THEORETICAL SYNTHESIS OF MACROMOLECULES FROM TRANSFERABLE FUNCTIONAL GROUPS

By

FERNANDO J. MARTÍN, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

THEORETICAL
SYNTHESIS
OF
MACROMOLECULES

Doctor of Philosophy (2001)
(Chemistry)

McMaster University
Hamilton, Ontario

TITLE: Theoretical Synthesis of Macromolecules from Transferable Functional Groups

AUTHOR: Fernando J. Martín, Bachelor of Science (York University, Canada)

SUPERVISOR: Professor R. F. W. Bader

NUMBER OF PAGES: xiv, 174

Abstract

The theory of Atoms In Molecules (AIM) generalizes quantum mechanics by deriving the physics of an atom in a molecule. This is accomplished by associating an atom with a quantum mechanical *open system*, a region of real space capable of exchanging electrons and properties with the rest of the molecule. AIM recovers the basic cornerstone of chemistry: that atoms and functional groups possess characteristic and additive properties. When the properties of an atom or functional group are found to be transferable, the theory makes possible the synthesis of macro-molecules from the corresponding pieces.

Many molecules of biological importance are composed of repeating structural units that appear to be highly transferable. These repeating units, when defined by AIM, are found to account for and recover this transferable behavior. As a consequence, theory can be used to construct a biological macromolecule from repeating units previously defined in simpler systems.

The properties of amino acid residues, the building blocks of proteins, are determined in this thesis by sandwiching each residue between two simpler amino acids to form a three-residue protein. The form and properties of the residues determined in this manner are found to be highly transferable. Thus, this thesis reports the tabulation of a library of amino acid residues that can then be linked together to form any desired protein or portion thereof. Such a synthesis makes possible the theoretical prediction of the properties of large molecules and does so using modest computing resources. The resulting availability of the electron density distribution, among other properties of the protein, enables one to apply in a quantitative manner the models and tools, such as the electrostatic field and the van der Waals shape, presently employed in the investigation of problems in molecular recognition and docking. Perhaps more importantly, it makes possible new approaches to the study of these problems.

Acknowledgements

I would like to express my deepest thanks to those who have directly and indirectly helped in defining the path that led me to work with professor Bader. Of these people, Freddy and Mari Paz deserve special mentioning as they extended their hand to me at a rather difficult stage of my life. This thesis is indeed the result of their unending support and encouragement.

I would also like to thank Professor Richard Bader, my thesis supervisor and above all my friend. His contagious enthusiasm for quantum mechanics has given intellectual life to all of those who have had the pleasure of working with him, as well as the curiosity and desire to continue extending the horizons of the theory of chemistry, the theory of Atoms In Molecules.

I am indebted to my past and present peers, *badoroids* like myself, in particular to Maggie Austen and Cherif Matta. Maggie has been instrumental in the removal of *spanglish* outbursts from this typescript and has always made herself available for theoretical and chemical discussions. Cherif has provided some of the data presented in this thesis, as well as invaluable commentary. In addition, I would like to thank Dr. David Bayles, Dr. Jesus Hernandez Trujillo and Dr. George Heard with whom I have shared many interesting discussions, generally accompanied by frosted pints and pizza at the "Camel".

I am extremely grateful to Professor R. S. Dumont, whom I think of as a friend, and who has taught me everything I know about numerical methods. His vast knowledge of mathematics has never ceased to amaze me.

I would also like to thank Professor D. Yang for his invaluable input in the underlying chemistry presented in this thesis.

Lastly, I would like to thank Johann Sebastian Bach for keeping me company during all those long sessions in the laboratory. His music, along with the development of practical applications of the theory of AIM, is the object of my obsession.

Esta tesis es dedicada a Freddy y Mari Paz

Contents

List of Tables

List of Figures

Contents

Chapter 1

Introduction

The birth of chemistry as we know it, one could argue, took place with the acceptance of the atom as the basic indivisible unit constituent of matter. Dalton's postulate that these indivisible units not only combined to give rise to everything that we see and touch, but that in doing so they also retained their identity with little change - which we now understand to be a consequence of having an unchanging nucleus - solidified what we know as chemistry today as a distinct branch of science. Since then, in our attempt to learn more about matter; how it affects everything that surrounds and defines us, the experimental chemist has built a vast knowledge of atomic *behavior*. This *behavior* has been characterized in terms of how atoms interact with one another. The theoretician has taken this vast amount of data and developed models to reproduce the experimental observations in a bid to extend our chemical understanding. In so doing he has created a modeling toolset that can reproduce and predict a great variety of chemical phenomena. This toolset, at its most advanced stages, is comprised of *ab initio* methodologies.

Much like modern chemistry evolved from the fixing of the atomic axiom, theoretical chemistry evolved from Schrödinger's derivation of wave mechanics in 1926. The realization that all the information about a system could be contained in a *wavefunction* is of paramount importance. Since then, electronic methods have evolved to tackle many-electron problems with great success. Unfortunately, despite great strides made towards the advancement of computer hardware and software to date, the modeling of macromolecules is prohibitory due to the lack of computational resources needed to implement the *ab initio* paradigm, even for small proteins. This hurdle has created the feeling that not until the computation of wavefunctions for large systems is accessible will the properties of these systems be studied at the quantum level. This thesis will attempt to prove otherwise.

The computational chemist is faced with two tasks:

(a) the generation of a wavefunction for a system of interest and

(b) the extraction of the chemical information contained therein.

Tackling the first task is accomplished with the use of commercial and freeware molecular electronics software. The ways in which to implement the second premise seems still the subject for much debate, although as this thesis demonstrates this is unreasonably so. This debate surfaces with the question of (a) where does an atom begin and end in a molecule and (b) whether its properties can be defined. The theory of Atoms In Molecules (AIM) has solved this ambiguity by providing a quantum framework which provides answers to these questions. Thus, the theory of AIM allows the computational chemist to recover the properties of atoms or groupings of these - functional groups - as they are constituents to a chemical system.

But how does AIM do it? The answer lies in the study of the electron density, a physically measurable quantity which has a definite value at each point in space. Like any non-homogeneous field it presents topological features succinctly summarised by its critical points; the locations in real space at which the gradient of the electron density vanishes. Further chemical insight is provided by the study of another field of the electronic density, that of its gradient. By mapping the gradient paths which expand the two-dimensional manifolds referred to as the "interatomic surfaces", AIM partitions a molecule into atoms. Such partitioning scheme is unbiased since it is fixed by the distribution of the electron density which in turn is fixed by the wavefunction representing the state of the system under study. But the topological atom need not be the quantum one. For that to be the case, the partitioning scheme must have a foundation in physics. AIM shows this to be possible by demonstrating that the topologically defined interatomic surfaces provide the boundary condition for the derivation of the physics of an *atom* in a *molecule*. A self-contained review of AIM, in full mathematical rigor, is presented on chapter 2.

We now know that the partitioning of a molecule into smaller systems is feasible, and that one can measure its properties, but how could this be of aid to us? Once again we turn to experimental observation to obtain answers. Chemists are very aware of the concept of functional group. Through observation, an empirical classification of the properties exhibited by linked sets of atoms has been, and still is, produced. In chemistry we recognize the presence of a group in a larger system and predict its effect upon the static and dynamic properties of the latter in terms of the characteristic set of properties assigned to that group. We can generally make very sound predictions of a yet to be synthesized compound on the basis of the functional groups it contains.

Theory can parallel experiment in this regard. In chapter 3 we examine functional groups and their properties. A functional group is recognized as a specific grouping of atoms which exhibits a particular set of characteristics. The functional groups of theory will be proven to be those of chemistry in that their classification and prediction of chemical behavior is contained within the framework of quantum mechanics. In some instances these groups can be transfered between systems without incurring significant changes in their form and properties. The lack in changes is commensurate with the group's *transferability*. This understanding provides us with the possibility of paralleling the experimental chemist in synthesizing molecules from functional groups, only that we no longer are limited to the use of beakers; we may use computers.

Perhaps the most important functional groups in chemistry are those that make up the molecules of life. Biological macromolecules are constituted of a small set of amino acid residues, nucleotides, simple sugars and phosphates. These building blocks are linked to form polymers in a nearly infinite number of ways; each variation serving a specific biological purpose. But the manifestation of such a small set of building blocks in such vast assortment of macro-structures seems to be indicative of the need that they be transferable. Chapter 4 studies the transferability exhibited by a subset of these building blocks: the amino acid residues. To accomplish this, wavefunctions for all amino acid residues are computed and their properties analyzed as these residues would be constituents in a mono-stranded perfect α-helix protein. To replicate the neighborhood of the residues we make use of a molecular template; the *glycine mold*. This mold is constituted of three linked amino acid residues, two glycines positioned at the termini, and the residue which we would like to model in the center. The cast is constrained to ensure that the middle residue locally matches the secondary structure it would exhibit should it be part of a macro-molecule with a specific secondary structure. The glycine mold is used to produce a wavefunction from which the properties and form of the middle fragment are extracted. By computing wavefunctions of the glycine mold with all the naturally occurring amino acid residues, and then calculating the properties of the middle residues, one can determine the degree of transferability of the properties exhibited by these residues. This chapter shows that amino acid residues are indeed transferable.

Following the line of reasoning outlined above;

(a) the realization that there is but one way of defining an atom in a molecule with the consequent retrieval of its properties,

(b) that these atoms form linked sets with specific identities as reflected by their properties and

(c) that these groups can be transferable, in particular in those instances where they make up

biological systems,

we find ourselves in a position to describe what it is meant by *theoretical* synthesis. Chapter 5 tackles this premise. Because the form of atoms in molecules is determined by their distribution of the electron density one should be able to model the residues which exhibited transferability for the purposes of creating proteins. This exercise entails the manipulation of the electron density distributions in real space. The reader will realize the elegance and simplicity of the methodology presented.

This thesis concludes with the author's view on the future work that naturally ensues from the material presented here. Chapter 6 proposes possible avenues of research which could be used to draw benefits from the synthetic methodology presented in this thesis, and geared towards the devising of predictive means of the reactivity of the modeled macromolecules.

Chapter 2

AIM background

Quantum mechanical computations require that the Hamiltonian matrix operator be diagonalized to render descriptions of the allowed states of a molecular system of interest. This exercise provides the computational chemist with three sources of information:

(a) an energy classification of the allowed states,

(b) the associated wavefunctions of these states and

(c) their molecular geometries.

Of these, the energy eigenvalue and the corresponding molecular geometry are necessary for the modeling of chemical behavior. In the past, the wavefunction was subjected to analysis in terms of its orbitals, but more recently it is used to compute the molecular electron density distribution of the system under study. The study of the topology of the electron density not only provides means of classifying the features of molecular structure in depth, but it also holds the key to the definition of an atom as a region of real space bounded by surfaces recovered from quantum mechanics. This background chapter provides a discussion centered around this issue.

Having understood that the electron density is the vehicle for the definition of an atom in a molecule, a more mathematically-rigorous section ensues, where the derivation of Schrödinger's energy functional for an *open system* is presented. The identification of an open system with a submolecular region capable of exchanging charge with the rest of the molecule it participates in is crucial in the development of the means required to recover its properties. As it will be demonstrated, an open system finds its most indivisible manifestation as an atom in a molecule. After a review of the principal tenets of the theory of AIM, it will be shown how the physics of an open system leads

to the generalization of quantum mechanics to allow for the recovery of atomic properties which are computed via implementing the associated *atomic theorem*.

This chapter is then concluded with a summary of the tools defined and used within the AIM framework for the description of molecular structure and properties, and which will be required for the understanding of the contents and objectives of this thesis.

2.1 The Electron Density

The first postulate of quantum mechanics dictates that the state of a chemical system is completely determined by a function Ψ, which depends on the position of the particles it represents, their associated spin and time. This function is generally called the "wavefunction" of the system and has the important property that $\Psi^*\Psi$ determines the probability density of a system possessing a given configuration. The wavefunction is more concisely depicted as $\Psi(\mathbf{x}, X, t)$, where $\mathbf{x}$ is the set of position and spin coordinates of the electrons, X the nuclear coordinates and t parameterizes this system's evolution in time. Certain assumptions can be made to simplify the wavefunction. When one is only interested in stationary states, as it is the case with this thesis, the wavefunction is not dependent on the time variable. In addition, one can fix the nuclear position variables with the Bohr-Oppenheimer approximation, thus reformulating $\Psi(\mathbf{x}, X, t)$ as $\psi(\mathbf{x}; X)$. The probability of finding *each one* of the electrons in a volume element $d\tau_i = dx_i\, dy_i\, dz_i$ with a spin coordinate σ_i becomes

$$\psi^*(\mathbf{x}; X)\psi(\mathbf{x}; X)d\mathbf{x}_1 d\mathbf{x}_2 \ldots d\mathbf{x}_N \quad , \tag{2.1}$$

where

$$for \quad i = 1 \ldots N$$

$$d\mathbf{x}_i = d\tau_i \sigma_i \; ; \quad \sigma_i = \alpha \; or \; \beta \quad .$$

The probability of finding *one* electron in the volume element $d\tau_1$, independent of spin, is obtained by summing over all the spins and integrating over all the spatial coordinates of all the electrons but one

$$Prob_e = \sum_{i=1}^{N} \sigma_i \left[\int d\tau_2 \int d\tau_3 \ldots \int d\tau_N \psi^*(\mathbf{x}; X)\psi(\mathbf{x}; X) \right] d\tau_1 \quad . \tag{2.2}$$

Since the wavefunction is antisymmetrized, it makes no difference which electron we wish to set as our reference; they all share the same probability $Prob_e$ of being in $d\tau$. Multiplying this

probability by the number of electrons in the system yields the expression for the total electron density

$$\rho(r;X) = N\sum_{i=1}^{N}\sigma_i\left[\int d\tau_2\int d\tau_3\ldots\int d\tau_N \quad \psi^*(\mathbf{x};X)\psi(\mathbf{x};X)\right] \quad , \tag{2.3}$$

a function independent of the spin variables. For simplicity eqn.[2.3] can be rewritten as

$$\rho(r) = N\int d\tau' \quad \psi^*\psi \quad , \tag{2.4}$$

where the mode of integration $\int d\tau'$ denotes the sum over all spins and the integration over the spatial coordinates of all the electrons but one. Unlike the wavefunction, the electron density is physically observable (Coppens 1997). It is in using the electron density that AIM sets the quantum foundation for the definition of an atom in a molecule.

2.2 Molecular topology and the electronic density.

Any distribution of molecular electronic density can be concisely characterized by the number, location and classification of its critical points. A critical point is classified according to its *rank* and *signature* using the expression (ω,σ). The rank ω of a critical point at r_c equals the number on non-zero eigenvalues (curvatures) of the Hessian matrix $\mathcal{A}$ where

$$\mathcal{A} = \sum_{i,j} a_{i,j}; \quad \forall \quad a_{i,j} = \frac{\partial^2\rho(r)}{\partial r_i \partial r_j} \quad . \tag{2.5}$$

Since $\mathcal{A}$ is real and symmetric, its diagonalization renders real eigenvalues and associated real eigenvectors. These eigenvalues quantify the curvatures in the directions of their associated eigenvectors. The signature (σ) of the critical point is the excess number of positive over negative eigenvalues. Critical points which rank $\omega < 3$, in three-dimensional fields, are degenerate critical points. These are unstable in the mathematical sense; any perturbation of the nuclear configuration will cause these degenerate points to bifurcate into stable critical points or vanish altogether. Inasmuch as this thesis is concerned, only stable critical points are encountered. In three dimensional space only four types of stable critical points are possible:

$(3,-3)$ All the curvatures are negative and ρ is a local maximum at r_c. Such points indicate nuclear positions, and thus represent the most important topological feature leading to the definition of an atom in a molecule. Although classifying nuclear "critical points" as such is a misnomer in the mathematical sense ($\partial\psi/\partial r$ is discontinuous at the position of a nucleus as the potential

energy becomes infinite), their occurrence serves as a cusp of all the gradient paths in the electronic density for practical purposes. Figure 2.1 shows a map of the gradient vector

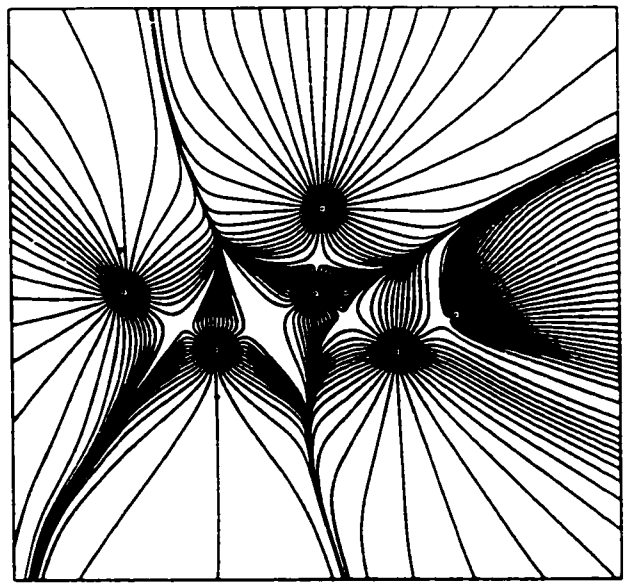


Figure 2.1: A collection of gradient paths for the plane containing the heavy atoms in glycine. The leftmost cusp corresponds to the amide Nitrogen whereas the rightmost is that of the acidic hydrogen. The position of the nuclei are depicted by "+" symbols.

field of the electron density for the plane containing the heavy nuclei of glycine.

$(3,-1)$ Two curvatures are negative and $\rho(r)$ is a maximum at r_c, the location of the critical point, in the plane defined by their corresponding axes. The pair of eigenvectors associated with the two negative eigenvalues of a $(3,-1)$ critical point generate a set of trajectories which terminate at the critical point and define a surface, as depicted in figure 2.2. The physical significance of this surface will be explored in greater depth in the following sections, but suffice it to say that it provides a natural partitioning of molecules into atoms. It is for this reason that such surface is also called an "interatomic surface". The unique pair of trajectories originating at the $(3,-1)$ critical point, and in the direction of the eigenvector associated with the positive eigenvalue, define the trajectories in $\nabla\rho(r)$ which terminate at the neighboring nuclei. This trajectory defines the line through space along which the electron density is a maximum with respect to any neighboring trajectory in $\nabla\rho(r)$. Such line is called as a bond path. The set of bond paths in a molecule at its equilibrium geometry constitute this molecule's *molecular*

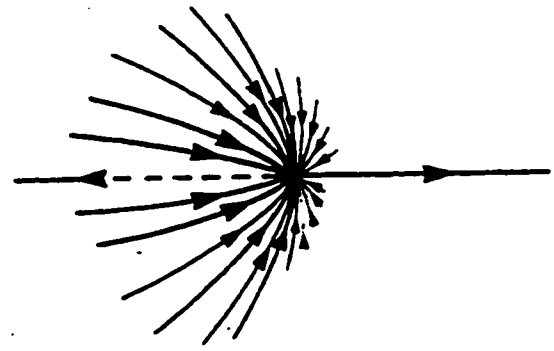


Figure 2.2: The local topology of $\nabla\rho(\mathbf{r})$ at a bond critical point.

graph, a universal indicator of bonding. Figure 2.3 depicts the molecular graph of the plane containing the heavy atoms in glycine. The curved trajectories which extend to the perimeter of the image correspond to the interatomic surfaces. The shorter lines connecting the nuclei (depicted by "+" symbols) are the computed bond paths.

$(3,+1)$ Two curvatures are positive and ρ is a minimum at $\mathbf{r}_c$ in the plane defined by their two axes. ρ is a maximum along the third axis, perpendicular to this plane. This type of critical point denotes the presence of a ring structure. Figure 2.4 illustrates the three types of critical points described so far via a portrayal of gradient paths for the plane of the heavy nuclei of the glycyl group in the C_s structure of N-formylglycine amide. A bond critical point, denoted by a solid dot, is the point of intersection of the bond path and the associated inter-atomic surface. The ring critical point is depicted by a solid triangle.

$(3,+3)$ All the curvatures are positive, and the density at this position is a minimum. This type of critical point denotes the presence of a cage structure. Such critical points arise in regions fully enclosed by ring structures. Critical points of this type are not encountered in this work.

The number of critical points and their type for a given nuclear configuration is regulated by the Pointcaré-Hopf relationship, which states that

$$n - b + r - c = 1 \tag{2.6}$$

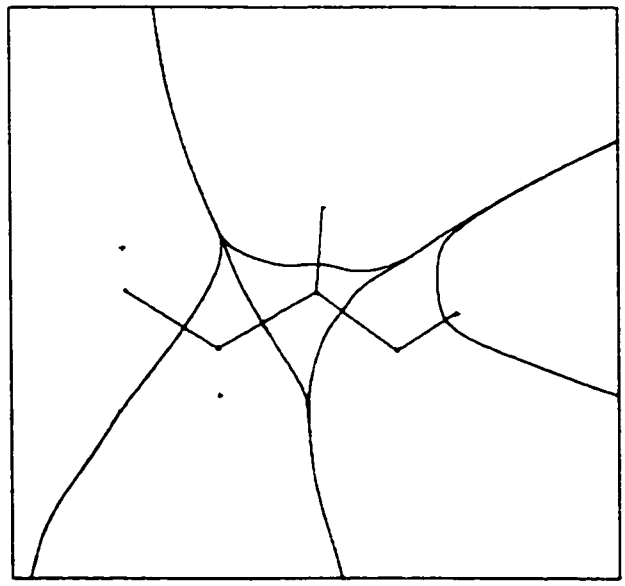


Figure 2.3: View of the surfaces of zero-flux and of the molecular graph of the plane containing the heavy atoms in glycine. The points at which the surfaces intersect with the bond paths correspond to the location of the bond critical points.

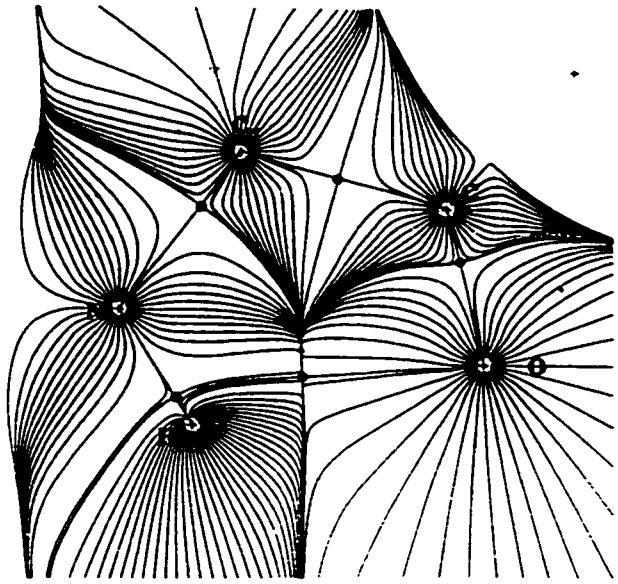


Figure 2.4: The local topology in $\nabla\rho(r)$ which which gives rise to the formation of a ring critical point, denoted by a solid triangle near the center of the image. The ring is a consequence of the presence of the hydrogen bond between the H of $N-H$ and the keto oxygen atom.

where n is the number of $(3,-3)$ critical points or nuclei, b is the number of bond critical points $(3,-1)$, r is the number of rings as denoted by those points of $(3,+1)$ signature, and c is the number of cage critical points or $(3,+3)$. The collection of integers $(n,\ b,\ r,\ c)$ is called the *characteristic set* of the molecule.

2.3 The Observation of an Atom in a Molecule.

Bader and co-workers initiated a study of the electron density in the 1960s. Through a collaboration with Mulliken-Roothan laboratory, the Bader group had access to wavefunctions for many diatomic molecules with energies very close to the Hartree-Fock limit. They observed in the initial studies of the topology of the electron density (and that of its gradient) that atoms in molecules are delimited from one another by their corresponding interatomic surfaces, introduced in the preceding section. It is to be noted that the theory thus has an observational basis in the topology exhibited by the measurable electron density distribution. These surfaces exhaustively partition a molecule into mononuclear regions dominated by maxima in the electron density at the positions of the nuclei. From the location of these maxima, one can follow trajectories in the gradient of the electron density through the atomic subspace, as depicted by figure 2.2. These gradient paths never cross any interatomic surface. On this basis, the interatomic surfaces can be characterized as one of "zero-flux". This condition is defined as

$$\nabla\rho(\boldsymbol{r})\cdot\mathbf{n}(\boldsymbol{r})=0; \qquad \forall \quad \boldsymbol{r}\in S(\Omega,\boldsymbol{r}) \quad , \tag{2.7}$$

where Ω refers to the Cartesian domain of the atom and S its bounding surface. Equation [2.7] tells us that no gradient trajectories in $\nabla\rho$ transverse the surface S. As it will be shown in ensuing sections, eqn.[2.7] will prove pivotal in defining the physics of an atom in a molecule. The space traversed by the set of gradient paths that terminate at a given nucleus define the basin associated with that nucleus, as illustrated in Figure 2.4. The definition of an *atom* is thus the union of a nucleus and its basin.

