *Int. J. Quantum Chem.*; **57**: 183-198.

King RB, Rouvray DH (1977). "Chemical applications of group-theory and topology. 7. Graph-theoretical interpretation of bonding topology in polyhedral boranes, carboranes, and metal clusters". *J. Am. Chem. Soc.*; **99**: 7834-7840.

King-Smith RD, Vanderbilt D (1993). "Theory of polarization of crystalline solids". *Phys. Rev. B*; **47**: 1651-1654.

Kittel C (1996). *Introduction to Solid State Physics, (Seventh Edition)*. New York: Wiley.

Koch W, Holthausen MC (2001). *A Chemist's Guide to Density Functional Theory, (Second Edition)*. New York: Wiley-VCH.

Kohn W, Sham LJ (1965). "Self consistent equations including exchange and correlation effects". *Phys. Rev. A*; **140 (4A)**: 1133-1138.

Koritsanszky TS, Coppens P (2001). "Chemical applications of X-ray charge-density analysis". *Chem. Rev.*; **101**: 1583-1628.

Kosov DS, Popelier PLA (2000a). "Atomic partitioning of molecular electrostatic potentials". *J. Phys. Chem. A*; **104**: 7339-7345.

Kosov DS, Popelier PLA (2000b). "Convergence of the multipole expansion for electrostatic potentials of finite topological atoms". *J. Chem. Phys.*; **113**: 3969-3974.

Krygowski TM (1993). "Crystallographic studies of inter- and intramolecular interactions reflected in aromatic character of π-electron systems". *J. Chem. Inf. Comput. Sci.*; **33**: 70-78.

Krygowski TM, Ciesielski A, Bird CW, Kotschy A (1995). "Aromatic character of the benzene ring present in various topological environments in benzenoid hydrocarbons. Nonequivalence of indices of aromaticity". *J. Chem. Inf. Comput. Sci.*; **35**: 203-210.

Lalitha V, Subramanian E, Bordner J (1984a). "Structure and conformation of linear peptides. II. Structure of glycyl-glycyl-L-isoleucine". *Int. J. Peptide Protein Res.*; **24**: 123-128.

Lalitha V, Subramanian E, Bordner J (1984b). "Structure and conformation of linear peptides. III. Structure of glycyl-glycyl-L-valine". *Int. J. Peptide Protein Res.*; **24**: 437-441.

Levine IN (2000). *Quantum Chemistry, (Fifth Edition)*. New Jersey: Prentice Hall.

Levinsky NG (1991). "Acidosis and alkalosis". In: Wilson JD, Braunwald E, Isselbacher KJ et al, (Eds.). *Harrison's Principles of Internal Medicine*. New York: McGraw-Hill, Inc., p.289.

Levy M (1979). "Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem". *Proc. Natl. Acad. Sci. USA*; **76**: 6062-6065.

Li S, Jiang Y (1995). "Bond lengths, reactivities, and aromaticities of benzenoid hydrocarbons based on the valence bond calculations". *J. Am. Chem. Soc.*; **117**: 8401-8406.

Liu B, McLean AD (1973). "Accurate calculation of the attractive interaction of two ground state helium atoms". *J. Chem. Phys.*; **59**: 4557-4558.

Loew GH, Villar HO, Cometta C, Perez J (1991). "Conformational and electronic properties of Met-enkephalin". *Int. J. Quantum Chem.: Quantum Biol. Symp.*; **18**: 165-181.

Macchi P, Proserpio DM, Sironi A (1998). "Experimental electron density in a transition metal dimer: metal-metal and metal-ligand bonds". *J. Am. Chem. Soc.*; **120**: 13429-13435.

Mager PP (1984). *Multidimensional Pharmacochemistry: Design of Safer Drugs*. London: Academic Press, Inc.

Martín FJ (2001). *"Theoretical Synthesis of Macromolecules from Transferable Functional Groups". Ph.D. Thesis*. Hamilton: McMaster University.

Martin RM (1974). "Comment on calculations of electric polarization in crystals". *Phys. Rev. B*; **9**: 1998-1999.

Massa L, Huang L, Karle J (1995). "Quantum crystallography and the use of kernel projector matrices". *Int. J. Quantum. Chem: Quantum Chem. Symp.*; **29**: 371-384.