Of course, these topological observations on their own do not necessarily provide a bridge between the atoms of chemistry and the atoms of theory. To establish such link one should examine whether the set of properties of atom other than the electron density, as defined within the framework of quantum mechanics as these atoms are constituents in different chemical systems. To answer this question, Bader and co-workers averaged the kinetic energy density as well as the electron density over these topologically defined atoms. One can define two local expressions for the kinetic energy

density:

i) the Schrödinger kinetic energy density

$$K(r) = \frac{-\hbar^2}{4m} N \int d\tau' \left[\psi^* \nabla^2 \psi + (\nabla^2 \psi^*)\psi\right] \quad , \tag{2.8}$$

and

ii) the gradient kinetic energy density

$$G(r) = \frac{\hbar^2}{2m} N \int d\tau' \nabla\psi^* \cdot \nabla\psi \quad . \tag{2.9}$$

It is readily demonstrated that both definitions of the kinetic energy density yield the same average value for the kinetic energy for an atom in a molecule. The difference in the two kinetic energy densities, for the one-electron case, is given by

$$\begin{aligned} K(r) - G(r) &= \frac{-\hbar^2}{4m}\left(\psi^* \nabla^2 \psi + \nabla^2 \psi^* \psi\right) - \frac{\hbar^2}{2m}\left(\nabla\psi^* \nabla\psi\right) \quad , \\ &= \frac{-\hbar^2}{4m}\left(\psi^* \nabla^2 \psi + \nabla^2 \psi^* \psi + 2\nabla\psi^* \nabla\psi\right) \quad . \end{aligned} \tag{2.10}$$

Since

$$\begin{aligned} \nabla^2 \rho &= \nabla^2(\psi^* \psi) \\ &= \nabla \cdot (\nabla\psi^* \psi + \psi^* \nabla\psi) \\ &= \nabla^2 \psi^* \psi + \psi^* \nabla^2 \psi + 2\nabla\psi^* \nabla\psi \quad , \end{aligned} \tag{2.11}$$

one finds

$$K(r) - G(r) = \frac{-\hbar^2}{4m} \nabla^2 \rho(r) \quad . \tag{2.12}$$

Because of the divergence theorem, when eqn. [2.12] is averaged over an atom in a molecule, the volume integral of the Laplacian function can be transformed into a surface integral. The latter should average to zero for an atom in a molecule because of the zero-flux condition. Thus, because of eqn.[2.7]

$$K(\Omega) - G(\Omega) = \frac{-\hbar^2}{4m} \int_\Omega d\tau \nabla \cdot \nabla\rho(r) \tag{2.13}$$

$$= \frac{-\hbar^2}{4m} \oint dS(\Omega, r) \nabla\rho(r) \cdot \mathbf{n}(r) \tag{2.14}$$

$$= 0 \quad , \tag{2.15}$$

demonstrating that the average electronic kinetic energy is well defined for an atom in a molecule. These studies of the kinetic energy density resulted in a most important observation; that both $K(r)$ and $G(r)$ exhibit the same degree of transferability over the basin of an atom up to its zero-flux surface as does the electron density $\rho(r)$. Thus, the conservation of $\rho(r)$, and its integrated population, on transfer was paralleled by a conservation in the electronic kinetic energy density and its atomic average. This paralleling behavior is the crucial observation that leads to the theory of atoms in molecules, as deduced from the following chain of reasoning.

The virial theorem for a system with Coulombic forces states that the total energy of a set of particles equals the negative of their average kinetic energy, $E = -T$. In terms the observation made by Bader and co-worker's described above, the energy of an atom in a molecule equals the negative of the averaged kinetic energy over such atom. If one could show that there is a virial theorem for an atom in a molecule then one could use this theorem to define the energy of an atom in a molecule in terms of its averaged kinetic energy, i.e.

$$T(\Omega) = -E(\Omega) \quad . \tag{2.16}$$

Since $T(\Omega)$ is additive, this definition of the energy of an atom would necessarily be additive as well, and the total electronic energy of a molecule would be the sum of its atomic contributions

$$E = \sum_{\Omega} E(\Omega) \quad . \tag{2.17}$$

Most importantly, because of the paralleling behavior of $\rho(r)$ and $G(r)$, one would know that when the form of the atom in real space remains unchanged on transfer between systems, so does its contribution to the total energy. That is, by observation, the energy of an atom would be transferable to the same extent as is its distribution of charge. Thus, the topological atom, as defined by eqn.[2.7], meet the very observational requirements of additivity and transferability discussed above which are essential for the definition of an atom in a molecule.

2.4 Derivation of the Hypervirial Theorem for an open system Ω

The virial theorem is but one of the theorems derived from the hypervirial theorem of quantum mechanics. Before proceeding with the development of the quantum mechanics of an atom in a molecule we first consider this theorem for both the total system and for an open system with

arbitrary boundaries. We do this to demonstrate the important differences between the physics of an open system and that of the total one.

The hypervirial theorem states that (Levine 1970)

$$\langle\psi|[\hat{H},\hat{A}]|\psi\rangle = 0 \quad , \tag{2.18}$$

or in words, the average of the commutator of $\hat{H}$, the Hamiltonian operator

$$\hat{H} = \frac{-\hbar^2}{2m}\nabla^2 + \hat{V}(r) \quad , \tag{2.19}$$

with any observable $\hat{A}$ vanishes for a stationary state, a consequence of $\hat{H}$ and $\hat{A}$ being hermitian[1] operators and that ψ is an eigenfunction of $\hat{H}$. To prove this theorem we consider

$$\langle\psi|[\hat{H},\hat{A}]|\psi\rangle = \langle\psi|\hat{H}\hat{A}|\psi\rangle - \langle\psi|\hat{A}\hat{H}|\psi\rangle \quad . \tag{2.20}$$

Because of the hermiticity of $\hat{H}$ the R.H.S. of eqn. [2.20] can be expressed as

$$\langle\hat{H}\psi|\hat{A}\psi\rangle - \langle\psi|\hat{A}\hat{H}|\psi\rangle \quad , \tag{2.21}$$

which because of Schrödinger's equation $\hat{H}\psi = E\psi$, the commutator becomes

$$\begin{aligned}\langle\psi|[\hat{H},\hat{A}]|\psi\rangle &= E\left\{\langle\psi|\hat{A}|\psi\rangle - \langle\psi|\hat{A}|\psi\rangle\right\} \\ &= 0 \quad . \end{aligned} \tag{2.22}$$

(2.23)

But, operators which are hermitian over all space may cease to be so in a molecular subsystem. This is because a complete system is enclosed exclusively by the *natural boundary condition*, i.e., the region of space where

$$\nabla\psi^* \cdot \mathbf{n} = 0 \quad or \quad \nabla\psi \cdot \mathbf{n} = 0 \quad . \tag{2.24}$$

This is not the case for an atomic subsystem where ψ has a finite value at the surfaces of zero-flux. If $\hat{H}$ is no longer hermitian over a subsystem, then eqn. [2.18] expression need not evaluate to zero. In this section we derive the hypervirial theorem for an atom in a molecule. For the sake of clarity, we will make use of the one-electron case without incurring a loss of generality.

We begin by evaluating eqn. [2.18] over a subsystem Ω as

$$\langle\psi|[\hat{H},\hat{A}]|\psi\rangle_\Omega = \langle\psi|\hat{H}\hat{A}|\psi\rangle_\Omega - \langle\psi|\hat{A}\hat{H}|\psi\rangle_\Omega \quad , \tag{2.25}$$

[1] An operator $\hat{A}$ is said to be hermitian if $\left(\hat{A}^*\right)^T = \hat{A}$.

if one adds and subtracts the term $\langle\hat{H}\psi|\hat{A}\psi\rangle_\Omega$, the R.H.S. of eqn. [2.25] becomes

$$\begin{aligned}\langle\psi|[\hat{H},\hat{A}]|\psi\rangle_\Omega &= \langle\psi|\hat{H}\hat{A}|\psi\rangle_\Omega - \langle\hat{H}\psi|\hat{A}\psi\rangle_\Omega \\ &\quad \langle\hat{H}\psi|\hat{A}\psi\rangle_\Omega - \langle\psi|\hat{A}\hat{H}|\psi\rangle_\Omega \quad . \end{aligned} \tag{2.26}$$

The last two terms on the R.H.S. of eqn.[2.26] vanish as a direct result of Schrödinger's equation as in eqn.[2.22]. Upon expanding the remaining terms in eqn.[2.26] one obtains

$$\begin{aligned}\langle\psi|[\hat{H},\hat{A}]|\psi\rangle_\Omega &= \frac{\hbar^2}{2m}\int_\Omega d\tau \quad \left[\nabla^2\psi^*\hat{A}\psi - \psi^*\nabla^2(\hat{A}\psi)\right] \\ &= \frac{\hbar^2}{2m}\int_\Omega d\tau \quad \nabla\cdot\left[\nabla\psi^*\hat{A}\psi - \psi^*\nabla\hat{A}\psi\right] \quad , \end{aligned} \tag{2.27}$$

By making use of Gauss' theorem, one can transform the volume integral to a surface one as

$$\langle\psi|[\hat{H},\hat{A}]|\psi\rangle_\Omega = \frac{\hbar^2}{2m}\oint dS(\Omega,r)\left[\nabla\psi^*\hat{A}\psi - \psi^*\nabla\hat{A}\psi\right]\cdot\mathbf{n}(r) \quad . \tag{2.28}$$

The R.H.S. of eqn.[2.28] corresponds to the flux in the current density of $\hat{A}$ through the zero-flux surfaces bounding Ω, and its integrand can be defined explicitly as the current density associated with the operator $\hat{A}$

$$\mathbf{j}_A(r) = \frac{\hbar}{2mi} \quad [\psi^*\nabla\hat{A}\psi - \nabla\psi^*\hat{A}\psi] \quad . \tag{2.29}$$

To ensure that a real quantity be obtained from the averaging of the commutator one must take the mean of eqn.[2.28] and its complex conjugate (*cc*). The hypervirial theorem for an open system reduces to

$$\frac{1}{2}\left[\left(\frac{i}{\hbar}\right)\langle\psi|[\hat{H},\hat{A}]|\psi\rangle_\Omega + cc\right] = \frac{1}{2}\left[\oint dS(\Omega,r)\mathbf{j}_A\cdot\mathbf{n}(r) + cc\right] \quad , \tag{2.30}$$

and is readily extendible to the many-electron case, as indicated in section 2.1. The appearance of the flux term is pivotal in describing the properties of an open system. When $\hat{H}$ and $\hat{A}$ don't commute there is a flux in the property $\hat{A}$ through the surface bounding the open system, of magnitude equal and opposite in sign to the value of the commutator. The current density operator for the many-electron case defined in eqn.[2.29] becomes

$$\mathbf{j}_A(r) = \frac{\hbar}{2mi} \quad N\int d\tau'\left[\psi^*\nabla\hat{A}\psi - \nabla\psi^*\hat{A}\psi\right] \quad . \tag{2.31}$$

When $\hat{A} = \hat{1}$, the unit operator, the current operator describes the flow of the electron density through space

$$\mathbf{j}(r) = \frac{\hbar}{2mi} \quad N\int d\tau'\left[\psi^*\nabla\psi - \nabla\psi^*\psi\right] \quad . \tag{2.32}$$

At this stage we introduce a more concise name for a subsystem. A subregion of a molecular space is not independent from the rest of the molecule but rather exchanges properties in the form of fluxes in the current densities of these properties through the subsystem's surfaces of zero-flux. Thus, it is more precise to rename a subsystem as an *open system*, or one that is free to exchange electrons and properties with the rest of the molecule.

2.5 The calculus of variations: A review

One could argue that today's understanding of quantum mechanics could not have been achieved had it not been for the work of Lagrange and Euler in the branch of mathematics called the *calculus of variations* (Smith 1974). This discipline deals with the problem of extremizing the value of a functional J of the form

$$J = \int_{x_1}^{x_2} f(y, y_x, x)dx \quad , \tag{2.33}$$

where f is a known function of $y \equiv y(x)$, $y_x \equiv \frac{dy(x)}{dx}$, and x, such that the dependence of y on x is not fixed. The implication of the latter statement is that although the integral in [2.33] is of a definite

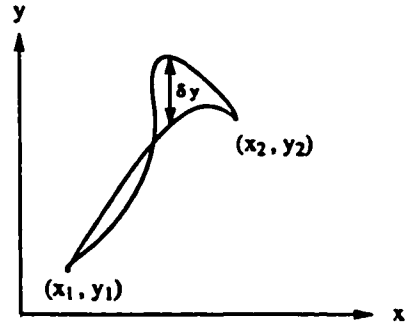


Figure 2.5: A varied path

form, the function $y(x)$ is not known, and hence neither is the path of integration. The extremizing of J may render stationary states of the functional, i.e. minima, maxima and saddle points. As an illustrated example, two possible paths for $y(x)$ are shown in figure 2.5, two of the infinite number of paths that $y(x)$ could map. By choosing two arbitrary paths one can then compute the variation of their difference. To this end, the difference between any two arbitrary paths δy constrained such that both paths pass through x_1 and x_2 can be mapped by the introduction of an arbitrary function $\eta(x)$ along with a scaling factor α.

This arbitrary function must comply with two prerequisites:

$$\eta(x_1) = \eta(x_2) = 0$$

and

$$\eta(x) \quad \textbf{must be differentiable.} \tag{2.34}$$

The extent of the variation between and *optimal unknown* path and a trial path can be quantified by making use of the function

$$y(x, \alpha) = y(x, 0) + \alpha\eta(x) \quad , \tag{2.35}$$

with a corresponding variation

$$\delta y = y(x, \alpha) - y(x, 0) = \alpha\eta(x) \quad . \tag{2.36}$$

By choosing $\alpha = 0$ to give rise to an unknown reference path $y(x, \alpha = 0)$ which will minimize J, we can use $y(x, \alpha)$ to describe any neighboring path. The redefined functional J in terms of the added parameter α takes the form

$$J(\alpha) = \int_{x_1}^{x_2} f[y(x, \alpha), y_x(x, \alpha), x]\, dx \quad . \tag{2.37}$$

This function is to be extremized at $\alpha = 0$, and much like in differential calculus, one first proceeds by taking the derivative

$$\frac{\partial J(\alpha)}{\partial \alpha} = \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y}\frac{\partial y}{\partial \alpha} + \frac{\partial f}{\partial y_x}\frac{\partial y_x}{\partial \alpha}\right] dx \quad . \tag{2.38}$$

Thanks to our definition of the adjacent path equation [2.35], we can readily obtain

$$\frac{\partial y(x, \alpha)}{\partial \alpha} = \eta(x) \tag{2.39}$$

$$\frac{\partial y_x(x, \alpha)}{\partial \alpha} = \frac{d\eta(x)}{dx} \tag{2.40}$$

rendering equation [2.38] in a more specific form

$$\frac{\partial J(\alpha)}{\partial \alpha} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y}\eta(x) + \frac{\partial f}{\partial y_x}\frac{d\eta(x)}{dx}\right) dx \quad , \tag{2.41}$$

which after integrating the second term by parts yields:

$$\int_{x_1}^{x_2} \frac{d\eta(x)}{dx}\frac{\partial f}{\partial y_x}\, dx = \eta(x)\frac{\partial f}{\partial y_x}\bigg|_{x_1}^{x_2} - \int_{x_1}^{x_2} \eta(x)\frac{d}{dx}\frac{\partial f}{\partial y_x}\, dx \quad . \tag{2.42}$$

Since it was one of our prerequisites that the introduced arbitrary function vanish at the points x_1 and x_2, the limits of the variation as denoted by [2.34], the integrated part in eqn. [2.42] vanishes. The extremized equation [2.41] becomes:

$$\int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y_x}\right] \eta(x)\, dx = 0 \quad . \tag{2.43}$$

Equation [2.43] can only be satisfied if the expression enclosed within square brackets vanishes at any point other than the extremes, where η vanishes. Thus, the variation reduces to finding solutions to

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y_x} = 0 \quad , \tag{2.44}$$

the so called "Euler" equation. Solving eqn.[2.44] may render one or more functions, i.e. $y(x)$, which minimize J.

This methodology was used by Schrödinger who made the energy functional $\mathcal{G}[\psi]$ stationary with respect to variations in the wavefunction ψ subject to the constraint that ψ remain normalized, to render

$$\mathcal{G}[\psi] = \int \left\{ \left(\frac{\hbar}{2m}\right) \nabla\psi^* \cdot \nabla\psi + (\hat{V} + \lambda)\psi^*\psi \right\} d\tau \quad . \tag{2.45}$$

As seen in our one-dimensional example, the integrated term of eqn. [2.42] gives rise to the natural boundary condition in Schrödinger's derivation. The integrand in eqn.[2.45] depicts the summation of the kinetic energy operator, $\left(\frac{\hbar}{2m}\right) \nabla\psi^* \cdot \nabla\psi$ and $\hat{V}$, the potential energy operator. This derivation yielded the familiar expression

$$\hat{H}\psi = E\psi \tag{2.46}$$

as the Euler solution to the variation, where $\hat{H}$ is the Hamiltonian operator and $E = -\lambda$ is the corresponding Lagrange multiplier required to ensure compliance with the constraint imposed on ψ. Further information about the calculus of variations can be found in (Arfken 1985, Feynman 1964, Gelfand and Fomin 2000).

2.6 Schrödinger's energy functional for an open system Ω

The exercise of deriving the hypervirial theorem for an open system proved that the properties of an open system can be measured through the flux in the current density of the associated observable through the open system's surfaces of zero-flux. It remains to be determined whether these open

systems have physical significance in a quantum mechanical sense. This section provides the answer to that question as obtained via the derivation of Schrödinger's energy functional for an open system.

Schrödinger demonstrated that making the energy functional

$$\mathcal{G}[\psi] = \int d\tau \left\{ \left(\frac{\hbar^2}{2m} \right) \nabla\psi^* \cdot \nabla\psi + V\psi^*\psi \right\} \tag{2.47}$$

stationary with respect to first-order variations in the state vector ψ and subject to the constraint that ψ remain normalized yielded the Euler relationship $\hat{H}\psi = E\psi$ as the solution. The eigenvalue E is the constant introduced as the Lagrange multiplier required to impose the constraint on ψ. We can now parallel Schrödinger's derivation only this time extended to an atom in a molecule. Let's define Schrödinger's energy functional for an open system Ω as (Srebrenik and Bader 1975)

$$\mathcal{G}[\psi, \Omega] = \int_\Omega \left\{ \left(\frac{\hbar^2}{2m} \right) \nabla\psi^* \cdot \nabla\psi + (\hat{V} - E)\psi^*\psi \right\} \, d\tau \quad , \tag{2.48}$$

where the undetermined multiplier for the normalization of ψ is made equal to $-E$, the total energy of the system obtained upon optimizing ψ. Since ψ does not satisfy the natural boundary condition on Ω eqn. [2.24], the variation of eqn.[2.48] must include a variation of the surface as well as that of ψ on the surface, to produce (Bader 1990)

$$\begin{aligned} \delta\mathcal{G}[\psi, \Omega] &= \int_\Omega \left\{ \left(\frac{\partial f}{\partial \psi} \right) \delta\psi + \left(\frac{\partial f}{\partial \nabla\psi} \right) \delta\nabla\psi \right\} \, d\tau \\ &+ \oint f(\psi, \nabla\psi) \; \delta S_\Omega(\Omega, r) dS(\Omega, r) + cc \quad , \end{aligned} \tag{2.49}$$

where the basin term finds a parallel with eqn.[2.41], and where

$$f(\psi, \nabla\psi) = \frac{\hbar^2}{2m} (\nabla\psi^* \cdot \nabla\psi) + (\hat{V} - E)\psi^*\psi \quad , \tag{2.50}$$

where $\delta\psi$ and $\delta\nabla\psi$ are infinitesimal variations of the wavefunction and its gradient. The surface enclosing the open system is denoted by $S(\Omega, r)$, a function of the position vector r. The term δS refers to the variation of this surface through a variation in the wavefunction. Lastly cc refers to the complex conjugate term of the entire preceding expression which arises from allowing the wavefunction be complex. Ridding eqn.[2.49] of the $\delta\nabla\psi$ terms by integrating by parts, and after some additional manipulations, the expression

$$\begin{aligned} \delta\mathcal{G}[\psi, \Omega] &= \int_\Omega \left\{ \hat{H}\psi - E\psi \right\} \delta\psi \; d\tau \\ &+ \oint \left\{ \left(\frac{\hbar^2}{2m} \right) \nabla\psi^* \cdot \mathbf{n}(r) \; \delta\psi + \delta S(\Omega, r) \; f(\psi, \nabla\psi) \right\} \, dS(\Omega, r) + cc \quad , \end{aligned} \tag{2.51}$$

is obtained where $\hat{H}$ is the Hamiltonian operator defined in eqn.[2.19]. Equation [2.51] is a generalization of the variation of Schrödinger's energy functional. If the system was exclusively enclosed by

the natural boundary condition, then the surface integral term in eqn. [2.51] vanishes so that the variation becomes

$$\delta\mathcal{G}[\psi] = \int d\tau \left\{ \hat{H}\psi^* - E\psi^* \right\} \delta\psi + cc = 0 \quad , \tag{2.52}$$

recovering Schrödinger's expression for the variation of the complete system. Expression [2.52] is made equal to zero in order to fulfill Schrödinger's requirement that the functional remain stationary to variations in the state vector. This equation can only be satisfied if

$$\hat{H}\psi^* - E\psi^* = 0 \quad \text{and} \quad \hat{H}\psi - E\psi = 0 \quad , \tag{2.53}$$

which are the already familiar Schrödinger equations. Paralleling the reasoning described above, the volume integral in eqn.[2.51] vanishes on the grounds that its integrand must comply with eqn.[2.53]. This renders the derivation of the open system as

$$\delta\mathcal{G}[\psi,\Omega] \;=\; \oint \left\{ \left(\frac{\hbar^2}{2m}\right) \nabla\psi^* \cdot \mathrm{n}(r)\, \delta\psi \;+ \delta S(\Omega,r)\; f(\psi,\nabla\psi) \right\} dS(\Omega,r) + cc \quad . \tag{2.54}$$

Unlike the case for the complete system, the variation of the energy functional for an open system does not vanish, but is instead proportional to the variation of ψ on the boundary together with the variation of the boundary itself $S(\Omega,r)$. The later expression is not operational since the surface term is a function which we do not know *a priori* in an analytical form. Such a term can be replaced on the grounds of its direct dependence on ψ. In order to make the replacement we will focus temporarily on the integrand $f(\psi,\nabla\psi)$ defined in eqn.[2.50]. One notices that when Schrödinger equations apply, eqn. [2.50] is expressible as

$$f(\psi,\nabla\psi) = \left\{ \psi^*\hat{H}\psi - E\psi^*\psi \right\} + \left(\frac{\hbar^2}{4m}\right) \nabla^2(\psi^*\psi) \tag{2.55}$$

and becomes

$$\begin{aligned} f(\psi,\nabla\psi) \;&=\; \frac{\hbar^2}{4m}\, \nabla^2(\psi^*\psi) \\ &=\; \frac{\hbar^2}{4m}\, \nabla^2\rho(r) \quad . \end{aligned} \tag{2.56}$$

A similar result has already been encountered in section 2.3 where the difference of both expressions for the kinetic energy density averaged to an expression of the Laplacian of the electronic density when $\hat{H}\psi = E\psi$. Making use of eqn.[2.56], eqn.[2.54] becomes

$$\delta\mathcal{G}[\psi,\Omega] \;=\; \oint \left(\frac{\hbar^2}{4m}\right) \left\{ 2\nabla\psi^* \cdot \mathrm{n}(r)\delta\psi + \delta S(\Omega,r)\nabla^2\rho(r) \right\} dS(\Omega,r) + cc \quad . \tag{2.57}$$

This expression is made operational by introducing a constraint which which will eliminate the term represented by the variation of the surface. Thus, one considers the entire variation to be carried out using a trial function ϕ, which, at the point of the variation reduces to the wavefunction ψ. One also imposes at all stages of the variation procedure the constraint that the region $\Omega(\phi)$, defined in terms of the trial function ϕ, be bounded by a surface of zero-flux such that

$$\nabla\rho_\phi \cdot n(r) = 0 \quad \forall r \in S(\Omega, r) \text{ and } \delta S = 0 \quad . \tag{2.58}$$

where ρ_ϕ is the trial density. The region $\Omega(\phi)$ represents the open system in the varied total system described by the trial function ϕ just as $\Omega(\psi)$ represents the open system in the state described by ψ. Requiring the fulfillment of eqn.[2.58] amounts to imposing the variational constraint that the divergence of $\nabla\rho_\phi$, $\nabla \cdot \nabla\rho_\phi$, integrates to zero at all stages of the variation, i.e.

$$\int_\Omega \nabla^2\rho(r)d\tau = 0 \tag{2.59}$$

for all admissible ϕ, which in turn implies that

$$\delta\left\{\int_\Omega \nabla^2\rho_\phi(r)\ d\tau\right\} = 0 \quad . \tag{2.60}$$

The variation denoted in eqn.[2.60] requires a variation of the surface as well as a variation of the integrand, as in eqn.[2.49] Since the sum of these two terms must vanish as a result of the constraint, one is able to equate the integral of the variation in the surface to the negative of the variation of the integrand to obtain

$$\left(\frac{\hbar^2}{4m}\right)\oint \delta S(\Omega, r)\nabla^2\rho(r)\ dS(\Omega, r) = -\left(\frac{\hbar^2}{4m}\right)\int_\Omega \delta\left\{\nabla^2\rho(r)\right\}\ d\tau \quad . \tag{2.61}$$

The term on the R.H.S. of eqn.[2.61] can be used to substitute the term involving the variation of the surface in eqn.[2.57]

$$\begin{aligned}\int_\Omega \delta\left\{\nabla^2\rho(r)\right\}\ d\tau &= \int_\Omega \nabla\cdot\left\{\nabla\psi^*\delta\psi + \psi^*\delta\nabla\psi\right\}\ d\tau \\ &= \oint \left\{\nabla\psi^*\delta\psi + \psi^*\delta\nabla\psi\right\}\cdot \mathrm{n}(r)\ dS(\Omega, r) \quad ,\end{aligned} \tag{2.62}$$

where Gauss' theorem is used to transform the volume integral into a surface integral. Making use of relationships (2.61) and (2.62) and varying ψ on the boundary, as well as varying the boundary itself, the variation of the functional becomes

$$\delta\mathcal{G}[\psi, \Omega] = \left(\frac{\hbar^2}{4m}\right)\oint \left\{\nabla\psi^*\delta\psi - \psi^*\delta\nabla\psi\right\}\cdot \mathrm{n}(r)\ dS(\Omega, r) \quad . \tag{2.63}$$

The variation of Schrödinger energy functional for an open system is now expressible as the surface integral of the current density

$$\mathbf{j}(r) = \frac{\hbar}{2mi}\{\psi^*\nabla\psi - (\nabla\psi^*)\psi\} \tag{2.64}$$

over this system's surface. The variation of this operator with respect to ψ is

$$\delta\mathbf{j}(r) = \frac{\hbar}{2mi}\{\psi^*\delta(\nabla\psi) - (\nabla\psi^*)\delta\psi\} \quad , \tag{2.65}$$

allowing one to re-express equation (2.63) in terms of the variation in the current to yield

$$\delta\mathcal{G}[\psi,\Omega] = -\left(\frac{i\hbar}{2}\right)\oint \delta\mathbf{j}(r)\cdot\mathbf{n}(r)\; dS(\Omega,r) + cc \quad . \tag{2.66}$$

The expression obtained is the result of retaining the variation of ψ on the open system's boundary while incurring a variation of he boundary itself.

It is at this point that Schwinger's action principle (Schwinger 1951) becomes pivotal in continuing the derivation at hand. This principle identifies the variations in the state vector of eqn.[2.65] with the generators of infinitesimal unitary transformations. These transformations, acting separately on the state vector and its observables, can be used to to generate all the possible changes in the dynamic variables of a quantum system. This methodology for the transformations is adopted by the theory of AIM as follows:

The operator for an infinitesimal transformation and its Hermitian conjugate are :

$$\hat{U} = \hat{1} - \frac{i\epsilon}{\hbar}\hat{G} \quad \text{and} \quad \hat{U}^{-1} = \hat{1} + \frac{i\epsilon}{\hbar}\hat{G} \quad , \tag{2.67}$$

where ϵ is an infinitesimal constant and $\hat{G}$ is the generator operator of the transformation. First order variations on the state vector and its complex conjugate can then be analytically expressed as

$$\delta\psi = -\left(\frac{i\epsilon}{\hbar}\right)\hat{G}\psi \quad \text{and} \quad \delta\psi^* = \left(\frac{i\epsilon}{\hbar}\right)\hat{G}\psi^* \quad . \tag{2.68}$$

One need only replace the expressions for the variation of the state vector on eqn.(2.68) to obtain

$$\delta\mathcal{G}[\psi,\Omega] = -\left(\frac{\epsilon}{2}\right)\left\{\oint \mathbf{j}_G\cdot\mathbf{n}(r)\; dS(\Omega,r) + cc\right\} \quad , \tag{2.69}$$

where $\mathbf{j}_G$ is the current density operator for the property G and has already been introduced in section 2.4. The variation of Schrödinger's energy functional has reached a form full of physical significance. When derived for an open system bounded by a surface of zero-flux, the energy functional is proportional to the surface flux of the current density of the infinitesimal generator causing the

change in the system. Making use of eqn.[2.30] one can substitute the surface term on the R.H.S. of eqn.[2.69] to obtain

$$\delta\mathcal{G}[\psi,\Omega] = -\left(\frac{\epsilon}{2}\right)\left\{\frac{i}{\hbar}\langle\psi|[\hat{H},\hat{G}(r)]|\psi\rangle_\Omega + cc\right\} \quad . \tag{2.70}$$

$\hat{H}$ is the many-electron Hamiltonian operator for the complete system, while $\hat{G}$, the generator, is a function whose coordinates are integrated over the open system Ω. This function provides the property density of the generator. Thus, equation [2.70] is of paramount importance. We now have a variational determination of the hypervirial theorem that applies to any system bounded by a zero-flux surface, which includes a total system or any topologically defined open system. This theorem provides the foundations for the physics of a total system or for an atom in a molecule.

2.7 Atomic Theorems

The variation of the energy functional for an open system bounded by surfaces of zero-flux in the gradient of the electronic density is proportional to the flux of the current density associated with the infinitesimal generator causing a change in the system. Expressions [2.69] and [2.70] demonstrate that the average value of the commutator is equated to and balanced by the negative of the flux of the corresponding current through the surface of the open system. One need only set $\hat{G}$ to the operator of interest and obtain the corresponding *atomic theorem*. These theorems define the quantum mechanics of an atom in a molecule, even in the presence of electric or magnetic fields (Bader and Popelier 1993). Of these, we will present two of great importance.

2.7.1 Atomic Force Theorem

Setting the generator $G(r)$ equal to $\hat{\mathbf{p}}$, the momentum operator, yields the time rate of change of momentum or the atomic *force* theorem, denoted by $F(\Omega)$, the Ehrenfest force acting on an atom. For a stationary state, the basin and surface forces balance each other, i.e. the Ehrenfest force exerted on the basin equals the pressure exerted on every element of its surface. Considering a

multi-electronic system, and because of eqn. [2.69] and eqn.[2.70]this force is averaged as

$$\begin{aligned} F(\Omega) &= \frac{N}{2}\left\{\frac{i}{\hbar}\langle\psi|[\hat{H},\hat{\mathbf{p}}]|\psi\rangle_\Omega + cc\right\} \\ &= -\oint dS(\Omega,r)\quad \mathbf{j}_p(r)\cdot\mathbf{n}(r) + cc \\ &= N\int_\Omega d\tau \int d\tau'\ \psi^*(-\nabla\hat{V})\psi \\ &= -\oint dS(\Omega,r)\ \overleftrightarrow{\sigma}\ (r)\cdot\mathbf{n}(r) \quad , \end{aligned} \tag{2.71}$$

where $\overleftrightarrow{\sigma}$ is the stress tensor, a symmetric dyad with dimensions of pressure. The stress tensor relates directly to the wavefunction according to the expression:

$$\overleftrightarrow{\sigma}\ (r) = \frac{N\hbar^2}{4m}\int\{\nabla(\nabla\psi^*)\psi + \psi^*\nabla\nabla\psi - \nabla\psi^*\nabla\psi - \nabla\psi\nabla\psi^*\}\ d\tau' \quad . \tag{2.72}$$

2.7.2 Atomic Virial Theorem

Another important theorem is derived from setting the generator equal to the operator $\hat{\mathbf{r}}\cdot\hat{\mathbf{p}}$. This generator yields the atomic statement of the virial theorem according to eqn.[2.69] and eqn.[2.70]

$$\frac{N}{2}\left\{\frac{i}{\hbar}\langle\psi|[\hat{H},\hat{\mathbf{r}}\cdot\hat{\mathbf{p}}]|\psi\rangle_\Omega + cc\right\} = -\frac{N}{2}\left\{\oint dS(\Omega,r)\mathbf{j}_{\mathbf{r}\cdot\mathbf{p}}(r)\cdot\mathbf{n}(r) + cc\right\} \quad , \tag{2.73}$$

where $\hat{\mathbf{r}}$ is referencing the nucleus. Expanding the L.H.S. of eqn. [2.73] one obtains

$$\begin{aligned} \frac{N}{2}\left\{\frac{i}{\hbar}\langle\psi|[\hat{H},\hat{\mathbf{r}}\cdot\hat{\mathbf{p}}]|\psi\rangle_\Omega + cc\right\} &= 2N\int_\Omega d\tau\int d\tau'\ \frac{-\hbar^2}{4m}\{\psi^*\nabla^2\psi + (\nabla^2\psi^*)\psi\} \\ &\quad +N\int_\Omega d\tau\int d\tau'\ \psi^*(-\mathbf{r}\cdot\nabla\hat{V})\psi \qquad (2.74) \\ &= 2T(\Omega) + \mathcal{V}_b(\Omega) \quad . \qquad (2.75) \end{aligned}$$

$\mathcal{V}_b(\Omega)$ is the integrated value of the virial of the Ehrenfest force acting on an electron in the basin of the atom. $T(\Omega)$ is the integrated value for the kinetic energy over the basin. Similarly, expanding the R.H.S. of eqn.[2.73] we obtain

$$\begin{aligned} \frac{1}{2}\left\{\oint dS(\Omega,r)\mathbf{j}_{\mathbf{r}\cdot\mathbf{p}}(r)\cdot\mathbf{n}(r) + cc\right\} &= -N\frac{\hbar^2}{4m}\{\oint dS(\Omega,r)\int d\tau'[\psi^*\nabla(\mathbf{r}\cdot\nabla\psi) \\ &\quad -\nabla\psi^*(\mathbf{r}\cdot\nabla\psi) + \psi\nabla(\mathbf{r}\cdot\nabla\psi^*) \\ &\quad -\nabla\psi(\mathbf{r}\cdot\nabla\psi^*)]\cdot\mathbf{n}(r)\} \\ &= -\oint dS(\Omega,r)\mathbf{r}\cdot\ \overleftrightarrow{\sigma}\ (r)\cdot\mathbf{n}(r) \\ &\quad -\frac{\hbar^2}{4m}\oint dS(\Omega,r)\nabla\rho(r)\cdot\mathbf{n}(r) \\ &= -\mathcal{V}_s(\Omega) \quad , \end{aligned} \tag{2.76}$$

where $\mathcal{V}_s(\Omega)$ is the virial of the surface enclosing the basin.

Equating the results obtained from expanding the R.H.S. of with those obtained from expanding the L.H.S. of eqn. [2.73] yields

$$-2T(\Omega) = \mathcal{V}_b(\Omega) + \mathcal{V}_s(\Omega) = \mathcal{V}(\Omega) \quad , \tag{2.77}$$

where the virial $\mathcal{V}(\Omega)$ is the sum of the basin and surface contributions.

For systems at electrostatic equilibrium (one which experiences no forces exerted on the nuclei), the virial equals the average potential energy of a molecule, $\mathcal{V} = V$. Under this condition one has

$$-2T(\Omega) = V(\Omega) \tag{2.78}$$

where $V(\Omega)$ is the potential energy of atom Ω. By adding $T(\Omega)$ to each side of eqn.[2.78] one obtains

$$-T(\Omega) = T(\Omega) + V(\Omega) = E(\Omega) \tag{2.79}$$

where $E(\Omega)$ is the total energy of the atom Ω, and the sum of the atomic contributions to the energy equals E, the total energy, just as postulated in eqn.[2.17], on the basis of the original observations on the paralleling transferability of the electron density and the kinetic energy densities.

2.8 Summary of AIM classification tools

2.8.1 Bond properties.

Having determined the molecular graph, the network of bond critical points and the bond paths which characterizes the interactions between nuclei, one can proceed with the classification of the bond critical points through an analysis of the properties of ρ at the location of the critical point. Those most relevant for this work are summarized below:

- *Bond Path Distances* r_A, r_B, denote the distance between a bond critical point and the bonded atoms A and B respectively.
- *The Electron Density at the critical point* ρ_b. It provides a measure of bond order for bonds of similar atoms.
- *The Ellipticity* $\epsilon(r_c)$. It provides a measure of the accumulation of the electron density in a given plane containing the bond critical point. This plane is determined by the eigenvectors

associated with the negative curvatures at the critical point. The extent of such accumulation is calculated according to

$$\epsilon = \frac{\lambda_1}{\lambda_2} - 1 \quad , \tag{2.80}$$

where

$$\lambda_1 \leq \lambda_2 < 0 \quad . \tag{2.81}$$

λ_1 is sometimes called the "hard" curvature whereas λ_2 is referred to as the "soft" one.

- *The Laplacian* $\nabla^2\rho_b$ is the sum of the three principal curvatures at the bond critical point. It provides a local measure of charge concentration or depletion. When used in combination with ρ_b at a bond critical point it becomes a diagnostic tool of the ionicity/covalency nature of such bond.
  - If $\nabla^2\rho_b < 0$ at r_c and ρ_b is large, the electron density is strongly concentrated along the bond path. This type of situation results in the sharing of electron density between the bonded atoms, and the interaction is characterized as *shared*.
  - If $\nabla^2\rho_b > 0$ and ρ_b is low at r_c, the electron density is locally depleted at the bond critical point and in the region of the interatomic surface, and accumulated in the basins of one or both of the interacting atoms. This is all a consequence of the effect of the Pauli exclusion principle on the approach of two closed shell atoms. This type of situation is typical of *closed shell* interactions.
- *The local Virial Theorem* The local expression of the virial theorem can be expressed as

$$\left(\frac{\hbar^2}{4m}\right)\nabla^2\rho(\mathbf{r}) = 2G(\mathbf{r}) + \mathcal{V}(\mathbf{r}), \tag{2.82}$$

where $\mathcal{V}(\mathbf{r})$ is the virial field

$$\mathcal{V}(\mathbf{r}) = -\mathbf{r}\cdot\nabla\cdot\overleftrightarrow{\sigma} + \nabla\cdot(\mathbf{r}\cdot\overleftrightarrow{\sigma}) \quad . \tag{2.83}$$

Since $G(r) > 0$ and $\mathcal{V}(r) < 0$, the sign of the Laplacian of the electron density $\nabla^2\rho$ determines which of these two contributions to the energy are dominant at a given point in space. When $\nabla^2\rho$ is negative, for example, the virial field $\mathcal{V}(r)$ which determined the electronic potential energy exceeds twice the kinetic energy density.

2.8.2 Atomic Properties.

The average of a property over an atomic basin is computed according to

$$A(\Omega) \equiv \langle \hat{A} \rangle_\Omega = \int_\Omega d\tau \int d\tau' \left(\frac{N}{2}\right) \left\{ \psi^* \hat{A} \psi + (\hat{A}\psi)^* \psi \right\} \quad . \tag{2.84}$$

Although some of the ensuing properties have been described in greater detail in this chapter, it is convenient to list them together in summary form for easy reference as they will be used in the rest of this thesis. In this work we will make use of the following properties:

- *Electron Population.* It is obtained by setting $\hat{A}$ in eqn. [2.84] to the unity operator $\hat{1}$. Averaging this operator yields the electronic charge over the basin Ω,

$$N(\Omega) = \int_\Omega \rho(r)\, d\tau \quad . \tag{2.85}$$

- *Net Charge.* Computed by subtracting the population $N(\Omega)$ from the nuclear charge Z_Ω,

$$q(\Omega) = Z_\Omega - N(\Omega) \quad . \tag{2.86}$$

- *Atomic Volume.* The space enclosed by the intersection of the atom's surfaces of zero-flux and a chosen isodensity envelope. Unless otherwise stated, this thesis uses $\rho = 0.001$ *au* as a default.

- *Schrödinger Kinetic Energy.* Schrödinger's formulation

$$K(\Omega) = \frac{-\hbar^2}{4m} N \int_\Omega d\tau \int d\tau' \left\{ \psi^* \nabla^2 \psi + \psi \nabla^2 \psi^* \right\}. \tag{2.87}$$

- *Gradient Kinetic Energy.* The other possible formulation, in its positive definite form, of the kinetic energy density

$$G(\Omega) = \frac{\hbar^2}{2m} N \int_\Omega d\tau \int d\tau' \left\{ \nabla\psi^* \cdot \nabla\psi \right\} \quad . \tag{2.88}$$

- *Laplacian.* Is used as an indicator of the accuracy of the numerical atomic integrations. This quantity is readiiy calculated after having computed both kinetic energy densities since

$$\begin{aligned} L(\Omega) &= K(\Omega) - G(\Omega) \\ &= -\frac{\hbar}{4m} \int_\Omega d\tau \nabla^2 \rho(r) \\ &= -\frac{\hbar}{4m} \oint_d S(\Omega, r) \nabla\rho(r) \cdot n(r) \\ &= 0 \quad . \end{aligned} \tag{2.89}$$

The greater the deviation of $L(\Omega)$ from zero, the greater the error incurred in the numerical integration procedure of Ω.

- *Total Atomic Energy.* As shown above in eqn.[2.78], for a system in electrostatic equilibrium, the energy of an atom in a molecule is given by

$$E(\Omega) = -T(\Omega) = \frac{1}{2}V(\Omega) \quad , \tag{2.90}$$

where $T(\Omega)$ is the average kinetic energy of the atom, and $V(\Omega)$ its average potential energy. Unfortunately, generating wavefunctions which comply with the virial theorem within a good accuracy - a deviation of 10^{-8} - is computationally expensive. Generally, electronic packages produce wavefunctions whose virial ratios start deviating from 2 at the fourth significant figure. Because the deviation is so small, one can correct for this according to the expression

$$E(\Omega) = K(\Omega)\left(1 - \frac{V(\Omega)}{K(\Omega)}\right) \quad . \tag{2.91}$$

- *First Moment.* The first moment, or dipole, is denoted by $\mu(\Omega)$ and is the atomic average of the electronic position vector $r_\Omega = r - X_\Omega$, where $r(\Omega)$ is referenced to the nucleus of atom Ω, and X_Ω is the position vector of the nucleus in the Ω in reference to a given arbitrarily chosen origin. The first moment is thus a vector, which in Cartesian space has the components

$$\mu_\alpha(\Omega) = -e\int_\Omega \alpha_\Omega\ \rho(r)\ d\tau; \quad \forall\ \alpha = x, y, z \quad . \tag{2.92}$$

Its physical significance lies in its quantifying the extent of distortion of the electronic density from sphericity; its polarization. The extent of the distortion is given by the magnitude of this vector. The dipole moment of a neutral molecule is calculated as

$$\mu = \sum_\Omega \{Z_\Omega X_\Omega + \mu(\Omega)\} \quad . \tag{2.93}$$

It is to be noted that although the product $Z_\Omega X_\Omega$ is origin dependent, the sum of the contributions over the whole system ceases this dependence.

- *Quadrupole Moment Tensor.* A symmetric traceless tensor with components

$$Q_{\alpha\beta}(\Omega) = -e\int_\Omega \left(3\alpha\beta - r_\Omega^2\delta_{\alpha\beta}\right)\rho(r)\ d\tau; \quad \forall\ \alpha, \beta = x, y, z \quad , \tag{2.94}$$

where the electronic coordinates are also referenced to the nucleus of the atom, as described for the dipole above. These elements reflect the degree of charge accumulation/depletion along

a given axis with respect to the perpendicular plane at a given point in space. Upon diagonalization, the diagonal elements provide the extent of polarization. If $Q_{zz}(\Omega) = Q_{yy}(\Omega) = Q_{zz}(\Omega) = 0$ the atom shows a spherical distribution. If, for example $Q_{zz}(\Omega) > 0$ then, and because the trace of a real symmetric matrix does not change due to unitary transformations, $Q_{yy}(\Omega) = Q_{xx}(\Omega) = -\frac{1}{2}\, Q_{zz}(\Omega)$. This results in an accumulation of charge in the xy plane with respect to the z coordinate axis.

Chapter 3

Transferable Functional Groups

The preceding chapter has shown that all chemical information about an atom in a molecule is contained in the molecule's wavefunction, and obtained with the electron density as its vehicle. The theory of AIM shows that the properties of an atom change in direct response to changes in its electron density distribution. As a result of this, an atom is expected to show varying values for its properties depending on the perturbations induced on it by its environment. A *functional group* is a linked set of atoms with a specific chemical identity and functionality. By extension, a *functional group* will similarly see its properties change according to the changes in the atoms it constitutes it. The amount of the change exhibited is a gauge of transfer. This change needs to be quantified and characterized on the basis of its chemical significance. This chapter is devoted to the demonstration that the theory of AIM recovers and explains the experimentally observed additivity of group properties and, in so doing, lays the theoretical foundation for the concept of *transfer* of a theoretical group.

Section 3.1 defines a quantum functional group. It is crucial that such concept be defined within the framework of quantum mechanics for only then the study of transferability can take place. Section 3.2 provides and discusses the means for quantifying the transfer of a functional group. This quantification will be shown to parallel the observation of chemistry in the laboratory; certain groups of atoms show the ability of preserving their properties and form despite their being components to distinctly different chemical systems. This chapter is concluded with a classification of the different types of transferability possible, section 3.3,

3.1 Functional Group: Definition as an Open System

To proceed further it is critical that the definition of a *functional group* within quantum mechanics be established. Such a definition builds from the following points:

1. All information related to the partitioning of a molecule into subsystems must be contained in the molecular wavefunction.
2. All the properties of the subsystems in point (1) must be defined.
3. The properties of all subsystems in a molecule must be additive to yield the molecular total for these properties.

The preceding chapter discussed the quantum mechanical framework which accounts for all of the conditions given above. In it we defined an open system as a region of real space, a domain of the scalar field of the electronic density and encapsulated surfaces of zero-flux. Most important, all of the atomic properties are defined by the physics of an open system. The atoms of theory are identified with the atoms of chemistry because they exhibit the following two important and necessary characteristics:

- If the electron density distribution of an atom is identical in two different molecules, then the atom makes the same contributions to all the properties, including the energy. It is because of this observation of properties following form that we can identify the atom to be one and the same in both systems.
- The atoms of theory are the most transferable pieces of a chemical system defined in real space, and thus maximize the transfer of chemical information between molecules using the electron density as the vehicle.

Even in the presence of external electric or magnetic fields one still observes the paralleling transferability of the atom's density distribution and its properties, including the field induced properties such as electronic polarizability and magnetic susceptibility.

A *functional group* is thus a grouping of linked atoms which comply with the two points indicated above and whose properties are determined by the physics of an open system.

3.2 The Quantification of Transfer

Functional groups make additive contributions to all the properties of the system in which they occur. Their form in real space is determined by the distribution of electron density. If this distribution is undisturbed upon transfer from one system to another then the functional group is transferable.