Matta, CF (2001a). *AIMDELOC: Program to calculate AIM localization and delocalization indices (QCPE0802)*. Quantum Chemistry Program Exchange, Indiana University: IN. (http://qcpe.chem.indiana.edu/)

Matta, CF (2001b). *FRAGDIP: Program to calculate functional group contributions to the molecular dipole moment (QCPE0801)*. Quantum Chemistry Program Exchange, Indiana University: IN. (http://qcpe.chem.indiana.edu/)

Matta CF (2001c). "Theoretical reconstruction of the electron density of large molecules from fragments determined as proper open quantum systems: the properties of the oripavine PEO, enkephalins, and morphine". *J. Phys. Chem. A*; **105**: 11088-11101.

Matta CF, Bader RFW (2000). "An atoms-in-molecules study of the genetically-encoded amino acids. I. Effects of conformation and of tautomerization on geometric, atomic, and bond properties.". *Proteins Struct. Funct. Genet.*; **40**: 310-329.

Matta CF, Bader RFW (2002a). "An atoms-in-molecules study of the genetically-encoded amino acids. II. Computational study of molecular geometries". *Proteins Struct. Funct. Genet.*; **48**: 519-538.

Matta CF, Bader RFW (2002b). "An atoms-in-molecules study of the genetically-encoded amino acids. III. Atomic and bond properties and quantitative structure-to-activity relationships". *Proteins Struct. Funct. Genet.*; [Submitted].

Matta CF, Cow CC, Sun S, Britten JF, Harrison PHM (2000). "Twisted amides: crystal and optimized structures, and molecular geometry analysis of 1-acetyl-3,4,7,8-tetramethylglycoluril and 1,6-diacetyl-3,4,7,8-tetramethylglycoluril". *J. Mol. Struct.*; **523**: 241-55.

Matta CF, Cow NC, Harrison PHM (2002). "X-ray crystallographic and theoretical study of two acetylated glycolurils. Correlation of electron density and delocalization with amide twist and chemical reactivity". *J. Mol. Struct.*; [Submitted].

Matta CF, Gillespie RJ (2002). "Understanding and interpreting electron density distributions". *J. Chem. Educ.*; **79**: [Accepted, in press].

Matta CF, Hernández-Trujillo J (2002). "A theoretical study of aromaticity in polycyclic aromatic

hydrocarbons in terms of the properties of the electron density and of the pair density: benzene, naphthalene, anthracene, phenanthrene, chrysene and related molecules". *J. Org. Chem.*; [Submitted].

Matta CF, Hernández-Trujillo J, Bader RFW (2002). "Proton spin-spin coupling and electron delocalisation". *J. Phys. Chem. A*; [Accepted, in press].

McConnell HM (1956). "Molecular orbital approximation to electron coupled interaction between nuclear spins". *J. Chem. Phys.*; **24**: 460-467.

McConnell HM (1957). "Contribution of π-electrons to nuclear spin-spin coupling of aromatic protons". *J. Mol. Spect.*; **1**: 11-17.

McWeeny R (1967). "The nature of electron correlation in molecules". *Int. J. Quantum Chem.*; **1s**: 351-359.

Memory JD, Parker GW, Halsey JC (1966). "High-resolution NMR spectra of polycyclic hydrocarbons". *J. Chem. Phys.*; **45**: 3567-3570.

Memory JD, Wilson NK (1982). *NMR of Aromatic Compounds*. New York: John Wiley and Sons.

Messer RR (1977). *"The Virial Partitioning Theory and its Applications to Molecular Systems". Ph. D. Thesis*. Hamilton: McMaster University.

Michel AG, Proulx M, Evrard G, Norberg B, Milchert E (1988). "Molecular structure of opiate alkaloids. Part I. Crystal structure of two 16-alkyl analogues of thebaine and oripavine". *Can. J. Chem.*; **66**: 2498-2505.

Millero FJ, Surodo AL, Shin C (1978). "The apparent molal volumes and adiabatic compressibilities of aqueous amino acids at 25 °C". *J. Phys. Chem.*; **82**: 784-792.