It is to be understood that perfect transferability is unattainable as dictated by the fundamental tenet of *density functional theory (DFT)*; the electron distribution of a total system is a unique functional of $\rho_D(r)$, the density contained in some domain D of the system, and fixed by this system's external potential $V(r)$. If $\rho_D(r)$ is identical over corresponding domains in two distinct systems, these systems are necessarily one and the same, negating the possibility for perfect transferability of a functional group and its properties. No restriction is placed on how closely this limit can be met however. From an operational point of view two functional groupings of atoms may appear to be identical to within computational and experimental accuracy. This is the context in which we define transferability; the observation that a functional group may show its set of properties and form deviating by less than the value of the experimental error associated with the measurement of the corresponding property.

The extent of transfer of an open system is assessed in terms of four inter-related measurables, namely the study of: (a) the geometry of the functional group, (b) the properties of its bond critical points, (c) the functional group's set of properties and (d) the properties of the group's surfaces and form. Such study can only be performed on systems which have been geometrically optimized at a high level of theory. These points are discussed in turn below.

3.2.1 Transfer of *Geometrical Parameters*:

The most obvious manifestation of transfer exhibited by a group is in its preservation of bond distances and angles. The smaller the deviation in these parameters the greater the possibilities for transfer. We illustrate this with previously conducted research by our group. Bader and Chang (Chang and Bader 1992) undertook a preliminary study of the transferability of a series of amino acid residues in small polypeptides such as formylglycine, formylglycinamide, glycylglycine, glycylglycylglycine, alaninamide, formilalanine, alanylglycine and glycylalanine at their optimal linear configuration. They found an average deviation of 0.001 Å and 0.17^0 of the corresponding set of bond lengths and angles in these molecules, with maximum deviations of 0.004 Å and 0.5^0 respectively. The extent of transfer of bond lengths was further illustrated with the prediction of the amide bond length to be recovered upon linking the complementary $-C(=O)|$ and $|NH-$ groups

upon superimposing their interatomic surfaces, denoted by "|". The resulting length is given by the sum of the bonded radii of the N and C atoms, $r_b(N)$ and $r_b(C)$, the distance from a nucleus to the bond critical point. The error in the predicted lengths in the synthesized planar di- and tri-peptides was 0.002 Å or less. The small geometrical deviations occurring in these systems, are so small that predictions can be made on these basis, clearly indicates possibilities for transfer of amino acid residues.

3.2.2 Transfer of *Bond Critical Point Data*:

The properties of a bond critical point located at r_c are clearly summarized in terms of the value of its density, $\rho(r_c)$, its Laplacian $\nabla^2\rho(r_c)$, and its ellipticity ϵ. Bader and Chang's study demonstrated that these values changed by such small amounts for a specific bond in a given group throughout the series of amino acid residues in their study that the authors characterized these groups as highly transferable. Subsequent studies (Popelier and Bader 1994) quantified small deviations in the properties of the similar bonds in $HC(=O)|Gly|Gly|Gly|NH_2$, a molecule whose geometry was optimized at a perfect $\alpha-$R helix configuration. As an example, the upper and lower bounds in the deviations for the value of ρ_b lie between 0.015 au, as corresponding to $C_\alpha - N$, and $C_\alpha - C$ with an associated value of 0.004 au. These decreases also corresponded with decreases in bond lengths, highlighting the relationship between geometrical parameters and bond critical point data. The other bond critical point properties showed paralleling changes upon transfer as well.

3.2.3 Transfer of *Basin Properties* and *Form*:

The atomic contributions to the energy, volume and multipole moments exhibited small changes within the set of molecules in Bader and Chang's study. The average deviation in the electron population was found to be 0.005 e in the corresponding atoms. The atomic energies, which for the case of C and O possess values of 23×10^3 kcal/mol and 47×10^3 kcal/mol respectively, varied by less than 3 kcal/mol (0.01%) upon transfer. The atoms which showed the greatest variation were the terminal C and N, as it would be expected. This is so since these atoms feel the immediate effects of a new amino acid residue as their neighbor. The rest of the atoms comprising the amino acid residue varied little in their contributions to the molecular energy, an indication of "electron density buffering" effects. It was also found that their volumes and form, as defined by their 0.001 au isodensity envelopes, showed little change upon transfer. Even the distribution of their local contributions to the scalar field of $\nabla^2\rho$, a highly sensitive measure for it makes use of second

derivative information of the density distribution, showed little variance.

This thesis presents a study of all the amino acid residues in their optimal α-helical conformation.

3.2.4 Transfer of *Surface Properties* and *Form*:

Although not explicitly demonstrated in the work of Bader and Chang, the extent of surface transfer was pictorially illustrated by overlapping contour maps in the electronic density of several amino acid residues. These maps were computed at the plane containing these amino acid residues' backbone atoms. The proteins were optimized at their β-sheet conformation. This pictorial approach served for the purposes of highlighting the high potential for surface transfer, easily illustrated by partitioning the two amide IAS to define the $H_2N|$, $C(=O)CH_2NH|$ and $|CHO$ groups.

This thesis extends the evaluation of surface transferability to a more quantitative level. To achieve this both properties and form of the zero-flux surfaces are to be computed. On this account, two surfaces will be considered transferable if the difference in their respective associated properties and form fall below a reasonable threshold. The difference in their forms will be quantified by a procedure which gives a measure of the extent to which surfaces superimpose onto one another. This method is discussed in detail in section 4.5.

3.3 Types of Transferability

The awareness of the existence of transferability in chemistry is not a recent phenomenon. The observation that variations in group properties through a series of molecules can be so slight as to enable one to establish a group additivity scheme played has played an important role in defining the concept of functional groups in chemistry. AIM contributes to the study of transferability by quantifying and comparing the contributions that functional groups make to different systems.

Our extensive past and present work in this area indicates that there is a predominant form of transferability, defined as *compensatory* transferability, and a lesser encountered *near-perfect* transferability. Both forms of transfer are discussed below in greater detail.

3.3.1 *Compensatory* Transferability

Compensatory Transferability is the the most commonly found exhibition of transfer. It occurs when the change occasioned in the properties of one group is offset by an equal and opposite change in the

properties of its new neighbors. Such behavior is exemplified below with the set of molecules of the form ACH_2CH_2B with $A, B = H, CH_3, NH_2, OH$ and F (Bader and Martín 1998). These system's geometries were optimized at a high level of theory HF/6-311++G(3df,2p) with a scaling of the electronic coordinates so as to satisfy $|2T/V|$ to an accuracy of 10^{-8} au. The satisfaction of the virial theorem for the complete system is only compromised by the accuracy of the atomic integrations, which upon summing over all the atomic contributions yielded errors in the total energies of less than 1.0 kcal/mol. This exercise revealed that the molecular energies of the hetero-molecular systems were the arithmetic mean of the corresponding homo-molecular counterparts, ACH_2CH_2A and BCH_2CH_2B, to within a few kcal/mol. Paralleling our results, experimental analogs with $A = H$ in the measured heats of formation of CH_3CH_2B have been reported in the literature (Benson *et al.* 1969) and are found to differ from the arithmetic mean of the heats of formation of ethane and BCH_2CH_2B by + 0.32, -0.53 and -0.17 kcal/mol respectively for $B = OH, Cl$ and I.

Table 3.1: Energy Additivity and properties of $HCH_2 - CH_2B$. Energy changes are in $kcal/mol$, and electron densities deviations in au.

B	ΔE	$\Delta E(HCH_2\|)$	$\Delta E(\|CH_2B)$	$q(HCH_2\|)$	ρ	$\Delta\rho$
CH_3	-0.05	-10.8	+10.6	-0.017	0.2588	0.0001
NH_2	-0.09	-8.6	+8.1	+0.008	0.2640	0.0025
OH	-0.33	-2.6	+1.9	+0.047	0.2692	0.0020
F	-1.53	-0.3	-1.4	+0.066	0.2727	0.0028

Our study demonstrated that the groups $ACH_2|$ and $BCH_2|$, where | denotes the location of the $C - C$ interatomic surface, undergo small changes in forming the heteromolecule, exhibiting a charge transfer between the groups of 0.07 e (table 3.1). The electron population is necessarily conserved in forming the hetero-molecule, i.e. $q(ACH_2|) = -q(|CH_2H)$ .The surprising fact is the extent to which other properties such as the energy are also conserved. This additivity scheme is an example of compensatory transferability, the small change in the properties of one group induced by the transfer is nullified by a compensatory change in its new neighbor.

The changes in energy

$$\Delta E = E(HCH_2CH_2B) - \frac{1}{2}\{E(HCH_2CH_2H) + E(BCH_2CH_2B)\} \tag{3.1}$$

in table (3.1) are produced from calculating the total SCF energies of the molecules, while the changes in the group energies are obtained from differences of the appropriately summed atomic

energies obtained from numerical integrations. The small differences between the values of ΔE and the sum of group energy changes, with a maximum value of 0.4 kcal/mol, are a measure of the total error in the integration of the atomic contributions to the energy. The changes to the group energies, $\Delta E(|CH_2B)$, decrease following the trend $B = CH_3$, despite the increase in the transfer of electronic charge through the series. This is so since the inclusion of more electronegative atoms in the groups withdraws electronic density away from the $C - C$ region of the molecule, and thus highlighting the occurrence of compensatory behavior; the changes in the Energy is found within each group and dominates the change for $B - F$. In the later case, the near zero-changes in $\Delta E(HCH_2|)$ and $\Delta E(|CH_2F)$ are the result of nearly equal and opposite changes in the energies of the carbon and hydrogen atoms in each group according to the amounts ± 22 and ± 13 kcal/mol respectively, with the carbon being stabilized in methyl and destabilized in CH_2F.

Table 3.2: Surface properties of $C - C$ bond in $HCH_2 - CH_2B$. All quantities are in au, save for $\Delta\mathcal{V}_S$ which is in kcal/mol.

B	$\mathcal{V}_S(HCH_2\|)$	$\Delta\mathcal{V}_S(H)$	$\mathcal{V}_S(\|CH_2B)$	$\Delta\mathcal{V}_S(B)$	$\Delta\mathcal{V}_S$	$\mathcal{N}_S$	$\Delta\mathcal{N}_S$	$\mathcal{G}_S$	$\Delta\mathcal{G}_S$
CH_3	-0.2063	-0.0279	-0.1690	+0.0267	-0.8	1.2193	0.0003	0.8203	-0.0001
NH_2	-0.2040	-0.0256	-0.1743	+0.0314	+3.6	1.2224	0.0095	0.8322	+0.0104
OH	-0.1906	-0.0122	-0.1852	+0.0065	-3.6	1.2181	0.0110	0.8380	+0.0062
F	-0.1836	-0.0053	-0.1949	-0.0032	-5.3	1.2102	0.0191	0.8348	+0.0130

Because of the interdeterminancy of a group's basin and its bounding surfaces of zero-flux denoted by eqn. [2.30], the compensatory behavior measured in computing the fragment's properties gets mirrored by a paralleling compensatory behavior in this fragment's surface properties. The surface energy of $HCH_2|$, as referenced to C, decreases by -0.0279 au (table 3.2), from its value in ethane upon transfer to propane, while the surface energy of $CH_3CH_2|$, referenced to the C of methylene, undergoes an increase of $+0.0267$ au relative to its value in butane. These compensatory changes result in a new surface whose energy $\mathcal{V}_S$ obtained by summing the two group contributions, differs by only -0.8 $kcal/mol$ from the mean of the two original surfaces, the quantity $\Delta\mathcal{V}_S$ in table 3.2. Other surface properties change in a similarly compensatory manner to yield small displacements from their mean values. Integrating $\rho(r)$ over the surface yields $\mathcal{N}_S$, the average number of electrons per unit length, and the integration of the gradient kinetic energy density $G(r)$ yields $\mathcal{G}_S$, the average kinetic energy per unit length. These properties show small changes from their mean values as well.

The data in Table(3.2) indicate that similar changes are found for the remaining molecules in this series. This suggests that a "Le Chatelier-like" principle underlies the observation of characteristic properties for functional groups, each group changing in response to the presence of another in such a way as to minimize the overall change.

3.3.2 *Near-Perfect* Transferability

One of the earliest and most extensive studies which presents the idea of transferability of properties in chemistry was provided by the measurement of heats of formation of cyclic and acyclic alkanes (Rossini 1946). The heats of formation of the hydrocarbons of the type $CH_3(CH_2)_nCH_3$ were shown to fit the linear relationship

$$\Delta H_f^\circ(298) = 2A + nB \quad ,$$

where A and B are the contributions from the methyl and the methylene groups respectively. The same additivity is recovered computationally, as provided by the SCF single determinant energies $E(N)$ which recover the relationship

$$E(N) = 2E(CH_3) + nE(CH_2) \quad . \tag{3.2}$$

Equation (3.2) is obeyed to within 10^{-4} *au* (0.06 kcal/mol) for the molecules ethane to heptane. The electron populations of these groups showed a small transfer of charge from a CH_2 group to its bonded CH_3, causing the methyl groups to acquire a negative charge in all its members from ethane onwards. The magnitude of this change is constant to within the integration error of 0.001 e, and the electron population of each such methyl group is increased by a constant amount, ΔN. One finds that the methyl group in these molecules, while differing slightly in ethane, exhibits a constant set of properties. One also finds that the charge transferred to a methyl group is taken from its bonded methylene group. Thus, the methylene groups bonded to methyl groups possess net positive charges, equal to $+2\Delta N$ in propane and to $+\Delta N$ in the remainder of the series. What accounts for the additivity of the energy in spite of these small changes in the methylene groups resulting from their changed environments is that the energy, as well as the charge, is conserved during transfer, an example of compensatory transferability. Thus, the methylene group in propane, which loses $2\Delta Ne$ of charge is destabilized by $-2\Delta E$ and has an energy equal to $E(CH_2) - 2\Delta E$. A methylene group which is bonded to a single methyl and loses ΔNe of charge is destabilized by $-\Delta E$ and has an energy equal to $E(CH_2) - \Delta E$. Thus $|\Delta E|$, which determines the increase in stability of a methyl group in a molecule relative to $E(CH_3)$, the energy of methyl in ethane, also determines

the decrease in the stability of a methylene group bonded to a single methyl relative to the energy $E(CH_2)$ in eqn. [3.2]. Thus, the central methylene group in pentane, hexane and heptane, which is bonded to other methylenes and bear no net charge, should possess energies equal to the increment $E(CH_2)$ in eqn. [3.2], as demonstrated by Bader and co-workers (Bader *et al.* 1992*b*). With these compensatory changes in energy as well as charge between the methyl and the methylene groups, the resulting groups exhibit perfect transferability. Thus all the methyl groups past ethane exhibit identical properties as do all methylene groups bonded to a single methyl, as do the methylene groups bonded to other methylenes.

Equally transferable are these fragment's volumes. It has been shown that the 0.002 *au* isodensity envelope provides a good measure of shape of a molecule in a crystal, while the slightly larger isodensity envelope of 0.001 *au* reproduces the equilibrium diameters of gas phase molecules as determined by viscosity data (Bader *et al.* 1987). In practical terms, the volume of a fragment is the amount of space enclosed by the fragment's surfaces of zero-flux and a chosen isodensity envelope. Bader *et al.* showed that the hydrocarbons under study exhibited the same degree of transferability as the one seen in the transfer of their energies and populations (Bader *et al.* 1992*b*). This exhibition of transfer, where the properties of a fragment change by negligible amounts upon transfer, is defined as *Near-Perfect* transferability. The same transferability is found for the methyl and methylene contributions to the electronic polarizability and magnetic susceptibility (Bader *et al.* 1992*b*). Thus, the methyl and methylene groups of the homologous series of hydrocarbons exemplify the property of near-perfect transferability.

Having defined and explored the concepts of functional groups and what is meant by transferability, we venture into the study the transferability of the building blocks of life, the amino acid residues.

Chapter 4

Amino Acid Residues as Transferable Functional Groups

Having provided the theoretical background required for the study of transferability within the framework of quantum mechanics we now proceed to analyze the extent to which the building blocks of life, the amino acid residues, are transferable. This analysis will demonstrate the feasibility of using these residues towards the use of the theoretical synthesis of proteins. This chapter presents a detailed account of the computations required for this assessment.

4.1 Background and Calculations

In the early 90s, Head-Gordon *et al* provided a preliminary survey on the feasibility of conducting *ab initio* computations on small blocked-peptides (Head-Gordon *et al.* 1991). The study on the conformational preferences of model glycine and alanine dipeptides at different levels of *ab initio* theory demonstrated that Hartree-Fock (HF) was suitable for the geometry optimization of wavefunctions for these systems. The authors demonstrated that the basis set 6-31+G* was a good choice in mapping the potential energy surface to the extent of recovering the occurrence of hydrogen bonding. A subsequent study (Popelier and Bader 1994) makes use of this level of theory in the geometry optimization of $N-$formyltriglycine amide at the local region of the potential energy surface corresponding to the 3.6_{13} $\alpha-$helix. They followed the optimization with a 6-311++G single point calculation of the wavefunction. The authors recovered the hydrogen bond expected to take place upon a full rotation of the backbone and characterized it on the basis of the properties of the bond**

critical point formed between $C = O \cdots H - N$. This study also showed that neutrality of charge is preserved by the amino acid residues in forming helical structures.

Our efforts extended from the work of Chang and Bader as well as Popelier and Bader by studying the transferability of all the amino acid residues, neutral or charged, which constitute proteins. This entailed the calculations of the density distributions of these functional groups replicating the environment they would have in a protein. Such calculations can be achieved within our computational resources by using a casting template which we defined as the "glycine mold". This mold takes the form $Gly'|Aa|Gly''$, where the "|" denote the surfaces of zero flux which delimit the residues "$|Aa|$ " from the glycines of the mold. A pictorial representation of the mold is presented by figure 4.1.

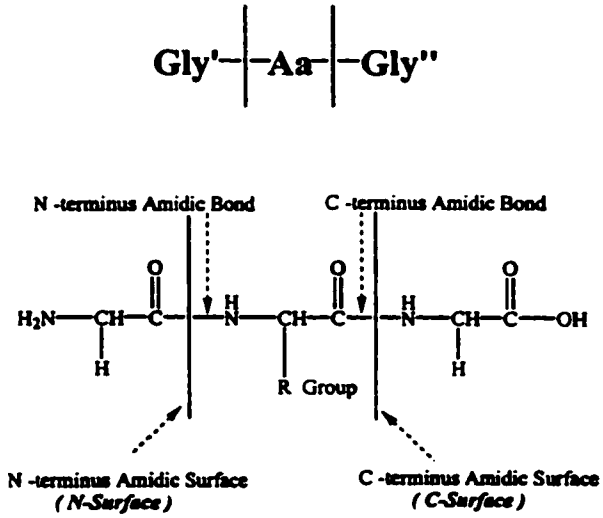


Figure 4.1: Definition of the Glycine Mold

The surface separating Gly′ from $|Aa|$ is henceforth referred to as the "N-Surface". Likewise, the surface delimiting $|Aa|$ from Gly″ is labeled the "C-Surface". Producing density representations of these $|Aa|$ residues in the mold and their subsequent extraction yields a library of fragments. For reasons which will become obvious in the next chapter we chose a perfect $\alpha - R$ helical conformation as the secondary structure for the mold to conform to. To this end the mold's dihedrals were constrained to the optimal values recovered from the Ramachandran secondary structure plots for residue stability (Ramakrishnan and Ramachandran 1965). The rest of the parameters, bond distances and angles, were optimized by a SCF procedure. As in the study of Popelier and Bader,

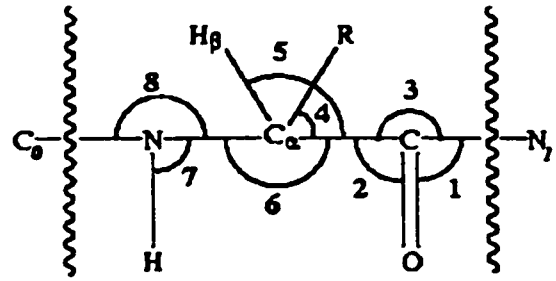


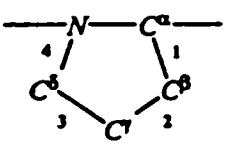
Figure 4.2: Labeling of the common atoms and backbone angles of the $|Aa|$.

the wavefunctions were produced at the HF/6-311++G**//6-31+G* level of theory using the *gaussian 94* package of routines (Frisch *et al.* 1994). These wavefunctions were then used to generate the electron density maps. The later were used to compute the topological features as well as the properties of the residues. Figure 4.2 provides a labeling for the atoms and bond angles of the $|Aa|$. We will use this classification scheme in the remainder of this thesis. Figures 4.3, 4.4 and 4.5 depict the side chains of the amino acid residues studied in this chapter.

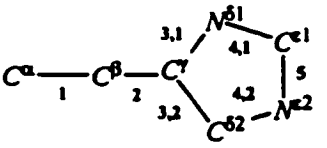
Alanine	$C^{\alpha}\underset{1}{\text{——}}C^{\beta}$
Serine	$C^{\alpha}\underset{1}{\text{——}}C^{\beta}\underset{2}{\text{——}}O^{\gamma}$
Cysteine	$C^{\alpha}\underset{1}{\text{——}}C^{\beta}\underset{2}{\text{——}}S^{\gamma}$
Methionine	$C^{\alpha}\underset{1}{\text{——}}C^{\beta}\underset{2}{\text{——}}C^{\gamma}\underset{3}{\text{——}}S^{\delta}\underset{4}{\text{——}}C^{\epsilon}$
Lysine	$C^{\alpha}\underset{1}{\text{——}}C^{\beta}\underset{2}{\text{——}}C^{\gamma}\underset{3}{\text{——}}C^{\delta}\underset{4}{\text{——}}C^{\epsilon}\underset{5}{\text{——}}N^{\zeta}$

Figure 4.3: Unbranched Side Chains

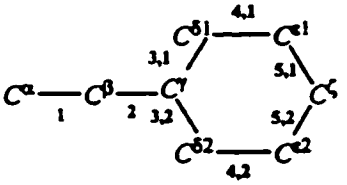
Proline



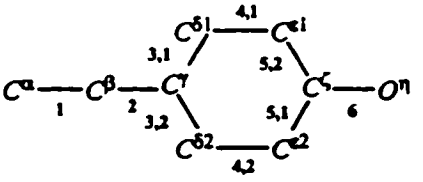
Histidine



Phenylalanine



Tyrosine



Triptophan

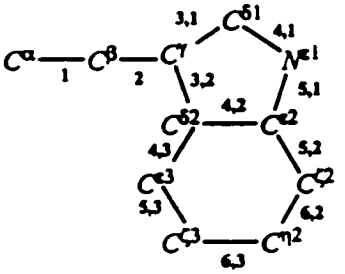
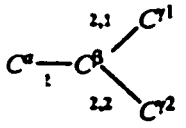
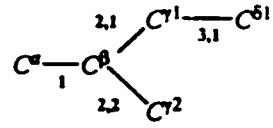


Figure 4.4: Cyclic Side Chains

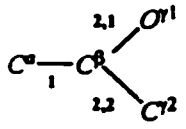
Valine



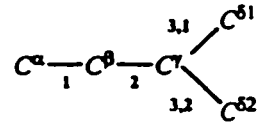
Isoleucine



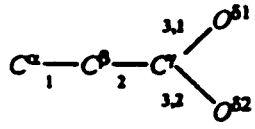
Threonine



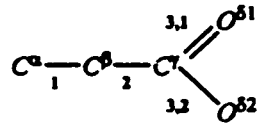
Leucine



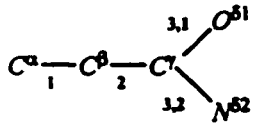
Aspartic Acid



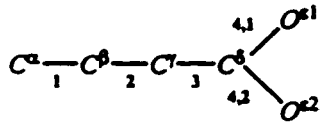
or



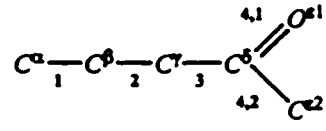
Asparagine



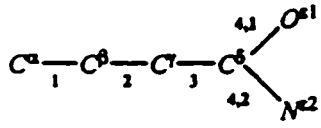
Gluatamic Acid



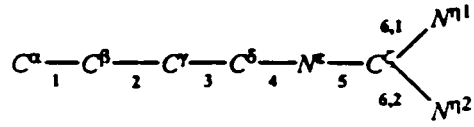
or



Glutamine



Arginine



or

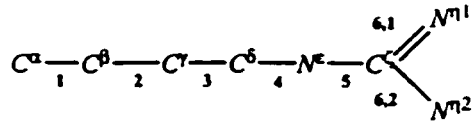


Figure 4.5: Branched Side Chains

In order to thoroughly study how transferable these residues are the following must be analyzed:

- *Geometry:* The commonly occurring bond distances and angles are to be compared among the residues as well as against experimental values, if available.
- *Bond Critical Point Data:* The study of the molecular graphs is used to reveal trends in the bond properties of the commonly occurring bonds throughout the series.
- *Atomic Properties:* The properties of the atoms constituting the residues are computed and compared.
- *Interatomic Surfaces Properties and Form:* The properties of the surfaces, as well as their form as determined by the glycine mold are computed and compared among the residues.

The ensuing sections address all of the points above and the observations made will serve a twofold purpose:

1. To quantify the transferability of the $|Aa|$ on the basis of the reproducibility of their properties.
2. To draw chemical insights from this quantification.

4.2 Transferability of $|Aa|$ Geometrical Parameters

Before a formal description of the calculated geometries is given, it should be noted that some systems exhibited the presence of internal hydrogen bonds. Although their occurrence has a minimal effect on the geometry and properties of the residues which exhibit them it is necessary that they be discussed first.

Internal hydrogen bonding, as it manifested itself using the mold, is the result of two possible types of interactions:

(1) Between the keto oxygen and a $C^{\gamma} - H$ hydrogen belonging to the R group or

(2) between the N-terminus hydrogen and an oxygen belonging to the R group.

These are bonding interactions since they are marked with the existence of a bond path linking the keto oxygen and hydrogen atom and its associated bond critical point. The first type of bonding was found in arginine, its cationic analog and histidine. Figure 4.6 displays the molecular graph

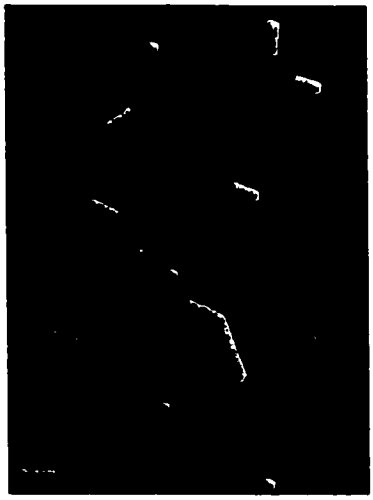


Figure 4.6: Molecular Graph of Arginine(+). The hydrogen bond takes place at the bottom of the image, between the keto oxygen (red), and one of the hydrogens bonded to C^{γ} in the side chain (grey). The formation of this hydrogen bond also results in the formation of a ring structure with $O, H, C^{\gamma}, C^{\beta}$ and C^{α}. The bond paths have been thickened to ease the visualization.

highlighting the bond path $C^{\gamma} - H \cdots O = C$ in arginine(+). A local view of the bonding region is presented in figure 4.7, which depicts the local field of the gradient of the electron density.

Aspartic acid, aspartate and asparagine exhibited hydrogen bonding of the second kind, also characterized with bond paths which the associated bond critical point. As an example, figure 4.8 shows the molecular graph of aspartate. The bonding interaction $N - H \cdots O - R$ is more clearly shown in figure 4.9, the local map of the gradient of the electron density in the plane containing the atoms $N - H \cdots O$.

Table 4.1 shows the properties of all of these hydrogen bonds. Arginine and asparagine exhibit very weak interactions, indicated by the low value of the density at the location of their respective hydrogen bond critical points. The high value of the ellipticity exhibited at these points indicate that the bonding interaction is highly unstable. In addition, the distances of the bond distance between the keto oxygen and H are rather large.

The remaining hydrogen bonds exhibit stronger interaction as the values for the density are slightly larger and the bonding radii are shorter. The values for ellipticity are smaller than those calculated for arginine and asparagine. Interestingly, the cationic residues exhibit a slightly higher conjugation than the anionic ones. The positive value of the Laplacian, along with the low value for the electron density, characterizes the bonds as closed shell interactions.

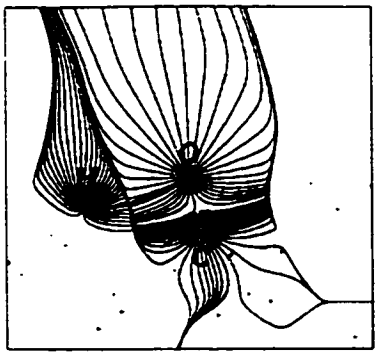


Figure 4.7: Close view of the field of $\nabla\rho$ which clearly recovers the presence of the hydrogen bond $C - O \cdots H - R$ in Arginine(+) as described in figure 4.6 .

Figure 4.8: Molecular Graph of Aspartate. The hydrogen bond takes place at the top-left of the image, between an acidic oxygen in the side chain (red), and H of the backbone (grey). The bond paths have been thickened to ease the visualization.

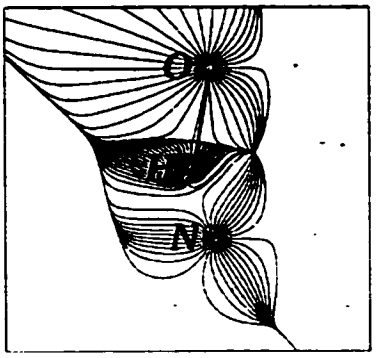


Figure 4.9: Local view of the gradient vector field of Aspartate which highlights the existence of a hydrogen bond between $N - H \cdots O - R$ as described in figure 4.8.

Table 4.1: Properties of the internal hydrogen bonds recovered in $|Aa|$ in au.

Aa	r_{1C-O}	$r_{1C\gamma-H}$	r_{1N-H}	$r_{1C\gamma-O}$	ϵ	ρ_b	$\nabla^2\rho_b$
arg	2.92	2.04			2.348	9.479E-03	3.588E-02
arg(+)	2.68	1.79			0.492	1.318E-02	4.814E-02
his	2.74	1.90			0.231	1.043E-02	3.880E-02
asn			1.98	2.77	1.297	1.043E-02	4.254E-02
asp			1.10	2.18	0.069	4.287E-02	1.492E-01
asp(−)			1.09	2.18	0.057	4.309E-02	1.508E-01

Table 4.2: |*Aa*| Common Back-Bone Bond Lengths.

Aa	$C_0 - N$	$N - H$	$N - C\alpha$	$C\alpha - H$	$C\alpha - C$	$C = 0$	$C - N_1$
ala	1.35	0.99	1.45	1.08	1.53	1.19	1.36
arg	1.35	0.99	1.45	1.08	1.53	1.20	1.36
asn	1.35	0.99	1.44	1.08	1.53	1.19	1.36
asp	1.35	0.99	1.44	1.08	1.53	1.20	1.36
cys	1.35	0.99	1.44	1.08	1.53	1.19	1.36
glu	1.35	0.99	1.44	1.08	1.53	1.19	1.36
gln	1.35	0.99	1.44	1.08	1.53	1.20	1.35
gly	1.35	0.99	1.44	1.07	1.52	1.19	1.36
his	1.35	0.99	1.45	1.08	1.53	1.20	1.35
ile	1.35	0.99	1.45	1.08	1.53	1.20	1.36
leu	1.35	0.99	1.45	1.07	1.53	1.19	1.36
lys	1.35	0.99	1.45	1.08	1.53	1.19	1.36
met	1.35	0.99	1.44	1.08	1.53	1.19	1.36
phe	1.35	0.99	1.45	1.08	1.53	1.19	1.36
ser	1.35	0.99	1.44	1.08	1.53	1.19	1.36
thr	1.35	0.99	1.44	1.08	1.53	1.20	1.35
trp	1.35	0.99	1.45	1.08	1.53	1.19	1.36
tyr	1.35	0.99	1.45	1.08	1.53	1.19	1.36
val	1.35	0.99	1.45	1.08	1.53	1.19	1.36
Ave	1.35	0.99	1.449	1.08	1.53	1.19	1.36
AveDev	1.15E-03	4.55E-04	2.71E-03	6.81E-04	1.13E-03	1.62E-03	1.94E-03
σ^2	2.85E-06	3.62E-07	1.03E-05	7.91E-07	3.56E-06	5.49E-05	1.02E-05
σ	1.69E-03	6.01E-04	3.21E-03	8.90E-04	1.89E-03	2.34E-03	3.19E-03
Max	1.35	0.99	1.45	1.08	1.53	1.20	1.36
Min	1.35	0.99	1.44	1.07	1.52	1.19	1.35
pro	1.36	1.47	1.46	1.07	1.52	1.20	1.35
arg(+)	1.36	0.99	1.44	1.08	1.52	1.20	1.35
lys(+)	1.36	0.99	1.44	1.08	1.53	1.19	1.36
his(+)	1.36	0.99	1.44	1.08	1.53	1.20	1.35
Ave	1.36	0.99	1.44	1.08	1.53	1.20	1.35
asp(−)	1.33	1.01	1.44	1.08	1.52	1.20	1.36
glu(−)	1.34	0.99	1.45	1.08	1.52	1.20	1.35
tyr(−)	1.34	0.99	1.45	1.08	1.52	1.19	1.36
Ave	1.34	1.00	1.45	1.08	1.52	1.20	1.36
lys(+)*Cl*(−)	1.35	0.99	1.44	1.08	1.53	1.19	1.36
asp(−)*Na*(+)	1.35	0.99	1.44	1.08	1.52	1.19	1.35

4.2.1 Comparison of the |Aα| Geometrical Parameters with Experimental Results

The statistical analysis of the data presented in table 4.2 indicates that the backbone bondlengths of the |Aα| exhibit very small variations in their respective series. With this in mind, we can safely use the average values for these bond lengths in the ensuing study. The first question to be answered is whether the geometries of the |Aα| obtained upon optimizing the mold are in agreement with the experimental data available. The standard geometrical parameters for the peptide units in a polypeptide are provided in figure 4.10 and are the result of the study of crystal structures of small polypeptides (Pauling *et al.* 1951), and which remain widely accepted (Schulz and Schirmer 1979, Murray *et al.* 2000). We start by comparing the |Aα| back-bone bond lengths, with the reported

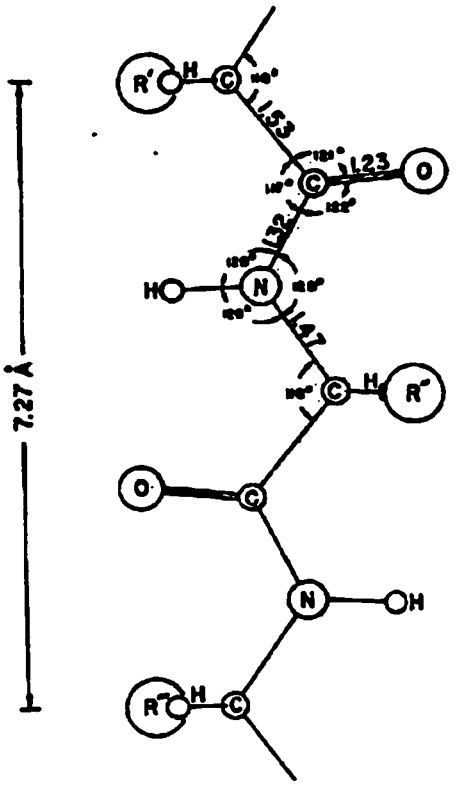


Figure 4.10: Average Dimensions of the Polypeptide Chain as proposed by Pauling, Corey and Branson (1951).

values of Pauling and co-workers. In general terms, the bond lengths of the casted residues, with

the exception of $C_\alpha - C$, are slightly longer than those provided by X-ray data. This is not an uncommon observation as the Hartree-Fock method tends to overestimate bond distances on the account of the lack of electron-electron correlation built into such procedure. Thus, the peptide bonds are 0.037 Å longer in the mold. $N - C_\alpha$ is 0.021 Å longer in the mold as well. The length of $C_\alpha - C$ in the $|A\alpha|$ is the same as in the experimental data. $C = O$ is slightly shorter in the mold, by 0.031 Å, the result of electronic effects which will be discussed in the next section. Pauling *et. al.* structural results are complemented with the neutron diffraction studies of Barret *et al*, who produced a compilation of structural parameters of the bonds involving hydrogen of several free amino acids in their switerionic forms (Barret 1985). The results of this study are summarized in figure 4.11. Comparing the results produced in the mold to these data is not as straight

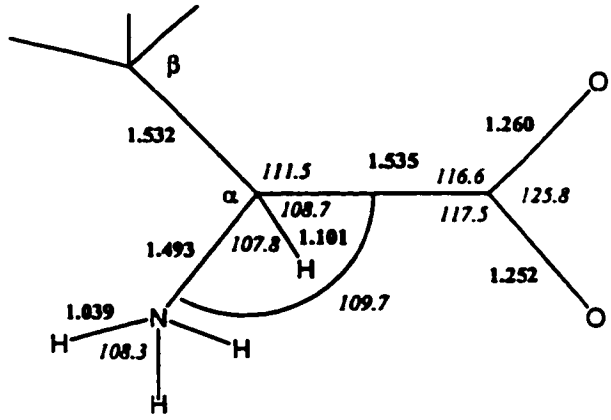


Figure 4.11: Mean geometry of Amino Acid main chain atoms taken from neutron diffraction results (Barret *et. al.* 1985). Bond distances are shown in bold font and angles in *italic*.

forward since the data in figure 4.11 does not reflect the structural effects on the residues as these are constituents of proteins. Nonetheless, one could assume that the lengths of $N - H$ and $C_\alpha - H_\beta$ will not vary significantly as the amino acids link into forming proteins. With this in mind the $N - H$ distance was found to be 0.042 Å shorter in the $|A\alpha|$, whereas that for $C_\alpha - H_\beta$ was found to be 0.07 Å longer, which are small deviations indeed.

In summary, the computed values the backbone bondlengths bonds are in close agreement with the available experimental data. This is perhaps the most stringent test of the validity of our proposed methodology for the study of transfer of the amino acid residues which constitute proteins.

The first measure of transfer of the $|A\alpha|$ is provided by the comparison of the amidic bond lengths in the set. Upon examining table 4.2 one notices that the C-amidic bond is 0.005 Å longer

than the N-amidic bond. It also exhibits nearly twice standard deviation than that calculated for the N-amidic bond, but of very small value. This bond length difference is extremely small, but nonetheless reflective of the fact that the $|Gly''$ group is slightly more electron-withdrawing than the $Gly'|$ group is electron-donating. By virtue of being more heavily polarized, the electron density between the bonded C and N_1 is locally depleted with respect to their N-terminus counterparts, and thus these atoms exhibit a slightly higher separation. These electronic effects will be revisited in more detail in the next section.

The remaining backbone bond lengths exhibit very small standard deviations as well, marked by both the $N-H$ and $C_\alpha - H$ bonds. This result hints that these bonds are not as afflicted as the rest of the common backbone bonds in accommodating the various R groups of the residues in this study. Because of this observation, which will be examined in more detail when we study the transfer of the bond critical point data in the $|Aa|$ in the ensuing section. Even those residues for which internal hydrogen bonding was recovered are not found to contribute large variations from the mean of their respective series.

The subset of residues with positively charged side chains revealed slightly longer N-amidic and unchanged C-amidic bond lengths when compared with their non-charged counterparts. Arginine(+) seems to break this trend as the N-amide bond shows the opposite of the effect described above but this is caused by the presence of the neighboring internal hydrogen bond $C=O\cdots H-C^{\gamma}$. Interestingly enough, this trend is reversed in the case of the cationic subset; the N-amidic bond is lengthened while the C-amidic bond remains very close to the mean of the non-charged counterparts. Aspartate breaks the trend by exhibiting a relative shortening of the $C_0 - N$ bondlength, but this is so because of the internal hydrogen bond found to involve H.

To study the extent to which variations in the properties of residues with ionized chains are affected by the stabilization of a counter-ion - as it would most likely be the case in nature - we complemented our study with calculations on $lys(+)Cl(-)$ and $asp(-)Na(+)$. It can be seen that the trends manifested by the lengthening/shortening of the afflicted backbone bonds of the charged $|Aa|$ disappear, causing the backbones of these residues to exhibit bondlengths approaching the average values of the non-charged series.

In addition to comparing the variations in the similar bonds among the $|Aa|$ series, we thought that it would be interesting to study how the free amino acids geometries change when linked into strands. This undertaking would entail paralleling the studies conducted on the $|Aa|$ but this time of their corresponding free amino acids. Fortunately, our laboratory has conducted a series of such calculations (Matta and Bader n.d.), and which data will be presented as needed

throughout this chapter. The authors paralleled the level of theory in their computations to that conducted in this thesis. A summary of the free amino acid backbone bond lengths is presented in table 4.3. The average values for the bond lengths of the free residues are close to those recovered

Table 4.3: Free Residue Bond Length Summary.

Aa	$N-H$	$N-C\alpha$	$C\alpha-H$	$C\alpha-C$	$C=0$
Free Aa	1.00	1.44	1.07	1.52	1.19

using the mold, but the comparison reveals a trend in which the bondlengths in the vicinity of the N-terminus are similar in the $|Aa|$ as in their free counterparts, but become slightly longer towards the vicinity of the C-terminus in the mold. Thus, $N-H$, $N-C_\alpha$ and $C_\alpha-H_\beta$ remain virtually unchanged, while $C_\alpha-C$ and $C=O$ are found to be 0.008 Å longer in the mold. These variations cannot be discussed in further detail without looking at the local properties of the electron density in the bonding regions. Such study is provided in the ensuing section, where the properties of the bond critical points and their effects on the lengths of the associated bonds are analyzed.

Just like the bond distances in our $|Aa|$ are in good agreement with the experimental values reported by Pauling (Pauling *et al.* 1951), the bond angles recovered from the $|Aa|$ series also show a remarkable agreement with their experimental counterparts (figure 4.10) as well. The only marked discrepancy, if it could be called so, is that of $\angle N - C_\alpha - C$ exhibits a difference of 3.37° between our values and Pauling's. The experimental parameters provided by Barret (Barret 1985) are also in good agreement with our values. The angle which showed the largest difference between the two sets was $\angle H_\beta - C_\alpha - C$, which is reported to be 3.8° larger in Barret's study.

Sterric effects manifest themselves in the form of bond angle fluctuations. Table 4.4 reflects such variations on the $|Aa|$. As anticipated, the bond angles which exhibit the highest standard deviations are those made with C_α, the result of accommodating the side chain R. The bond angle most susceptible to change is $\angle R - C_\alpha - C$ with values[1] of 104.9°(1.86) followed by $\angle H - C_\alpha - C$ with 113.4°(0.92) and $\angle N - C_\alpha - C$ with 107.3°(0.75). The constancy in values for the neutral and ionized residues further re-enforces the notion that steric effects are exhibited mainly in the form of bond angles variations.

The last set of geometrical parameters to be discussed in this section are those corresponding to the bond distances and angles of the R groups. Because a comprehensive account of experimental data for the geometries of these groups seems not to be available we will focus our attention to the study of the effects produced on the backbone of the various R groups by studying the data produced in the mold and that provided by Matta and Bader. In general terms, the geometries of the side-chains produced using the mold exhibit a general shortening of the bond distances in the immediacy of C_α when compared to the corresponding values reported in Matta and Bader's data (Matta and Bader n.d.). These decreases in bond lengths tapper off as one compares the values of the corresponding bonds deeper into the side-chain. Because the direct relationship between electronic effects and bond distances, such observation indicates that there is a high polarization of the electron density from the R group towards the residue's backbone. The rest of the geometrical parameters - bond and torsion angles - did not vary significantly from the data reported in the above mentioned study.

[1] mean(standard deviation)

Table 4.4: Common Back-Bone Bond Angles

Aa	1	2	3	4	5	6	7	8
ala	122.1	121.0	116.3	110.0	104.6	113.9	119.5	122.7
arg	121.7	123.1	114.7	109.1	106.3	110.7	121.1	120.5
asn	122.6	121.0	115.8	108.4	104.7	113.5	119.7	121.5
asp	122.6	120.6	116.3	109.0	104.9	113.4	120.6	121.0
cys	122.4	120.9	116.1	108.6	104.8	113.9	119.9	122.2
glu	122.1	121.0	116.3	111.3	104.8	113.5	119.7	122.5
gln	122.0	121.2	116.3	111.4	104.7	113.3	119.5	122.6
gly	122.2	121.2	115.9	107.5	106.4	115.5	119.9	122.2
his	121.8	121.6	116.3	112.6	105.2	112.3	119.5	122.7
ile	121.5	121.0	117.0	113.9	103.7	113.0	120.2	122.5
leu	122.0	121.0	116.4	108.9	104.1	113.1	119.6	122.6
lys	121.9	121.3	116.2	111.5	104.9	113.2	119.5	122.7
met	122.0	121.2	116.3	111.5	104.8	113.4	119.6	122.6
phe	121.9	121.6	116.1	111.9	105.0	113.1	119.5	122.8
ser	122.3	121.0	116.0	108.8	105.6	113.7	119.5	122.4
thr	122.1	120.5	117.1	111.1	103.6	114.4	118.6	122.0
trp	121.8	121.8	116.0	111.5	104.9	113.0	119.5	122.9
tyr	122.0	121.6	116.0	111.9	104.9	113.2	119.5	122.8
val	121.6	121.0	116.9	113.9	103.7	113.0	120.0	122.6
Ave	122.0	121.2	116.2	110.7	104.8	113.3	119.7	122.3
AveDev	2.21E-01	3.65E-01	3.18E-01	1.58E+00	4.98E-01	5.51E-01	3.51E-01	4.68E-01
σ^2	8.34E-02	3.06E-01	2.61E-01	3.46E+00	5.57E-01	8.43E-01	2.64E-01	4.15E-01
σ	2.89E-01	5.53E-01	5.11E-01	1.86E+00	7.46E-01	9.18E-01	5.14E-01	6.44E-01
Max	122.6	123.1	117.1	113.9	106.4	115.5	121.1	122.9
Min	121.5	120.5	114.7	107.5	103.6	110.7	118.6	120.5
pro	122.0	120.0	117.7	110.8	105.7	114.2	111.7	118.5
arg(+)	122.1	121.9	114.9	108.3	105.6	111.7	122.2	119.3
lys(+)	122.4	120.4	116.4	110.9	104.3	114.2	120.4	121.3
his(+)	122.7	119.9	116.7	111.8	104.6	114.2	120.8	120.4
Ave	122.4	120.8	116.0	110.4	104.8	113.3	121.1	120.3
asp(−)	121.1	123.1	115.6	110.0	105.9	110.1	111.3	122.0
Ave	121.5	123.1	115.2	109.7	106.1	111.6	114.2	123.6
lys(+)*Cl*(−)	122.0	121.3	116.2	111.1	104.9	113.3	119.7	122.5
asp(−)*Na*(+)	121.9	121.8	116.0	108.9	105.9	111.0	118.1	120.6

Table 4.5: Unbranched Side Chains: Optimized Bond Distances.

BOND DISTANCES(Å)	1	2	3	4	5
ala	1.52				
ser	1.52	1.40			
cys	1.52	1.82			
met	1.53	1.52	1.81	1.80	
lys	1.53	1.53	1.52	1.53	1.45

Table 4.6: Unbranched Side Chains: Optimized Bond Angles.

BOND ANGLES(degrees)	∠12	∠23	∠34	∠45
ser	108.0			
cys	110.4			
met	113.6	109.5	99.9	
lys	114.6	112.1	112.8	115.6

Table 4.7: Branched Side Chains: Optimized Bond Distances.

BOND DISTANCES(Å)	1	2	2,1	2,2	3,1	3,2
val	1.54		1.53	1.53		
ile	1.54		1.53	1.53	1.53	
thr	1.53		1.41	1.52		
leu	1.53	1.54			1.53	1.53
asp(−)	1.54	1.56			1.21	1.25
asp	1.53	1.50			1.18	1.32
ans	1.53	1.51			1.20	1.34

Table 4.8: Branched Side Chains: Optimized Bond Distances *(continued)*.