Minitab (2000). *MINITAB, Release 13 for Windows*. Minitab Inc.: PA, U.S.A.

Moffitt W (1954). "The electronic structure of bis-cyclopentadienyl compounds". *J. Am. Chem. Soc.*; **76** : 3386-3392.

Momany FA, McGuire RF, Burgess AW, Scheraga HA (1975). "Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, non-bonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids". *J. Phys. Chem.*; **79**: 2361-2381.

Nelson D. (Ed.) (1998). *The Penguin Dictionary of Mathematics, (Second Edition)*. London: Penguin Books.

Némethy G (1967). "Hydrophobic interactions". *Angew. Chem. Int. Ed. Engl.*; **6**: 195-206.

O'Brien SE, Popelier PLA (2001). "Quantum molecular similarity. 3. QTMS descriptors". *J. Chem. Inf. Comput. Sci.*; **41**: 764-775.

Parr RG, Yang W (1989). *Density-Functional Theory of Atoms and Molecules*. Oxford: Oxford University Press.

Pascal P, Gallais F, Labarre J-F (1961). "Sur la susceptibilité diamagnétique des carbures éthyléniques et acétyléniques". *Compte Rend. Acad. Sci.*; **252**: 2644-2649.

Pelc SR (1965). "Correlations between coding-triplets and amino-acids". *Nature*; **207**: 597-599.

Pichon-Pesme V, Lecomte C, Lachekar H (1995). "On building a data bank of transferable experimental density parameters: application to polypeptides". *J. Phys. Chem.*; **99**: 6242-6250.

Pichon-Pesme V, Lecomte C, Wiest R, Bénard M (1992). "Modeling fragments for the ab initio determination of electron density in polypeptides. An experimental and theoretical approach to the electron distribution in Leu-enkephalin trihydrate". *J. Am. Chem. Soc.*; **114**: 2713-2715.

Pilar FL (1990). *Elementary Quantum Chemistry, (Second Edition)*. New York: Dover Publications.

Popelier PLA (1996). "Integration of atoms in molecules: a critical examination". *Mol. Phys.*; **87**: 1196-1187.

Popelier PLA (1998). "Characterization of a dihydrogen bond on the basis of the electron density". *J. Phys. Chem. A*; **102**: 1873-1878.

Popelier PLA (2000). *Atoms in Molecules: An Introduction*. London: Prentice Hall.

Popelier PLA (2001). "A fast algorithm to compute atomic charges based on the topology of the electron density". *Theor. Chem. Acc.*; **105**: 393-399.

Popelier PLA, Bader RFW (1992). "The existence of an intramolecular C-H-O hydrogen bond in creatine and carbamoyl sarcosine". *Chem. Phys. Lett.*; **189**: 542-548.

Popelier PLA, Bader RFW (1994). "Effect of twisting a polypeptide on its geometry and electron distibution". *J. Phys. Chem.*; **98**: 4473-4481.

Popelier PLA, Joubert L, Kosov DS (2001). "Convergence of the electrostatic interaction based on topological atoms". *J. Phys. Chem. A*; **105**: 8254-8261.

Popelier PLA, Logothetis G (1998). "Characterization of an agostic bond on the basis of the electron density". *J. Organomet. Chem.*; **555**: 101-111.

Pople JA, Beveridge DL (1970). *Approximate Molecular Orbital Theory*. New York: McGraw-Hill, Inc.

Pople JA, Schneider WG, Bernstein HJ (1959). *High-Resolution Nuclear Magnetic Resonance*. New York: McGraw-Hill Book Co. Inc.

Price SL, Faerman CH, Murray CW (1991). "Toward accurate transferable electrostatic models for polypeptides: a distributed multipole study of blocked amino acid residue charge distribution". *J. Comput. Chem.*; **12**: 1187-1197.

Ramek M, Cheng VKW (1992). "On the role of polarization functions in SCF calculations of glycine and related systems with intramolecular hydrogen bonding". *Int. J. Quantum Chem.: Quantum Biol. Symp.*; **19**: 15-26.

Ramsey NF (1953). "Electron coupled interactions between nuclear spins in molecules". *Phys. Rev.*; **91**: 303-307.