(Continued) BOND DISTANCES (Å)	1	2	3	4	4,1	4,2	5	6,1	6,2
glu	1.53	1.53	1.50		1.18	1.33			
glu(−)	1.53	1.52	1.55		1.23	1.23			
gln	1.53	1.52	1.51		1.20	1.35			
arg	1.54	1.54	1.53	1.46			1.38	1.26	1.38
arg(+)	1.54	1.54	1.53	1.48			1.31	1.32	1.32

Table 4.9: Branched Side Chains: Optimized Bond Angles.

BOND ANGLES (degrees)	∠12	∠12,1	∠12,2	∠2,12,2	∠23,1	23,2	∠2,13,1	∠3,13,2
val		111.2	115.2	110.3				
ile		110.6	114.5	111.8			113.9	
thr		106.7	112.5	111.1				
leu	117.5				114.5	112.5		110.9
asp	110.8				114.5	112.5		123.1
asp(−)	115.6				116.3	114.7		128.8
asn	111.4				121.2	116.0		122.7

Table 4.10: Branched Side Chains: Optimized Bond Angles *(continued)*.

(Continued) BOND ANGLES (degrees)	∠12	∠23	∠34	∠34,1	∠34,2	∠4,14,2	∠45	∠56,1	∠56,2	∠6,16,2
glu	113.2	111.3		125.4	112.4	122.0				
glu(−)	116.2	112.5		114.9	115.9	129.0				
gln	113.3	112.0		123.1	114.7	122.0				
arg	116.6	111.5	110.2				125.2	126.7	113.9	119.2
arg(+)	109.2	112.3	108.2				125.3	120.5	120.2	119.2

Table 4.11: Cyclic Side Chains: Optimized Bond Distances.

BOND DISTANCES(Å)	1	2	3	3,1	3,2	4	4,1	4,2	4,3
pro	1.53	1.52	1.52			1.47			
his	1.53	1.50		1.37	1.35		1.29	1.37	
phe	1.54	1.51		1.39	1.39		1.35	1.35	
tyr	1.54	1.51		1.39	1.39		1.38	1.38	
trp	1.54	1.49		1.44	1.44		1.31	1.44	1.39

Table 4.12: Cyclic Side Chains: Optimized Bond Distances *(continued)*.

BOND DISTANCES (Å) *(Continued)*	5	5,1	5,2	5,3	6,2	6,3	6
pro							
his	1.34						
phe		1.38	1.38				
tyr		1.38	1.38		1.35		
trp		1.37	1.39	1.37		1.37	1.40

Table 4.13: Cyclic Side Chains: Optimized Bond Angles.

BOND ANGLES (degrees)	∠12	∠23	∠23,1	∠23,2	∠3,4	∠3,14,1	∠3,13,2
pro	103.1	103.5			103.7		
his	117.0		118.2	131.9		105.8	109.8
phe	115.0		120.9	120.5		120.7	118.5
tyr	115.2		119.7	121.2		121.5	119.0
trp	114.9		126.8	126.9		110.4	106.1

Table 4.14: Cyclic Side Chains: Optimized Bond Angles *(continued 1)*.

BOND ANGLES (degrees) *(Continued)*	∠4,15	∠4,15,1	∠4,25	∠4,25,2	∠4,24,3	∠4,35,3	∠4,25,1
his	111.9		107.0				
phe		120.0		120.0			
tyr		119.7		119.6			
trp		119.5		119.1	122.2	117.4	110.4

Table 4.15: Cyclic Side Chains: Optimized Bond Angles *(continued 2)*.

BOND ANGLES (degrees) *(Continued)*	∠5,15,2	∠5,16	∠5,26	∠5,26,2	∠6,26,3
phe	119.5				
tyr	119.7	122.6	117.6		
trp	130.4			109.0	120.8

4.3 $|Aa|$ Bond Critical Point Analysis

If steric effects manifest themselves on the $|Aa|$ as fluctuations of the bond angles, electronic effects manifest themselves via the polarization of the electron density inducing bonds to lengthen or shorten. As a general rule of thumb, the decrease in electron density in a bonding region causes the bond to lengthen and vice-versa.

We commence by comparing the properties of the amidic bonds at their respective bond critical points within the $|Aa|$ series. It was indicated in section 4.2 that the neutral residues exhibited a slightly shorter $C_0 - N$ bond distance over its $C - N_1$ counterpart. Upon examining tables 4.18 and 4.30 one notices that the bond critical point (bcp) for $C_0 - N$ shows a slightly larger value for the density than that for $C - N_1$, a difference of 0.0016 au. The increase in density is accompanied with a decrease in ellipticity of 0.0081 au (tables 4.17 and 4.29). Thus, the C-terminus amidic bond is slightly of more conjugated character than its N-terminus counterpart. This conjugation takes place with the keto bond, as verified by comparing the properties of the bond critical points of the keto bond in the $|Aa|$ to those for the same bond in the free amino acid series, which values are shown in table 4.16. The keto bond exhibits a higher ellipticity in the free series than in the $|Aa|$ thus indicating that this bond acquires more of a "single-bond" character when the residue is a constituent in a protein.

Conjugation effects become more apparent in the charged $|Aa|$ as the ellipticity of the N-amidic bond increases slightly for both the anionic as well as the cationic residues. The ellipticity of the C-amidic bond remains unchanged for the cationic species but exhibits an increase on the anionic $|Aa|$.

Table 4.16: Summary of the free residue bond critical point data.

Bond	r_A	r_B	ϵ	ρ_c	$\nabla^2\rho_c$
$N-H$	1.39	0.49	0.051	0.355	-1.818
$N-C_\alpha$	1.71	1.02	0.018	0.281	-0.905
$C_\alpha-H$	1.29	0.74	0.027	0.295	-1.102
$C_\alpha-C$	1.29	1.58	0.075	0.269	-0.800
$C=O$	0.81	1.70	0.162	0.308	-0.066

The comparison between the $|Aa|$ bond critical point data of the backbone bonds and those

calculated for the free residues, shown in table 4.16, produced the following observations:

- $N - H$, tables 4.19 and 4.20. This bond is found to be longer in the mold than in the free residues. The distance between the critical point and N is lengthened while that with H is shortened. The gain in electron density by this bond places the bcp slightly closer to the lesser electro-negative hydrogen. The ellipticity is slightly increased as well. The bond seems relatively unaffected to variations in charge of the side chain. The occurrence of internal hydrogen bonds does not impart a significant effect of this bond's geometry.

- $N - C_\alpha$, tables 4.21 and 4.22. Charge polarization is more apparent in this bond. Compared against the free residues, the distance between the bcp and carbon is lengthened whilst shortening with respect to nitrogen. The ellipticity is slightly higher while the density is reduced. The bcp exhibits a slightly more positive Laplacian. These are indications of an increase in conjugation from the observed values in the free residues. The extent of the conjugation is further increased in the ionic species.

- $C_\alpha - H_\beta$, tables 4.23 and 4.24. This bond is recipient of electron density. While this bond length is comparable to that of the free amino acid series occurring analog, this bond exhibits an increase in ellipticity, density and more negative Laplacian in the mold. The properties of this bond seem unaltered by the transfer of charged side-chains. The short distance between the bond critical point and C_α indicate that the interaction is dominated by this atom.

- $C_\alpha - C$, tables 4.25 and 4.26. The bcp recovered is situated closer to C and more distant to C_α than in the free residues. The ellipticity is reduced as well as the value for the point density. The Laplacian is slightly more positive. This effect is more dramatically exhibited by the cationic species. The anionic depict a more stabilized state than its cationic counterparts, but the values for the ellipticity and density are still lower than those measured in the free residues.

- $C = O$, tables 4.27 and 4.28. The keto bond is the most electron-rich bond in the backbone, and as such, the most susceptible to polarization. It comes as no surprise that the bcp of the $|Ac|$ is closer to oxygen than what has been found on the free residues. Those residues which exhibited hydrogen bond exhibit small deviation in their properties from the mean of their respective series, indicating that these hydrogen bonds are weak interactions. The ellipticity is greatly reduced while the density increased as well as the Laplacian. This indicates a reduction

in π-character of this bond from the free series. The ionized species do not indicate a different trend in properties.

The summary of bond critical properties highlights two important properties, a) that the resides exhibit a high degree of transfer as characterized by the low fluctuations in their properties, and b) the peptide bonds experience a conjugation with the neighboring keto bonds creating electronically-buffered regions at the $|A\alpha|$ boundaries. These electronic effects will be explored in more detail in the next section, as we study the atomic properties of the backbone atoms.

Table 4.17: $C_0 - N$ Bond Critical Point Properties.

Aa	r_A	r_B	ϵ	λ_1	λ_2	λ_3
ala	0.86	1.69	0.0074	-0.7319	-0.7266	0.6341
arg	0.86	1.69	0.0070	-0.7298	-0.7247	0.6248
asn	0.86	1.69	0.0120	-0.7326	-0.7239	0.6480
asp	0.86	1.69	0.0059	-0.7196	-0.7153	0.6375
cys	0.86	1.69	0.0028	-0.7256	-0.7236	0.6424
glu	0.86	1.69	0.0039	-0.7265	-0.7237	0.6336
gln	0.86	1.69	0.0070	-0.7320	-0.7268	0.6375
gly	0.86	1.69	0.0056	-0.7302	-0.7262	0.6380
his	0.86	1.69	0.0078	-0.7324	-0.7267	0.6397
ile	0.86	1.69	0.0107	-0.7339	-0.7261	0.6347
leu	0.86	1.69	0.0109	-0.7354	-0.7274	0.6351
lys	0.86	1.69	0.0083	-0.7323	-0.7263	0.6344
met	0.86	1.69	0.0058	-0.7295	-0.7253	0.6344
phe	0.86	1.69	0.0064	-0.7297	-0.7251	0.6317
ser	0.86	1.69	0.0143	-0.7375	-0.7272	0.6609
thr	0.86	1.69	0.0079	-0.7348	-0.7291	0.6317
trp	0.86	1.69	0.0070	-0.7386	-0.7277	0.6308
tyr	0.86	1.69	0.0068	-0.7305	-0.7256	0.6317
val	0.86	1.69	0.0102	-0.7334	-0.7260	0.6331
Ave	0.86	1.69	0.0078	-0.7314	-0.7254	0.6366
AveDev	6.39E-04	1.43E-03	2.09E-03	3.12E-03	1.74E-03	5.08E-02
σ^2	9.55E-07	4.10E-06	7.84E-06	1.89E-05	7.99E-06	5.84E-05
σ	9.77E-04	2.03E-03	2.80E-03	4.34E-02	2.83E-03	7.64E-03
Max	0.86	1.69	0.0143	-0.7196	-0.7153	0.6609
Min	0.86	1.69	0.0028	-0.7386	-0.7291	0.6248
pro	0.88	1.69	0.0280	-0.7224	-0.7027	0.5262
arg(+)	0.87	1.70	0.0096	-0.7129	-0.7062	0.6127
lys(+)	0.87	1.70	0.0116	-0.7134	-0.7052	0.6187
his(+)	0.87	1.71	0.0262	-0.7038	-0.6859	0.6073
Ave	0.87	1.70	0.0158	-0.7100	-0.6991	0.6129
asp(−)	0.86	1.66	0.0374	-0.7954	-0.7667	0.6365
glu(−)	0.86	1.67	0.0335	-0.7682	-0.7433	0.6488
tyr(−)	0.86	1.67	0.0313	-0.7626	-0.7394	0.6430
Ave	0.86	1.67	0.0341	-0.7754	-0.7498	0.6428
lis(+)*Cl*(−)	0.86	1.69	0.0044	-0.7277	-0.7245	0.6367
asp(−)*Na*(+)	0.86	1.68	0.0222	-0.7602	-0.7436	0.6527

Table 4.18: $C_0 - N$ Bond Critical Point Properties *(continued)*.

Aa	σ_1	σ_2	σ_3	tr_σ	ρ_c	$\nabla^2\rho_c$
ala	-0.3636	-0.3452	-0.0995	-0.8083	0.3275	-0.8244
arg	-0.3628	-0.3443	-0.0985	-0.8056	0.3271	-0.8296
asn	-0.3650	-0.3460	-0.1006	-0.8117	0.3272	-0.8085
asp	-0.3612	-0.3407	-0.0984	-0.8003	0.3239	-0.7975
cys	-0.3631	-0.3440	-0.0998	-0.8069	0.3260	-0.8068
glu	-0.3624	-0.3435	-0.0989	-0.8049	0.3262	-0.8166
gln	-0.3640	-0.3455	-0.0998	-0.8093	0.3275	-0.8213
gly	-0.3637	-0.3450	-0.0999	-0.8086	0.3271	-0.8183
his	-0.3642	-0.3459	-0.1001	-0.8101	0.3275	-0.8195
ile	-0.3639	-0.3460	-0.0996	-0.8095	0.3279	-0.8253
leu	-0.3644	-0.3465	-0.0998	-0.8106	0.3282	-0.8277
lys	-0.3636	-0.3454	-0.0995	-0.8085	0.3275	-0.8242
met	-0.3631	-0.3445	-0.0993	-0.8069	0.3269	-0.8204
phe	-0.3628	-0.3443	-0.0990	-0.8061	0.3269	-0.8230
ser	-0.3670	-0.3488	-0.1025	-0.8183	0.3283	-0.8039
thr	-0.3639	-0.3455	-0.0994	-0.8089	0.3281	-0.8321
trp	-0.3657	-0.3441	-0.0989	-0.8087	0.3236	-0.8239
tyr	-0.3629	-0.3445	-0.0991	-0.8065	0.3271	-0.8244
val	-0.3636	-0.3457	-0.0994	-0.8087	0.3278	-0.8263
Ave	-0.3637	-0.3450	-0.0996	-0.8083	0.3270	-0.8197
AveDev	8.66E-04	1.08E-03	5.68E-03	2.23E-02	8.62E-04	7.02E-03
σ^2	1.57E-06	2.49E-06	7.78E-07	1.21E-05	1.63E-06	8.55E-05
σ	1.25E-03	1.58E-03	8.82E-04	3.48E-03	1.28E-03	9.25E-03
Max	-0.3612	-0.3407	-0.0984	-0.8003	0.3283	-0.7975
Min	-0.3670	-0.3488	-0.1025	-0.8183	0.3236	-0.8321
pro	-0.3483	-0.3324	-0.0856	-0.7663	0.3242	-0.8989
arg(+)	-0.3571	-0.3364	-0.0954	-0.7888	0.3219	-0.8064
lys(+)	-0.3569	-0.3359	-0.0957	-0.7885	0.3215	-0.7999
his(+)	-0.3520	-0.3293	-0.0932	-0.7744	0.3172	-0.7824
Ave	-0.3553	-0.3339	-0.0948	-0.7839	0.3202	-0.7962
asp(−)	-0.3807	-0.3644	-0.1053	-0.8504	0.3430	-0.9255
glu(−)	-0.3721	-0.3576	-0.1041	-0.8338	0.3355	-0.8627
tyr(−)	-0.3701	-0.3553	-0.1029	-0.8283	0.3342	-0.8590
Ave	-0.3743	-0.3591	-0.1041	-0.8375	0.3376	-0.8824
lis(+)*Cl*(−)	-0.3630	-0.3442	-0.0994	-0.8066	0.3265	-0.8155
asp(−)*Na*(+)	-0.3730	-0.3555	-0.1038	-0.8322	0.3343	-0.8511

Table 4.19: $N - H$ Bond Critical Point Properties.

Aa	r_A	r_B	ϵ	λ_1	λ_2	λ_3
ala	1.41	0.46	0.0632	-1.4045	-1.3211	0.7579
arg	1.41	0.46	0.0628	-1.4015	-1.3187	0.7507
asn	1.42	0.45	0.0588	-1.4279	-1.3487	0.7329
asp	1.42	0.46	0.0615	-1.4147	-1.3328	0.7416
cys	1.41	0.46	0.0621	-1.4088	-1.3264	0.7505
glu	1.41	0.46	0.0631	-1.4038	-1.3204	0.7564
gln	1.41	0.46	0.0628	-1.4056	-1.3225	0.7551
gly	1.41	0.46	0.0631	-1.4150	-1.3311	0.7560
his	1.41	0.46	0.0632	-1.4050	-1.3215	0.7567
ile	1.41	0.46	0.0626	-1.4114	-1.3283	0.7591
leu	1.41	0.46	0.0628	-1.4087	-1.3254	0.7581
lys	1.41	0.46	0.0632	-1.4040	-1.3205	0.7585
met	1.41	0.46	0.0631	-1.4045	-1.3211	0.7563
phe	1.41	0.46	0.0633	-1.4030	-1.3196	0.7579
ser	1.41	0.46	0.0618	-1.4139	-1.3316	0.7541
thr	1.42	0.45	0.0590	-1.4237	-1.3445	0.7385
trp	1.41	0.46	0.0630	-1.4031	-1.3187	0.7577
tyr	1.41	0.46	0.0632	-1.4030	-1.3196	0.7580
val	1.41	0.46	0.0627	-1.4110	-1.3278	0.7587
Ave	1.41	0.46	0.0624	-1.4091	-1.3263	0.7534
AveDev	3.32E-03	2.77E-03	9.18E-04	5.67E-03	6.37E-03	5.56E-03
σ^2	2.03E-05	1.55E-05	1.75E-06	5.33E-05	7.22E-05	5.69E-05
σ	4.50E-03	3.94E-03	1.32E-03	7.30E-03	8.49E-03	7.54E-03
Max	1.42	0.46	0.0633	-1.4015	-1.3187	0.7591
Min	1.41	0.45	0.0588	-1.4279	-1.3487	0.7329
pro	...	...	...	...	...	...
arg(+)	1.42	0.46	0.0626	-1.3996	-1.3172	0.7447
lys(+)	1.41	0.46	0.0631	-1.4006	-1.3175	0.7532
his(+)	1.41	0.46	0.0626	-1.3981	-1.3159	0.7470
Ave	1.41	0.46	0.0628	-1.3994	-1.3169	0.7483
asp(−)	1.50	0.42	0.0416	-1.4021	-1.3461	0.6589
glu(−)	1.41	0.46	0.0619	-1.4102	-1.3280	0.7590
tyr(−)	1.41	0.47	0.0628	-1.4050	-1.3129	0.7644
Ave	1.44	0.45	0.0554	-1.4058	-1.3290	0.7274
lys(+)*Cl*(−)	1.41	0.46	0.0631	-1.4043	-1.3209	0.7564
asp(−)*Na*(+)	1.43	0.44	0.0537	-1.4390	-1.3658	0.7183

Table 4.20: $N - H$ Bond Critical Point Properties *(continued)*.

Aa	σ_1	σ_2	σ_3	tr_σ	ρ_c	$\nabla^2\rho_c$
ala	-0.3944	-0.3802	0.1779	-0.5967	0.3549	-1.9677
arg	-0.3928	-0.3787	0.1763	-0.5952	0.3530	-1.9695
asn	-0.3974	-0.3837	0.1725	-0.6086	0.3533	-2.0437
asp	-0.3954	-0.3813	0.1745	-0.6022	0.3534	-2.0059
cys	-0.3948	-0.3806	0.1764	-0.5990	0.3542	-1.9847
glu	-0.3942	-0.3800	0.1775	-0.5967	0.3546	-1.9679
gln	-0.3943	-0.3802	0.1773	-0.5972	0.3545	-1.9730
gly	-0.3969	-0.3826	0.1779	-0.6016	0.3559	-1.9901
his	-0.3944	-0.3802	0.1776	-0.5970	0.3548	-1.9699
ile	-0.3964	-0.3824	0.1780	-0.6008	0.3562	-1.9805
leu	-0.3955	-0.3814	0.1779	-0.5990	0.3558	-1.9760
lys	-0.3943	-0.3802	0.1780	-0.5965	0.3550	-1.9660
met	-0.3943	-0.3801	0.1775	-0.5969	0.3547	-1.9693
phe	-0.3942	-0.3800	0.1778	-0.5964	0.3548	-1.9647
ser	-0.3961	-0.3819	0.1776	-0.6004	0.3551	-1.9915
thr	-0.3970	-0.3837	0.1734	-0.6073	0.3541	-2.0297
trp	-0.4075	-0.3935	0.1793	-0.6217	0.3605	-2.0091
tyr	-0.4074	-0.3932	0.1793	-0.6213	0.3603	-2.0083
val	-0.3964	-0.3823	0.1779	-0.6008	0.3561	-1.9801
Ave	-0.3965	-0.3824	0.1771	-0.6019	0.3553	-1.9867
AveDev	2.49E-03	2.58E-03	1.30E-03	5.45E-03	1.35E-03	1.80E-02
σ^2	1.63E-05	1.66E-05	3.22E-06	6.08E-05	4.00E-06	5.20E-04
σ	4.04E-03	4.07E-03	1.79E-03	7.79E-03	2.00E-03	2.28E-02
Max	-0.3928	-0.3787	0.1793	-0.5952	0.3605	-1.9647
Min	-0.4075	-0.3935	0.1725	-0.6217	0.3530	-2.0437
pro	...	...	...	...	...	...
arg(+)	-0.3921	-0.3778	0.1748	-0.5951	0.3518	-1.9721
lys(+)	-0.3934	-0.3790	0.1767	-0.5958	0.3538	-1.9649
his(+)	-0.3927	-0.3780	0.1751	-0.5956	0.3527	-1.9672
Ave	-0.3927	-0.3783	0.1755	-0.5955	0.3528	-1.9681
asp(−)	-0.3816	-0.3729	0.1425	-0.6120	0.3323	-2.0892
glu(−)	-0.3951	-0.3817	0.1781	-0.5986	0.3556	-1.9793
tyr(−)	-0.3947	-0.3810	0.1792	-0.5965	0.3561	-1.9620
Ave	-0.3905	-0.3785	0.1666	-0.6024	0.3480	-2.0102
lys(+)*Cl*(−)	-0.3943	-0.3800	0.1776	-0.5967	0.3546	-1.9869
asp(−)*Na*(+)	-0.3981	-0.3857	0.1683	-0.6155	0.3520	-2.0867

Table 4.21: $C_\alpha - N$ Bond Critical Point Properties.

Aa	r_A	r_B	ϵ	λ_1	λ_2	λ_3
ala	1.82	0.91	0.0357	-0.4749	-0.4585	0.3683
arg	1.82	0.92	0.0280	-0.4706	-0.4578	0.3408
asn	1.80	0.92	0.0481	-0.5079	-0.4846	0.3303
asp	1.80	0.92	0.0509	-0.5065	-0.4820	0.3310
cys	1.81	0.92	0.0470	-0.4920	-0.4699	0.3478
glu	1.81	0.91	0.0329	-0.4791	-0.4638	0.3550
gln	1.82	0.91	0.0304	-0.4764	-0.4624	0.3608
gly	1.81	0.91	0.0284	-0.4800	-0.4667	0.3673
his	1.82	0.92	0.0318	-0.4727	-0.4581	0.3537
ile	1.82	0.91	0.0391	-0.4670	-0.4494	0.3663
leu	1.82	0.91	0.0372	-0.4706	-0.4537	0.3691
lys	1.82	0.91	0.0297	-0.4698	-0.4562	0.3651
met	1.82	0.91	0.0313	-0.4760	-0.4616	0.3576
phe	1.82	0.91	0.0300	-0.4714	-0.4577	0.3584
ser	1.81	0.91	0.0302	-0.4867	-0.4724	0.3566
thr	1.80	0.92	0.0254	-0.4898	-0.4776	0.3381
trp	1.82	0.91	0.0281	-0.4711	-0.4581	0.3613
tyr	1.82	0.91	0.0287	-0.4702	-0.4571	0.3606
val	1.82	0.91	0.0395	-0.4680	-0.4502	0.3668
Ave	1.81	0.91	0.0343	-0.4790	-0.4630	0.3555
AveDev	6.61E-03	2.51E-03	6.01E-03	9.38E-03	7.97E-03	9.68E-03
σ^2	6.56E-05	9.99E-06	5.56E-05	1.50E-04	1.02E-04	1.52E-04
σ	8.10E-03	3.16E-03	7.45E-03	1.22E-02	1.01E-02	1.23E-02
Max	1.82	0.92	0.0509	-0.4670	-0.4494	0.3691
Min	1.80	0.91	0.0254	-0.5079	-0.4846	0.3303
pro	1.81	0.94	0.0182	-0.4463	-0.4543	0.2755
arg(+)	1.80	0.93	0.0479	-0.5004	-0.4776	0.3128
lys(+)	1.80	0.92	0.0423	-0.5004	-0.4801	0.3310
his(+)	1.78	0.93	0.0590	-0.5240	-0.4948	0.2907
Ave	1.79	0.93	0.0497	-0.5083	-0.4842	0.3115
asp(−)	1.80	0.93	0.0337	-0.4933	-0.4772	0.3096
glu(−)	1.84	0.91	0.0528	-0.4428	-0.4206	0.3973
tyr(−)	1.84	0.90	0.0461	-0.4350	-0.4159	0.3999
Ave	1.83	0.91	0.0442	-0.4570	-0.4379	0.3689
lys(+)*Cl*(−)	1.81	0.91	0.0331	-0.4789	-0.4635	0.3569
asp(−)*Na*(+)	1.81	0.92	0.0399	-0.4293	-0.4737	0.3306

Table 4.22: $C_\alpha - N$ Bond Critical Point Properties *(continued)*.

Aa	σ_1	σ_2	σ_3	tr_σ	ρ_c	$\nabla^2\rho_c$
ala	-0.2757	-0.2704	-0.0485	-0.5946	0.2622	-0.5651
arg	-0.2706	-0.2665	-0.0446	-0.5817	0.2613	-0.5875
asn	-0.2771	-0.2713	-0.0443	-0.5927	0.2698	-0.6622
asp	-0.2775	-0.2709	-0.0444	-0.5928	0.2694	-0.6575
cys	-0.2763	-0.2702	-0.0463	-0.5928	0.2658	-0.6141
glu	-0.2750	-0.2700	-0.0469	-0.5919	0.2635	-0.5879
gln	-0.2754	-0.2705	-0.0476	-0.5935	0.2631	-0.5779
gly	-0.2784	-0.2734	-0.0480	-0.5998	0.2649	-0.5794
his	-0.2739	-0.2688	-0.0466	-0.5893	0.2620	-0.5771
ile	-0.2741	-0.2686	-0.0482	-0.5909	0.2603	-0.5501
leu	-0.2751	-0.2691	-0.0482	-0.5924	0.2611	-0.5553
lys	-0.2746	-0.2696	-0.0480	-0.5922	0.2614	-0.5609
met	-0.2748	-0.2699	-0.0472	-0.5919	0.2629	-0.5800
phe	-0.2739	-0.2691	-0.0474	-0.5904	0.2617	-0.5707
ser	-0.2776	-0.2724	-0.0479	-0.5979	0.2661	-0.6024
thr	-0.2762	-0.2717	-0.0456	-0.5935	0.2675	-0.6294
trp	-0.2743	-0.2694	-0.0478	-0.5915	0.2615	-0.5671
tyr	-0.2740	-0.2693	-0.0477	-0.5910	0.2616	-0.5668
val	-0.2743	-0.2688	-0.0482	-0.5913	0.2605	-0.5515
Ave	-0.2752	-0.2700	-0.0470	-0.5922	0.2635	-0.5865
AveDev	1.33E-03	1.14E-03	1.10E-03	2.11E-03	2.37E-03	2.48E-02
σ^2	3.15E-06	2.38E-06	1.86E-06	1.26E-05	8.56E-06	1.09E-03
σ	1.78E-03	1.54E-03	1.36E-03	3.55E-03	2.93E-03	3.29E-02
Max	-0.2706	-0.2665	-0.0443	-0.5817	0.2698	-0.5501
Min	-0.2784	-0.2734	-0.0485	-0.5998	0.2603	-0.6622
pro	-0.2608	-0.2559	0.0345	-0.4822	0.2603	0.6414
arg(+)	-0.2721	-0.2669	-0.0410	-0.5800	0.2675	-0.6652
lys(+)	-0.2758	-0.2705	-0.0440	-0.5903	0.2683	-0.6460
his(+)	-0.2745	-0.2670	-0.0377	-0.5792	0.2730	-0.7280
Ave	-0.2741	-0.2681	-0.0409	-0.5832	0.2696	-0.6797
asp(−)	-0.2720	-0.2664	-0.0406	-0.5790	0.2669	-0.6609
glu(−)	-0.2736	-0.2661	-0.0511	-0.5908	0.2538	-0.4662
tyr(−)	-0.2721	-0.2656	-0.0515	-0.5892	0.2523	-0.4511
Ave	-0.2726	-0.2660	-0.0477	-0.5863	0.2577	-0.5261
lys(+)*Cl*(−)	-0.2753	-0.2702	-0.0471	-0.5926	0.2635	-0.5855
asp(−)*Na*(+)	-0.2741	-0.2688	-0.0439	-0.5868	0.2662	-0.6356

Table 4.23: $C_\alpha - H_\beta$ Bond Critical Point Properties.

Aa	r_A	r_B	ϵ	λ_1	λ_2	λ_3
ala	1.31	0.72	0.0305	-0.8194	-0.7952	0.4792
arg	1.31	0.72	0.0305	-0.8188	-0.7946	0.4790
asn	1.31	0.72	0.0288	-0.8134	-0.7906	0.4771
asp	1.32	0.71	0.0273	-0.8310	-0.8089	0.4937
cys	1.32	0.71	0.0281	-0.8253	-0.8028	0.4893
glu	1.32	0.72	0.0295	-0.8170	-0.7936	0.4793
gln	1.31	0.72	0.0307	-0.8124	-0.7882	0.4745
gly	1.32	0.71	0.0338	-0.8155	-0.7888	0.4807
his	1.31	0.72	0.0287	-0.8173	-0.7945	0.4797
ile	1.31	0.72	0.0315	-0.8182	-0.7931	0.4775
leu	1.31	0.72	0.0316	-0.8200	-0.7949	0.4789
lys	1.31	0.72	0.0310	-0.8150	-0.7905	0.4765
met	1.31	0.72	0.0300	-0.8160	-0.7922	0.4778
phe	1.31	0.72	0.0298	-0.8195	-0.7957	0.4798
ser	1.32	0.71	0.0315	-0.8211	-0.7960	0.4854
thr	1.31	0.72	0.0291	-0.8159	-0.7928	0.4779
trp	1.31	0.72	0.0310	-0.8181	-0.7937	0.4773
tyr	1.31	0.72	0.0303	-0.8178	-0.7938	0.4778
val	1.31	0.72	0.0316	-0.8173	-0.7923	0.4772
Ave	1.31	0.72	0.0303	-0.8184	-0.7943	0.4799
AveDev	2.21E-03	2.30E-03	1.15E-03	2.79E-03	2.95E-03	3.09E-03
σ^2	9.93E-06	1.09E-05	2.27E-06	1.76E-05	2.22E-05	2.19E-05
σ	3.15E-03	3.31E-03	1.51E-03	4.19E-03	4.72E-03	4.68E-03
Max	1.32	0.72	0.0338	-0.8124	-0.7882	0.4937
Min	1.31	0.71	0.0273	-0.8310	-0.8089	0.4745
pro	1.32	0.71	0.0295	-0.8197	-0.7963	0.4854
arg(+)	1.31	0.72	0.0280	-0.8178	-0.7955	0.4800
lys(+)	1.32	0.72	0.0279	-0.8157	-0.7936	0.4791
his(+)	1.32	0.71	0.0229	-0.8205	-0.8026	0.4864
Ave	1.32	0.72	0.0263	-0.8180	-0.7972	0.4818
asp(−)	1.31	0.73	0.0279	-0.7999	-0.7782	0.4651
glu(−)	1.31	0.72	0.0356	-0.8101	-0.7822	0.4687
tyr(−)	1.31	0.72	0.0350	-0.8177	-0.7900	0.4751
Ave	1.31	0.72	0.0328	-0.8092	-0.7835	0.4696
lys(+)*Cl*(−)	1.31	0.72	0.0302	-0.8151	-0.7912	0.4777
asp(−)*Na*(+)	1.32	0.72	0.0284	-0.8144	-0.7919	0.4790

Table 4.24: $C_\alpha - H_\beta$ Bond Critical Point Properties *(continued)*.

Aa	σ_1	σ_2	σ_3	tr_σ	ρ_c	$\nabla^2\rho_c$
ala	-0.2343	-0.2314	0.1119	-0.3538	0.3001	-1.1354
arg	-0.2342	-0.2310	0.1119	-0.3533	0.3000	-1.1345
asn	-0.2328	-0.2305	0.1109	-0.3524	0.2992	-1.1270
asp	-0.2357	-0.2337	0.1157	-0.3537	0.3016	-1.1463
cys	-0.2346	-0.2323	0.1145	-0.3524	0.3006	-1.1389
glu	-0.2338	-0.2308	0.1119	-0.3527	0.2997	-1.1313
gln	-0.2331	-0.2300	0.1105	-0.3526	0.2990	-1.1260
gly	-0.2328	-0.2299	0.1128	-0.3499	0.2980	-1.1236
his	-0.2340	-0.2309	0.1120	-0.3529	0.2998	-1.1321
ile	-0.2348	-0.2317	0.1112	-0.3553	0.3004	-1.1338
leu	-0.2350	-0.2320	0.1116	-0.3554	0.3006	-1.1360
lys	-0.2335	-0.2304	0.1112	-0.3527	0.2994	-1.1290
met	-0.2337	-0.2306	0.1115	-0.3528	0.2996	-1.1304
phe	-0.2344	-0.2312	0.1121	-0.3535	0.3002	-1.1354
ser	-0.2335	-0.2314	0.1134	-0.3515	0.2996	-1.1317
thr	-0.2339	-0.2309	0.1117	-0.3531	0.2996	-1.1307
trp	-0.2341	-0.2308	0.1117	-0.3532	0.3005	-1.1337
tyr	-0.2342	-0.2309	0.1116	-0.3535	0.3000	-1.1338
val	-0.2345	-0.2314	0.1112	-0.3547	0.3002	-1.1324
Ave	-0.2340	-0.2311	0.1121	-0.3531	0.2999	-1.1327
AveDev	5.61E-04	6.23E-04	8.59E-04	8.70E-04	5.47E-04	3.49E-03
σ^2	5.43E-07	7.59E-07	1.59E-06	1.58E-06	5.67E-07	2.46E-05
σ	7.37E-04	8.71E-04	1.26E-03	1.26E-03	7.53E-04	4.96E-03
Max	-0.2328	-0.2299	0.1157	-0.3499	0.3016	-1.1236
Min	-0.2357	-0.2337	0.1105	-0.3554	0.2980	-1.1463
pro	-0.2347	-0.2314	-0.1139	-0.5800	0.2996	-1.1306
arg(+)	-0.2340	-0.2312	0.1120	-0.3532	0.3000	-1.1334
lys(+)	-0.2337	-0.2310	0.1117	-0.3530	0.2996	-1.1302
his(+)	-0.2347	-0.2320	0.1137	-0.3530	0.3005	-1.1368
Ave	-0.2341	-0.2314	0.1125	-0.3531	0.3000	-1.1335
asp(−)	-0.2305	-0.2281	0.1078	-0.3508	0.2972	-1.1129
glu(−)	-0.2328	-0.2290	0.1092	-0.3526	0.2985	-1.1236
tyr(−)	-0.2341	-0.2302	0.1110	-0.3533	0.2996	-1.1326
Ave	-0.2325	-0.2291	0.1093	-0.3522	0.2984	-1.1230
lys(+)*Cl*(−)	-0.2335	-0.2305	0.1114	-0.3525	0.2993	-1.1287
asp(−)*Na*(+)	-0.2327	-0.2304	0.1152	-0.3479	0.2991	-1.1273

Table 4.25: $C_\alpha - C$ Bond Critical Point Properties.

Aa	r_A	r_B	ϵ	λ_1	λ_2	λ_3
ala	1.37	1.51	0.0665	-0.5510	-0.5167	0.2950
arg	1.36	1.51	0.0674	-0.5633	-0.5278	0.2923
asn	1.39	1.49	0.0629	-0.5528	-0.5201	0.2990
asp	1.37	1.51	0.0639	-0.5515	-0.5184	0.2949
cys	1.39	1.50	0.0637	-0.5471	-0.5143	0.2978
glu	1.38	1.51	0.0652	-0.5486	-0.5151	0.2957
gln	1.37	1.51	0.0677	-0.5489	-0.5141	0.2938
gly	1.38	1.50	0.0694	-0.5540	-0.5180	0.2945
his	1.37	1.51	0.0628	-0.5461	-0.5138	0.2945
ile	1.37	1.52	0.0693	-0.5440	-0.5088	0.2934
leu	1.37	1.51	0.0690	-0.5491	-0.5137	0.2949
lys	1.37	1.51	0.0674	-0.5491	-0.5145	0.2944
met	1.37	1.51	0.0659	-0.5492	-0.5153	0.2949
phe	1.38	1.50	0.0641	-0.5489	-0.5158	0.2965
ser	1.38	1.50	0.0683	-0.5488	-0.5137	0.2944
thr	1.39	1.50	0.0640	-0.5448	-0.5120	0.2971
trp	1.38	1.50	0.0651	-0.5497	-0.5160	0.2960
tyr	1.38	1.50	0.0647	-0.5491	-0.5158	0.2963
val	1.37	1.52	0.0697	-0.5445	-0.5090	0.2932
Ave	1.38	1.51	0.0662	-0.5495	-0.5154	0.2952
AveDev	5.79E-03	5.23E-03	2.03E-03	2.66E-03	2.66E-03	1.27E-03
σ^2	5.24E-05	3.98E-05	5.41E-06	1.81E-05	1.68E-05	2.71E-06
σ	7.24E-03	6.31E-03	2.33E-03	4.25E-03	4.10E-03	1.65E-03
Max	1.39	1.52	0.0697	-0.5440	-0.5088	0.2990
Min	1.36	1.49	0.0628	-0.5633	-0.5278	0.2923
pro	1.37	1.51	0.0969	-0.5574	-0.5082	0.2930
arg(+)	1.37	1.50	0.0636	-0.5612	-0.5277	0.2949
lys(+)	1.38	1.50	0.0627	-0.5474	-0.5151	0.2973
his(+)	1.40	1.49	0.0527	-0.5436	-0.5164	0.3006
Ave	1.38	1.50	0.0597	-0.5507	-0.5197	0.2976
asp(−)	1.33	1.53	0.0697	-0.5574	-0.5211	0.2833
glu(−)	1.35	1.53	0.0758	-0.5502	-0.5114	0.2878
tyr(−)	1.36	1.51	0.0719	-0.5517	-0.5147	0.2197
Ave	1.35	1.53	0.0725	-0.5531	-0.5157	0.2636
lys(+)*Cl*(−)	1.37	1.51	0.0668	-0.5486	-0.5143	0.2943
asp(−)*Na*(+)	1.37	1.51	0.0657	-0.5564	-0.5221	0.2945

Table 4.26: $C_\alpha - C$ Bond Critical Point Properties *(continued)*.

Aa	σ_1	σ_2	σ_3	tr_σ	ρ_c	$\nabla^2\rho_c$
ala	-0.1781	-0.1706	0.0564	-0.2923	0.2672	-0.7727
arg	-0.1830	-0.1758	0.0560	-0.3028	0.2717	-0.7988
asn	-0.1781	-0.1702	0.0571	-0.2912	0.2677	-0.7739
asp	-0.1788	-0.1708	0.0564	-0.2932	0.2675	-0.7750
cys	-0.1764	-0.1686	0.0568	-0.2882	0.2658	-0.7636
glu	-0.1774	-0.1699	0.0565	-0.2908	0.2666	-0.7680
gln	-0.1779	-0.1705	0.0561	-0.2923	0.2667	-0.7692
gly	-0.1789	-0.1713	0.0568	-0.2934	0.2679	-0.7775
his	-0.1769	-0.1692	0.0565	-0.2896	0.2658	-0.7655
ile	-0.1764	-0.1690	0.0561	-0.2893	0.2650	-0.7594
leu	-0.1777	-0.1706	0.0563	-0.2920	0.2666	-0.7679
lys	-0.1777	-0.1703	0.0563	-0.2917	0.2667	-0.7692
met	-0.1777	-0.1702	0.0564	-0.2915	0.2667	-0.7696
phe	-0.1771	-0.1694	0.0568	-0.2897	0.2665	-0.7681
ser	-0.1772	-0.1693	0.0566	-0.2899	0.2663	-0.7681
thr	-0.1757	-0.1682	0.0568	-0.2871	0.2652	-0.7597
trp	-0.1769	-0.1697	0.0570	-0.2896	0.2671	-0.7613
tyr	-0.1772	-0.1695	0.0568	-0.2899	0.2666	-0.7685
val	-0.1765	-0.1691	0.0561	-0.2895	0.2652	-0.7603
Ave	-0.1777	-0.1701	0.0565	-0.2913	0.2668	-0.7693
AveDev	9.40E-04	9.75E-04	2.71E-04	2.01E-03	8.87E-04	5.46E-03
σ^2	2.34E-06	2.54E-06	1.06E-07	1.06E-05	2.09E-06	7.73E-05
σ	1.53E-03	1.59E-03	3.25E-04	3.25E-03	1.45E-03	8.79E-03
Max	-0.1757	-0.1682	0.0571	-0.2871	0.2717	-0.7594
Min	-0.1830	-0.1758	0.0560	-0.3028	0.2650	-0.7988
pro	-0.1770	-0.1734	0.0563	-0.2941	0.2674	-0.7726
arg(+)	-0.1822	-0.1749	0.0563	-0.3008	0.2712	-0.7940
lys(+)	-0.1770	-0.1694	0.0567	-0.2897	0.2663	-0.7652
his(+)	-0.1754	-0.1673	0.0573	-0.2854	0.2652	-0.7593
Ave	-0.1782	-0.1705	0.0568	-0.2920	0.2676	-0.7728
asp(−)	-0.1835	-0.1763	0.0540	-0.3058	0.2703	-0.7951
glu(−)	-0.1791	-0.1718	0.0551	-0.2958	0.2669	-0.7737
tyr(−)	-0.1785	-0.1710	0.0562	-0.2933	0.2673	-0.7748
Ave	-0.1804	-0.1730	0.0551	-0.2983	0.2682	-0.7812
lys(+)*Cl*(−)	-0.1777	-0.1703	0.0562	-0.2918	0.2666	-0.7687
asp(−)*Na*(+)	-0.1803	-0.1724	0.0564	-0.2963	0.2691	-0.7839

Table 4.27: $C = O$ Bond Critical Point Properties.

Aa	r_A	r_B	ϵ	λ_1	λ_2	λ_3
ala	0.75	1.50	0.0477	-1.2199	-1.1644	2.4362
arg	0.76	1.53	0.0413	-1.1368	-1.0917	2.1216
asn	0.75	1.50	0.0480	-1.2262	-1.1700	2.4589
asp	0.75	1.51	0.0267	-1.1863	-1.1555	2.3233
cys	0.75	1.50	0.0513	-1.2250	-1.1652	2.4487
glu	0.75	1.51	0.0486	-1.2191	-1.1626	2.4274
gln	0.75	1.51	0.0446	-1.2124	-1.1607	2.4117
gly	0.75	1.50	0.0513	-1.2264	-1.1666	2.4462
his	0.75	1.51	0.0443	-1.2121	-1.1606	2.4064
ile	0.75	1.51	0.0480	-1.2173	-1.1616	2.4196
leu	0.75	1.51	0.0458	-1.2173	-1.1640	2.4282
lys	0.75	1.50	0.0488	-1.2206	-1.1639	2.4338
met	0.75	1.51	0.0479	-1.2179	-1.1623	2.4272
phe	0.75	1.50	0.0558	-1.2319	-1.1668	2.4673
ser	0.75	1.50	0.0509	-1.2232	-1.1640	2.4405
thr	0.75	1.51	0.0441	-1.2058	-1.1549	2.3696
trp	0.75	1.50	0.0561	-1.2390	-1.1671	2.4513
tyr	0.75	1.50	0.0552	-1.2307	-1.1664	2.4640
val	0.75	1.51	0.0482	-1.2176	-1.1616	2.4234
Ave	0.75	1.51	0.0476	-1.2150	-1.1595	2.4108
AveDev	1.70E-03	3.48E-03	4.09E-03	1.28E-02	8.03E-03	4.45E-02
σ^2	8.89E-06	3.62E-05	4.17E-05	4.82E-04	2.83E-04	6.05E-03
σ	2.98E-03	6.01E-03	6.46E-03	2.19E-02	1.68E-02	7.78E-02
Max	0.76	1.53	0.0561	-1.1368	-1.0917	2.4673
Min	0.75	1.50	0.0267	-1.2390	-1.1700	2.1216
pro	0.75	1.51	0.0336	-1.1951	-1.1565	2.3403
arg(+)	0.76	1.53	0.0357	-1.1227	-1.0841	2.0822
lys(+)	0.75	1.51	0.0459	-1.2138	-1.1605	2.4167
his(+)	0.75	1.51	0.0414	-1.2031	-1.1553	2.1616
Ave	0.76	1.52	0.0410	-1.1799	-1.1333	2.2202
asp(−)	0.75	1.51	0.0387	-1.2067	-1.1617	2.3971
glu(−)	0.75	1.51	0.0429	-1.2113	-1.1615	2.3988
tyr(−)	0.75	1.50	0.0521	-1.2333	-1.1687	2.4631
Ave	0.75	1.51	0.0446	-1.2171	-1.1640	2.4197
lys(+)*Cl*(−)	0.75	1.51	0.0465	-1.2168	-1.1627	2.4242
asp(−)*Na*(+)	0.75	1.50	0.0488	-1.2208	1.1640	2.4220

Table 4.28: $C = O$ Bond Critical Point Properties *(continued)*.

Aa	σ_1	σ_2	σ_3	tr_σ	ρ_c	$\nabla^2\rho_c$
ala	-0.6201	-0.6163	-0.2668	-1.5032	0.4276	0.0520
arg	-0.5803	-0.5749	-0.2372	-1.3924	0.4123	-0.1068
asn	-0.6232	-0.6196	-0.2695	-1.5123	0.4289	0.0628
asp	-0.6048	-0.5952	-0.2538	-1.4538	0.4210	-0.0186
cys	-0.6220	-0.6192	-0.2689	-1.5101	0.4286	0.0585
glu	-0.6192	-0.6158	-0.2660	-1.5010	0.4274	0.0458
gln	-0.6166	-0.6120	-0.2636	-1.4922	0.4260	0.0386
gly	-0.6217	-0.6191	-0.2686	-1.5094	0.4287	0.0531
his	-0.6157	-0.6113	-0.2630	-1.4900	0.4258	0.0337
ile	-0.6180	-0.6144	-0.2650	-1.4974	0.4269	0.0407
leu	-0.6192	-0.6148	-0.2661	-1.5001	0.4272	0.0469
lys	-0.6198	-0.6164	-0.2664	-1.5026	0.4276	0.0492
met	-0.6188	-0.6151	-0.2656	-1.4995	0.4271	0.0470
phe	-0.6240	-0.6227	-0.2704	-1.5171	0.4296	0.0687
ser	-0.6207	-0.6178	-0.2676	-1.5061	0.4281	0.0533
thr	-0.6130	-0.6083	-0.2617	-1.4830	0.4253	0.0090
trp	-0.6239	-0.6221	-0.2699	-1.5159	0.4231	0.0563
tyr	-0.6235	-0.6220	-0.2700	-1.5155	0.4294	0.0669
val	-0.6182	-0.6147	-0.2652	-1.4981	0.4269	0.0443
Ave	-0.6170	-0.6132	-0.2645	-1.4947	0.4262	0.0374
AveDev	5.74E-03	6.79E-03	4.54E-03	1.71E-02	2.48E-03	2.44E-02
σ^2	9.91E-05	1.24E-04	5.84E-05	8.24E-04	1.58E-05	1.65E-03
σ	9.96E-03	1.11E-02	7.64E-03	2.87E-02	3.97E-03	4.06E-02
Max	-0.5803	-0.5749	-0.2372	-1.3924	0.4296	0.0687
Min	-0.6240	-0.6227	-0.2704	-1.5171	0.4123	-0.1068
pro	-0.6104	-0.6034	-0.2585	-1.4723	0.4241	-0.0112
arg(+)	-0.5750	-0.5679	-0.2330	-1.3759	0.4097	-0.1246
lys(+)	-0.6177	-0.6133	-0.2646	-1.4956	0.4265	0.0424
his(+)	-0.6127	-0.6071	-0.2606	-1.4804	0.4242	0.0263
Ave	-0.6018	-0.5961	-0.2527	-1.4506	0.4201	-0.0186
asp(−)	-0.6140	-0.6075	-0.2602	-1.4817	0.4245	0.0287
glu(−)	-0.6147	-0.6101	-0.2621	-1.4869	0.4255	0.0260
tyr(−)	-0.6232	-0.6223	-0.2698	-1.5153	0.4296	0.0609
Ave	-0.6173	-0.6133	-0.2640	-1.4946	0.4265	0.0385
lys(+)*Cl*(−)	-0.6184	-0.6143	-0.2651	-1.4978	0.4268	0.0447
asp(−)*Na*(+)	-0.6197	-0.6160	-0.2661	-1.5018	0.4279	0.0372

Table 4.29: $C - N_1$ Bond Critical Point Properties.