Reisine T, Pasternak G (1996). "Opioid analgesics and antagonists". In: Hardman JG, Limbird LE, Molinoff PB, Ruddon RW, Gilman AGE, (Eds.). *Goodman and Gilman's: The Pharmacological Basis of Therapeutics*. New York: McGraw-Hill, p.527.

Resta R (1993). "Macroscopic electric polarization as a geometric quantum phase". *Europhys. Lett.*; **22** : 133-138.

Resta R (1998). "Quantum-mechanical position operator in extended systems". *Phys. Rev. Lett.*; **80**: 1800-1803.

Resta R (2000). "Manifestations of Berry's phase in molecules and condensed matter". *J. Phys. : Condens. Matter*; **12**: R107-R143.

Rhodes G (1993). *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. San Diego: Academic Press, Inc.

Roe BP (1992). *Probability and Statistics in Experimental Physics*. New York: Springer. pp.23-25.

Roothaan CCJ (1951). "New developments in molecular orbital theory". *Rev. Mod. Phys.*; **23**: 69-89.

Russell B (1945). *A History of Western Philosophy*. New York: Simon and Schuster. p.165.

Salem L (1966). *The Molecular Orbital Theory of Conjugated Systems*. New York: W. A. Benjamin, Inc.

Sands DE (1975). *Introduction to Crystallography*. New York: Dover Publications, Inc.

Sapse A-M (2000). *Molecular Orbital Calculations for Amino Acids and Peptides*. Boston: Birkhäuser.

Schäfer L, Klup-Newton SQ, Siam K, Klimkowski VJ, Van Alsenoy C (1990a). "Ab initio studies of structural features not easily amenable to experiment. Part 71. Conformational analysis and structural study of valine and threonine.". *J. Mol. Struct.(Theochem)*; **209**: 373-385.

Schäfer L, Siam K, Klimkowski VJ, Ewbank JD, Van Alsenoy C (1990b). "Ab initio studies of structural features not easily amenable to experiment. Part 69. Conformational analysis and structural study of cysteine". *J. Mol. Struct. (Theochem)*; **204**: 361-372.

Scherer W, Hieringer W, Spiegler M, Sirsch P, McGrady GS, Downs AJ, Haaland A, Pedersen B (1998). "Characterisation of agostic interactions by a topological analysis". *Chem. Commun.*; 2471-2472.

Schiller PW (1984). "Conformational analysis of enkephalin and conformation-activity relationships". In: Udenfriend S, Meienhofer JE, (Eds.). *The Peptides: Analysis, Synthesis, Biology (Vol.6)*. New York: Academic Press, p.219-268.

Schleyer P.v-R., Maerker C, Dransfeld A, Jiao H, Hommes NJRvE (1996). "Nucleus-Independent chemical shifts: A simple and efficient aromaticity probe". *J. Am. Chem. Soc.*; **118**: 6317-6318.

Schleyer P.v-R., Manoharan M, Wang Z-X, Kiran B, Jiao H, Puchta R, Hommes NJRvE (2001). "Dissected nucleus-independent chemical shift analysis of π-aromaticity and antiaromaticity". *Org. Lett.*; **3**: 2465-2468.

Schleyer P.v-R. (Guest Ed.) (2001). "Aromaticity". *Chem. Rev.*; **101**.

Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, Windus TL, Dupuis M, Montgomery JA (1993). "General atomic and molecular electronic-structure system (GAMESS)". *J. Comput. Chem.*; **14**: 1347-1363.

Schrödinger E (1926a). "Quantisierung als eigenwertproblem". *Ann. D. Phys.*; **79**: 361-376.

Schrödinger E (1926b). "Quantisierung als eigenwertproblem". *Ann. D. Phys.*; **81**: 109-139.

Schrödinger E (1967). *What is Life?* Cambridge: Cambridge University Press.

Schwinger J (1951). "The theory of quantized fields. I.". *Phys. Rev.*; **82**: 914-927.

Shirazian S, Gronert S (1997). "The gas-phase conformations of valine: an ab initio study". *J. Mol. Struct.(Theochem)*; **397**: 107-112.

Siam K, Klimkowski VJ, Ewbank JD, Van Alsenoy C, Schäfer L (1984). "Ab initio studies of structural features not easily amenable to experiment. Part 39. Conformational analysis of glycine and alanine". *J. Mol. Struct. (Theochem)*; **110**: 171-182.

Slater JC (1930). "Note on Hartree's method". *Phys. Rev.*; **35**: 210-211.

Smith GV, Kriloff H (1963). "NMR coupling constants and conformations of cycloolefins". *J. Am. Chem. Soc.*; **85**: 2016.

Smith VH, Absar I (1977). "Basic concepts of quantum chemistry for electron density studies". *Israel J. Chem.*; **16**: 87-102.

Snyder SH (1977). "Opiate receptors and internal opiates". *Sci. Am.*; **236**: 44-56.

Sokalski WA, Maruszewski K, Hariharan PC, Kaufman JJ (1989). "Library of cumulative atomic multipole moments: II. Neutral and charged amino acids". *Int. J. Quantum Chem.: Quantum Biol. Symp.*; **16**: 119-164.

Sokalski WA, Poirier RA (1983). "Cumulative atomic multipole representation of the molecular charge distribution and its basis set dependence". *Chem. Phys. Lett.*; **98**: 86-92.

Sordo JA, Probst M, Corongiu G, Chin S, Clementi E (1987). "Ab initio pair potentials for the interactions between aliphatic amino acids". *J. Am. Chem. Soc.*; **109**: 1702-1708.

Spackman MA (1998). "Charge densities from x-ray diffraction data". *R. Soc. Chem. Ann. Rep. Section C*; **94**: 177-207.

Springford MEd (1997). *Electron: A Centenary Volume.* Cambridge: Cambridge University Press.

Srebrenik S, Bader RFW (1975). "Towards the development of the quantum mechanics of a subspace". *J. Chem. Phys.*; **63**: 3945-3961.

Steiner T (2002). "The hydrogen bond in the solid state". *Angew. Chem. Int. Ed. Engl.*; **41**: 48-76.

Steiner T, Koellner G (2001). "Hydrogen bonds with π-acceptors in proteins: frequencies and role in stabilizing local 3D structures". *J. Mol. Biol.*; **305**: 535-557.

Stephen MJ (1957). "A variational method for calculating nuclear spin-spin interactions in molecules". *Proc. Roy. Soc. (London) A*; **243**: 274-280.

Stewart RF (1976). "Electron population analysis with rigid pseudoatoms". *Acta Cryst.*; **A32**: 565-574.

Stewart RF (1977). "Total x-ray scattering and two-electron density functions". *Israel J. Chem.*; **16**:

111-114.

Stone AJ (1981). "Distributed multipole analysis, or how to describe a molecular charge distribution". *Chem. Phys. Lett.*; **83**: 233-239.

Stone AJ, Alderton M (1985). "Distributed multipole analysis methods and applications". *Mol. Phys.*; **56**: 1047-1064.

Stone AJ, Hättig C, Jansen G, Ángyán JG (1996). "Transferability of topologically partitioned polarizabilities - the case of n-alkanes". *Mol. Phys.*; **89**: 595-605.

Stout GH, Jensen LH (1989). *X-Ray Structure Determination: A Practical Guide, (Second Edition)*. New York: John-Wiley and Sons.

Subramanian E, Lalitha V, Bordner J (1984). "Structure and conformation of linear peptides. I. Structure of L-tyrosyl-L-tyrosine". *Int. J. Peptide Protein Res.*; **24**: 55-59.

Sun S, Britten JF, Cow CN, Matta CF, Harrison PHM (1998). "The crystal structure of 3,4,7,8-tetramethylglycoluril". *Can. J. Chem.*; **76**: 301-306.

Szabo A, Ostlund NS (1989). *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. New York: Dover Publications, Inc.

Takigawa T, Ashida T, Sasada Y, Kakudo M (1966). "The crystal structure of L-tryptophan hydrochloride and hydrobromide". *Bull. Chem. Soc. Japan*; **39**: 2369-2378.

Tarasov LV (1980). *Basic Concepts of Quantum Mechanics, (English Translation)*. Moscow: Mir Publishers.

Thomas LH (1927). "The calculation of atomic fields". *Proc. Cambridge Philos. Soc.*; **23**: 542-548.

Thomson JJ (1897). "Cathode rays". *Philos. Magaz. S. 5*; **44**: 293-316.

Tomaszewski R, Hyla-Kryspin I, Mayne CL, Arif AM, Gleriter R, Ernst R (1998). "Shorter non-bonded than bonded contacts or nonclassical metal-to-saturated carbon atom interactions?". *J. Am. Chem. Soc.*; **120**: 2959-2960.

Tóth G, Watts CR, Murphy RF, Lovas S (2001). "Significance of aromatic-backbone interactions in protein structure". *Proteins: Structure, Function and Genetics*; **43**: 373-381.

Tsirelson V, Zou PF, Bader RFW (1995). "Topological definition of crystal structure: determination of the bonded interactions in solid molecular chlorine". *Acta Cryst. A*; **51**: 143-153.

Van Alsenoy C, Klup S, Siam K, Klimkowski VJ, Ewbank JD, Schäfer L (1988). "Ab initio studies of structural features not easily amenable to experiment. Part 63. Conformational analysis and structure of serine". *Mol. Struct. (Theochem)*; **181**: 169-178.

van Duijneveldt FB, van Duijneveldt-van de Rijdt JGCM, van lenthe JH (1994). "State of the art in counterpoise theory". *Chem. Rev.*; **94**: 1873-1885.

Vilkov LV, Mastryukov VS, Sadova NI (1983). *Determination of the Geometrical Structure of Free Molecules, (English translation)*. Moscow: Mir Publishers.

Volkenstein MV (1965). "Coding of polar and non-polar amino acids". *Nature*; **207**: 294-295.

Volkenstein MV (1966). "The genetic coding of protein structure". *Biochim. Biophys. Acta*; **119**: 421-424.

Volkenstein MV (1977). *Molecular Biophysics*. New York: Academic Press, Inc.

Volkov A, Gatti C, Abramov Y, Coppens P (2000). "Evaluation of net atomic charges and atomic and molecular electrostatic moments through topological analysis of the experimental charge density". *Acta Cryst.*; **A56**: 252-258.

Watanabe N, Kamata Y, Yamauchi K, Udagawa Y (2001). "Calculation of x-ray scattering intensities by means of coupled cluster singles and doubles model". *J. Comput. Chem.*; **22**: 1315-1320.

Watanabe T, Hashimoto K, Takase H, Kikuchi O (1997). "Monte Carlo simulation study on the conformation and interaction of the glycine zwitterion in aqueous solution". *J. Mol. Struct. (Theochem)*; **397**: 113-119.

Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986). "An all atom force-field for simulation of proteins and nucleic acids". *J. Comput. Chem.*; **7**: 230-252.

Wheland GW (1955). *Resonance in Organic Chemistry*. New York: John Wiley and Sons.

White SH, Wimley WC (1999). "Membrane protein folding and stability: physical principles". *Annu. Rev. Biophys. Biomol. Struct.*; **28**: 319-365.

Wiberg KB, Bader RFW, Lau CDH (1987). "Theoretical analysis of hydrocarbons properties. 2. Additivity of group properties and the origin of strain energy". *J. Am. Chem. Soc.*; **109**: 1001-1012.

Wiest R, Pichon-Pesme V, Bénard M, Lecomte C (1994). "Electron distributions in peptides and related molecules. Experimental and theoretical study of Leu-enkephalin trihydrate". *J. Phys. Chem.*; **98**: 1351-1362.

Wilson EK (2001). "Cover Story. Computers and chemistry: symposium speakers showcase new developments in computational chemistry". *Chem. Eng. News*; **79**: 39-44.

Woese CR (1965). "Order in the genetic code". *Proc. Natl. Acad. Sci. USA*; **54**: 71-75.

Wolfenden R, Andersson L, Cullis PM, Southgate CCB (1981). "Affinities of amino acid side chains for solvent water". *Biochem.* ; **20**: 849-855.

Zou PF, Bader RFW (1994). "A topological definition of a Wigner-Seitz cell and the atomic scattering factor". *Acta Cryst. A*; **50**: 714-725.