Aa	r_A	r_B	ϵ	λ_1	λ_2	λ_3
ala	0.87	1.70	0.0191	-0.7250	-0.7113	0.5888
arg	0.87	1.69	0.0151	-0.7265	-0.7157	0.5826
asn	0.87	1.69	0.0149	-0.7318	-0.7211	0.5821
asp	0.86	1.68	0.0141	-0.7529	-0.7424	0.6578
cys	0.87	1.69	0.0154	-0.7273	-0.7163	0.5818
glu	0.87	1.69	0.0152	-0.7273	-0.7164	0.5948
gln	0.87	1.69	0.0154	-0.7283	-0.7172	0.6024
gly	0.87	1.69	0.0124	-0.7254	-0.7166	0.5869
his	0.86	1.69	0.0159	-0.7316	-0.7201	0.6148
ile	0.87	1.69	0.0140	-0.7241	-0.7141	0.6038
leu	0.87	1.70	0.0195	-0.7251	-0.7112	0.5950
lys	0.87	1.70	0.0188	-0.7238	-0.7104	0.5927
met	0.87	1.69	0.0160	-0.7269	-0.7154	0.5971
phe	0.87	1.70	0.0202	-0.7195	-0.7052	0.5789
ser	0.87	1.69	0.0153	-0.7276	-0.7166	0.5850
thr	0.86	1.68	0.0102	-0.7411	-0.7336	0.6732
trp	0.86	1.69	0.0198	-0.7179	-0.7043	0.5980
tyr	0.87	1.70	0.0203	-0.7199	-0.7056	0.5807
val	0.87	1.69	0.0139	-0.7240	-0.7141	0.6017
Ave	0.87	1.69	0.0159	-0.7299	-0.7185	0.6028
AveDev	2.50E-03	2.97E-03	2.06E-03	7.16E-03	7.60E-03	1.82E-02
σ^2	1.36E-05	2.25E-05	7.26E-06	1.11E-04	1.24E-04	6.83E-04
σ	3.69E-03	4.75E-03	2.69E-03	1.05E-02	1.11E-02	2.61E-02
Max	0.87	1.70	0.0203	-0.7195	-0.7052	0.6732
Min	0.86	1.68	0.0102	-0.7529	-0.7424	0.5789
pro	0.86	1.68	0.0116	-0.7435	-0.7350	0.6656
arg(+)	0.87	1.69	0.0107	-0.7393	-0.7315	0.5931
lys(+)	0.87	1.69	0.0142	-0.7336	-0.7233	0.5753
his(+)	0.87	1.68	0.0195	-0.7552	-0.7407	0.5906
Ave	0.87	1.68	0.0148	-0.7427	-0.7318	0.5863
asp(−)	0.86	1.70	0.0213	-0.7190	-0.7040	0.6236
glu(−)	0.86	1.70	0.0212	-0.7191	-0.7042	0.6500
tyr(−)	0.87	1.71	0.0269	-0.7089	-0.6903	0.6127
Ave	0.86	1.70	0.0232	-0.7157	-0.6995	0.6288
lis(+)*Cl*(−)	0.87	1.69	0.0182	-0.7259	-0.7129	0.5877
asp(−)*Na*(+)	0.86	1.69	0.0083	-0.7254	-0.7194	0.6338

Table 4.30: $C - N_1$ Bond Critical Point Properties *(continued)*.

Aa	σ_1	σ_2	σ_3	tr_σ	ρ_c	$\nabla^2\rho_c$
ala	-0.3561	-0.3357	-0.0939	-0.7857	0.3240	-0.8475
arg	-0.3590	-0.3391	-0.0957	-0.7938	0.3254	-0.8440
asn	-0.3577	-0.3388	-0.0942	-0.7907	0.3267	-0.8709
asp	-0.3727	-0.3540	-0.1043	-0.8310	0.3330	-0.8375
cys	-0.3561	-0.3369	-0.0937	-0.7867	0.3253	-0.8618
glu	-0.3577	-0.3380	-0.0949	-0.7907	0.3252	-0.8489
gln	-0.3589	-0.3388	-0.0958	-0.7935	0.3254	-0.8431
gly	-0.3558	-0.3365	-0.0937	-0.7860	0.3246	-0.8552
his	-0.3609	-0.3410	-0.0974	-0.7993	0.3262	-0.8369
ile	-0.3577	-0.3378	-0.0954	-0.7908	0.3242	-0.8345
leu	-0.3569	-0.3362	-0.0946	-0.7877	0.3240	-0.8412
lys	-0.3562	-0.3359	-0.0942	-0.7863	0.3237	-0.8415
met	-0.3578	-0.3378	-0.0951	-0.7907	0.3249	-0.8452
phe	-0.3532	-0.3332	-0.0923	-0.7787	0.3224	-0.8458
ser	-0.3566	-0.3373	-0.0939	-0.7878	0.3253	-0.8593
thr	-0.3694	-0.3518	-0.1045	-0.8257	0.3297	-0.8016
trp	-0.3544	-0.3337	-0.0925	-0.7806	0.3267	-0.8401
tyr	-0.3535	-0.3334	-0.0925	-0.7794	0.3225	-0.8448
val	-0.3574	-0.3375	-0.0951	-0.7901	0.3242	-0.8364
Ave	-0.3583	-0.3386	-0.0955	-0.7924	0.3254	-0.8440
AveDev	3.09E-03	3.34E-03	2.15E-03	8.46E-03	1.59E-03	8.75E-03
σ^2	2.39E-05	2.94E-05	1.15E-05	1.87E-04	6.03E-06	1.95E-04
σ	4.89E-03	5.42E-03	3.39E-03	1.37E-02	2.46E-03	1.40E-02
Max	-0.3532	-0.3332	-0.0923	-0.7787	0.3330	-0.8016
Min	-0.3574	-0.3540	-0.1045	-0.8310	0.3224	-0.8709
pro	-0.3694	-0.3497	-0.1034	-0.8225	0.3297	-0.8128
arg(+)	-0.3623	-0.3438	-0.0967	-0.8028	0.3296	-0.8777
lys(+)	-0.3576	-0.3387	-0.0935	-0.7897	0.3272	-0.8816
his(+)	-0.3644	-0.3477	-0.0972	-0.8093	0.3330	-0.9052
Ave	-0.3614	-0.3434	-0.0958	-0.8006	0.3299	-0.8882
asp(−)	-0.3603	-0.3374	-0.0973	-0.7950	0.3227	-0.7994
glu(−)	-0.3608	-0.3388	-0.0993	-0.7989	0.3219	-0.7733
tyr(−)	-0.3532	-0.3315	-0.0945	-0.7792	0.3185	-0.7866
Ave	-0.3581	-0.3359	-0.0970	-0.7910	0.3210	-0.7864
lis(+)*Cl*(−)	-0.3565	-0.3362	-0.0939	-0.7866	0.3244	-0.8510
asp(−)*Na*(+)	-0.3635	-0.3438	-0.0998	-0.8071	0.3263	-0.8110

4.4 Transferability of $|Aa|$ Atomic Properties

We commence our study of the properties of the $|Aa|$ by presenting a survey of the numerical errors produced upon computing the atomic properties of these residues. These errors are characterized according to the following criteria:

1. *The difference between the electron population of the complete mold and the sum of the integrated populations over the atomic basins in the same mold.* This difference is computed as

$$\Delta N = N_{mold} - \sum_{i}^{M} N(\Omega_i) \quad , \tag{4.1}$$

where the subscript i indexes the individual atomic basins and M is the total number of atoms. Table 4.32 indicates that the computation of the electron populations over all the $Gly'|Aa|Gly''$ systems produced a small error, considering the fact that N_{mold} varies from 72 to 141 electrons. For example, the relative error in triptophan - the ratio $\frac{\Delta N}{N}$ - is only .002%. The average of computing eqn.[4.1] over the systems studied in this thesis was calculated to be -.0030 e, with a corresponding standard deviation of 8.89E-3. These are small figures and fall within the numerical errors of computing the corresponding atomic integrations.

2. *The difference between the energy obtained in the SCF procedure which produced the wavefunction for the mold and the sum of the atomic contributions to the molecular energy.* This deviation is represented as

$$\Delta E = E_{SCF} - \sum_{i}^{M} E(\Omega_i) \quad , \tag{4.2}$$

where the subscript i references to each and every one of the atoms in the mold. Table 4.32 shows that eqn.[4.2] averages to -.0016 au or 1.004 kcal/mol for the neutral series, with a standard deviation of 2.25E-3. The cationic residues averaged a deviation of .0002 au (.1255 kcal/mol) and the anionic ones .0009 au, or .5647 kcal/mol. These low values are indicative of the high level of accuracy of the atomic integrations.

3. *The evaluation of the integrated Laplacian of the electron density over the atoms in the* $|Aa|$. The quantity

$$L = \sum_{i}^{M} |L(\Omega_i)| \quad , \tag{4.3}$$

is the sum of all the atomic contributions $L(\Omega_i)$ to the residue's Laplacian. It was indicated in section 2.8.2 that each $L(\Omega_i)$ should evaluate to zero; the direct consequence of Gauss' theorem for a system bounded by surfaces of zero-flux. Thus, the extent of the deviation is the extent to which the integration procedure failed to

(a) calculate the properties of the i-ith atom,

(b) map the distribution of this atom's bounding surfaces of zero-flux or

(c) a combination of (a) and (b).

Table 4.32 shows the values obtained for L for the various $|Aa|$, which average to 0.0049 au. In fact, all the residues exhibited integrated values for L in the 10^{-3} region. The mean of L in the neutral series corresponds to 3.075 Kcal/mol in energetic terms. These are good results, considering that these residues contain between 7 and 24 atoms[2] .

Having understood the validity of the data presented in table 4.32 as accurate, within numerical errors, we continue our study on the transferability of the $|Aa|$ with the analysis of the properties of the three units in the glycine mold, namely Gly', $|Aa|$ and Gly''.

Table 4.31: Summary of Free Amino Acid backbone integration data (in au).

Atom	q	E	Vol(0.001)
H	0.373	-0.478	30.6
N	-1.167	-54.910	112.9
C_α	0.585	-37.493	42.3
H_β	0.002	-0.625	43.4
C	1.775	-36.702	30.9
O	-1.358	-75.677	135.1

The first feature drawn from the data presented is the extent to which the $|Aa|$ with non-ionized side-chains remain non-charged when in a protein. As it can be seen on table 4.32, the charge q of the $|Aa|$ averages to 0.0023 e. Similarly, the residues with ionized side-chains exhibit preservation of charge, i.e. the charge of the side chain produces little polarization on the bounding glycines of the mold. Nonetheless there is a measurable polarization on the bounding glycines, which absorb/donate charge to stabilize the middle amino acid residue. Thus, the residues $arg(+)$, $lys(+)$,

[2]The author firmly believes that atomic integrations could be improved with the deployment of better numerical techniques, an area of research pursuit by some of the members of our group.

exhibit a loss of charge of 0.035 *e* to the neighboring glycines. Similarly $asp(-)$, and $glu(-)$ shows a loss of 0.070 *e* as well. This *Le-Chatelier*-like behavior is a mean that the residue uses to gain stabilization, and is deployed in conjunction with the rearrangement of the geometry of its side chain as indicated in the preceding section. Nonetheless, because the charged $|Aa|$ are not likely to remain unstabilized *in vivo*, we conducted two test cases where counter-ions were placed in the neighborhood of the residue which exhibited the greatest error in the computation of its properties. This was performed in a bid to verify if the ionized residues exhibited the same transferable behavior as their non-charged counterparts when in the presence of a counter-ion. The $Gly'| \leftrightarrow |Aa| \leftrightarrow |Gly''$ charge-transfer effects described above disappeared when a counter-ion is placed closed to the source of the charge in these chains, as exemplified by the integrated charge of 0.0015 *e* for $|lys(+)|Cl(-)$ and 0.0023 *e* for $|asp(-)|Na(+)$. Thus, by complying with the pre-requisite that the $|Aa|$ need bear a zero net charge (Popelier and Bader 1994), the amino acid residues in proteins exhibit transferability.

The second important feature highlighted in table 4.32 is the equal and opposite charges of Gly′and Gly″. This result is complementary to the observation of neutrality of the $|Aa|$. Since the charge for a mold is the sum of the contributions to the charge by its components, that is

$$q_{Gly'} + q_{|Aa|} + q_{Gly''} = \text{Total charge of the Mold} \quad , \tag{4.4}$$

if the molecular charge as well as that of the residue is zero, then

$$q_{Gly'} = -q_{Gly''} \quad . \tag{4.5}$$

This is indeed the observation one makes from examining table 4.32. Gly′bears on average a +0.50 au whereas Gly″bears the negative of that number (columns $q_{Gly'}$ and $q_{Gly''}$ in table 4.32). It is unclear at this stage whether the quantity "0.50" bears physical significance, but it is particularly interesting that this value should be recovered. A recent study by Bader and Matta (Bader and Matta n.d.) analyzed the polarization effects in polymers of the form $CH_2O|CH_2O|_nCH_2O$. The authors calculated the charge of the $CH_2O|$ end-group to be -0.642 e while that of $|CH_2O$ was calculated to be +0.652 e. The inner groups $|CH_2O|$ are non-charged. The difference in the net charges of the end fragments is the result of the transfers through the $|CH_2O|$ surfaces such that the flux in the electric field through one surface is equal in magnitude and opposite in direction to the flux through the second. This physical effect renders both surfaces nearly identical in shape and properties, the difference being commensurate with the difference in charge of the $|CH_2O|$ fragment from zero. Surfaces which exhibit this behavior are said to *complement* each other. It seems plausible to think that these charge transfers arise from the ability of each of the fragments to polarize its neighbors, or in turn of being polarized by them.

At a more detailed level, the atomic properties of the backbone atoms show a great transferability throughout the $|Aa|$ series. If fluctuations in the specific properties of an atom are exhibited its neighboring atoms show compensatory changes in the same properties. This observation is more acutely observed in those systems which exhibited internal hydrogen bonding; despite its presence, the net properties of the $|Aa|$ are close to the mean for the entire residue series for the same properties.

Tables 4.33, 4.34, 4.35, 4.36, 4.37 and 4.38 present the integrated values for $H, N, C_\alpha, H_\beta, C$ and O, the backbone atoms in the $|Aa|$. The properties of these atoms are to be compared with the data reported for the integrations of the equivalent free amino acid analogs (Matta and Bader n.d.), which values are shown in table 4.31. The following conclusions are drawn from examining these data:

- H is polarized by the backbone donating on average 0.5543 e to the backbone. The integrated values for this atom in the $|Aa|$ series (table 4.33), shows one of the highest standard deviations in its properties, the result of the relative ease in being polarized, as well as the result of the statistical tinting produced by the contributions of the residues which exhibited internal hydrogen bonding, namely *asn*, *asp* and *asp*$(-)$. Nonetheless, the standard deviations for the charge, and in particular for the energy, are low. It is interesting to notice the physical effects imparted on H when participating in hydrogen bonding. As we can see in table 4.33, the charge for the H atoms in *asn* and *asp* are more positive than the mean of the neutral series. This is paralleled with a drop in the energy and volume of these atoms, again with respect to their corresponding means. This trend indicates that hydrogen-bonding further polarizes the hydrogen in the backbone, further withdrawing electron density from it. This effect is taken to an extreme in *asp*$(-)$ where the atom exhibits a 0.5568 e charge, and a volume 41 % smaller than in *asp*.

  Comparing H properties against the values of the corresponding atomic integrations in the free amino acids reveals that H in the $|Aa|$ is slightly less electron-rich, by an amount of 0.0727 e. Similarly its energy and volumes are higher in the free amino acids than in the mold.

- N (table 4.34), like H, also shows changes in its properties in *asn*, *asp* and *asp*$(-)$ as a result of its participation in internal hydrogen bonding on these residues. The atom bears an average charge of -1.507 e which increases in the hydrogen-bonded species. Thus, in the neutral series, N exhibits the highest values for the charge in *asn* and *asp*, with -1.519 e and -1.514 e respectively. These increases are accompanied by increases in energy (more negative)

and reduced volumes. In the anionic species, $asp(-)$ shows a greater increase in charge, -1.582 e, with a smaller volume. Interestingly, the volume of this atom seems greatly affected by the bulkiness of the $|Aa|$ side groups. On this account *gly* exhibits the largest volume in the series, closely followed by *ala*. The atom exhibits sp^2 character as the negative eigenvector of its diagonalized Quadrupolar Polarization matrix indicates a the presence of non-bonded π density perpendicular to the amidic plane.

Charge transfers along the back-bone play a very important role in stabilizing this atom, judging by the increase of 0.34 e when N is part of a polypeptide chain as opposed to its unlinked analogs. By virtue of being more electro-negative than each of the atoms it is bonded to in the $|Aa|$, this gain in electron density is likely to be contributed by all the atoms N is bonded to. This translates into fluxes in the electric field through the $N|H$, $Gly'|$ and $N - C_\alpha$ surfaces.

- The atom with the highest variations in its properties is C_α (table 4.35) as it would expected. Such observation can be rationalized on the basis that C_α is the atom in the backbone which also belongs to the side-chain in the $|Aa|$, and thus the atom most susceptible to feel the electronic and steric effects of the remainder of the R group as well as those of the backbone. On this account, the range in the energy obtained for this atom in the neutral series is 0.042 au, or 26 kcal/mol. Similarly, the range in the charge is 0.066 e, exhibiting a maximum in *gly*. Glycine also exhibited the largest volume for this atom in the neutral series as well, thus highlighting the great dependency on the steric effects imparted by the side-chain.

  Upon examining the anionic series data, one notices the deviation from the mean of 0.045 exhibited by $asp(-)$. Such difference is most likely to be the result of the polarization induced by the R group on C_α withdrawing electron density from the later. Aspartate has a very short side chain with two lone oxygens at C_γ, thus being highly electron-withdrawing as well as at a very short distance from C_α. The volume for this atom in aspartate does not show a significant change from its counterparts in the rest of the anionic series.

  C_α is on the average slightly more electron-rich in the $|Aa|$ than in the free residues indicating that there is a charge transfer from the backbone into this atom. This is rationalized on the basis that the side-chains are identical in both sets of calculations, so the difference must be accountable to the extended backbone chain. The volumes and energies remain essentially the same for both sets.

- H_β, table 4.36, is not as susceptible to charge transfers away from it as its backbone counterpart

H. Its charge is essentially zero, even in the charged residues. The standard deviations calculated for all the series are extremely small. In fact there are no differences between the properties measured in the free residues and their counterparts in the $|Aa|$. This atom seems unaffected when it is part of a protein. Its role is thus likely to stabilize C_α.

- C, table 4.37, is heavily polarized by the more electro-negative keto oxygen. This can be confirmed by the marked difference in charge between C_α and C, making the later 1.187 e poorer over the former in the neutral series. The lowest values for the charge are observed in arg, in the neutral series, $arg(+)$, in the cationic series and $asp(-)$ in the anionic. The significantly lower values than the mean of their respective series are rationalized for arginine and its cationic analog on the basis of its bonding to the keto oxygen, which in turn is hydrogen-bonded to $C_\gamma H_1$ in the side-chain. With regards to aspartate, the low charge measured is likely to be the result of large polarizations induced by both the R-group, as well as by the keto-oxygen, both bonded to C at the latter's α position. Otherwise, C shows a strong constancy in its properties, denoting a large potential for transfer.

- O is one of the least afflicted to changes in the side chain. This atom's properties are shown in table 4.38. As anticipated, there is a slight decrease in the charge of the keto oxygen in arg and $arg(+)$ with respect to the means of their respective series, the result of these residue's exhibition of internal hydrogen bonding with the side-chain, but generally, this atom's properties remain virtually constant in the $|Aa|$. This is principally due to this atom's high electro-negative nature, which ensures O an almost constant distribution of its electron density. In fact its properties in the free residues are essentially identical as in the $|Aa|$, although O is slightly more charged in the former. Otherwise, this atom seems virtually unaffected when in a free amino acids or in a polypeptide.

The study of the atomic properties of the $|Aa|$ has revealed the extent to which these residues, as a function of the small variations in their respective atomic properties, show transferability of their properties. When significant variations from the mean of the data took place, these have been traced to internal hydrogen bonding. But even in the presence of these internal secondary effects, the $|Aa|$ preserved their charge as a whole, indicating that the residues exhibit an equilibrium-like behavior in a bid to maintain chemical identity. Such observation clearly solidifies the concept of the amino acid residues as transferable pieces.

Table 4.32: Summary of $Gly'|Aa|Gly''$ Fragment Properties.

Aa	$q_{Gly'}$	$q_{Gly''}$	q	L	$E_{\|Aa\|}$	E_{SCF}	ΔE	ΔN	$\|\mu\|_{origin}$	Vol(0.001
ala	.4999	-.5032	.0054	.0010	-245.9147	-735.709	.0015	.0022	2.5199	582.5061
arg	.5093	-.5011	.0030	.0009	-527.0587	-1016.803	.0034	.0113	3.4599	1317.2392
asn	.5027	-.4983	-.0034	.0040	-413.7548	-903.543	.0000	.0010	2.6162	851.2789
asp	.5090	-.5033	-.0112	.0011	-433.5962	-923.379	-.0015	-.0055	2.4519	804.6949
cys	.5056	-.5011	-.0088	.0010	-643.5299	-1133.240	-.0005	-.0043	3.0917	774.3580
glu	.5024	-.5051	-.0062	.0049	-472.6318	-962.426	-.0029	-.0090	3.0266	967.2371
gln	.4989	-.5058	.0048	.0025	-452.7871	-942.560	.0013	-.0020	2.8375	1006.6108
gly	.5030	-.5118	.0059	.0045	-206.8746	-696.663	.0004	-.0029	2.9061	427.1171
his	.5032	-.5084	.0094	.0051	-469.6203	-959.424	.0007	.0042	2.9336	1056.1474
ile	.4999	-.5147	-.0098	.0196	-363.0186	-852.835	-.0056	-.0246	2.4734	1032.8572
leu	.4988	-.5131	.0056	.0021	-363.0345	-852.836	.0004	-.0087	2.4819	1038.0707
lys	.5050	-.5094	.0089	.0003	-418.0802	-907.881	.0009	.0045	2.3503	1157.2417
met	.5056	-.5100	.0083	.0072	-721.6067	-1211.327	.0013	.0040	2.5933	1081.5564
phe	.5035	-.5101	.0017	.0051	-475.4804	-965.299	.0006	-.0049	2.4999	1249.4375
ser	.5013	-.5051	-.0066	.0071	-320.8042	-810.585	-.0016	-.0105	3.0952	648.4262
thr	.4996	-.5034	.0137	.0119	-359.8309	-849.636	.0029	.0099	2.6018	759.4194
trp	.5045	-.5169	.0072	.0023	-606.4034	-1095.876	-.0012	-.0052	3.4773	1506.2727
tyr	.4960	-.5166	.0026	.0068	-550.3702	-1040.102	-.0038	-.0180	1.9784	1312.4016
val	.4986	-.5112	.0139	.0057	-323.9855	-813.793	.0002	.0013	2.5074	878.6172
Ave	.5025	-.5079	.0023	.0049			-.0016	-.0030		
AveDev	.0028	.0046	.0064	.0031			.0017	.0067		
σ^2	1.27E-5	2.97E-5	6.14E-5	2.13E-5			5.05E-6	7.91E-5		
σ	3.57E-3	5.45E-3	7.83E-3	4.61E-3			2.25E-3	8.89E-3		
Max	.5093	-.4983	.0139	.0197			.0035	.0113		
Min	.4960	-.5169	-.0112	.0003			-.0056	-.0246		
pro	.4740	-.5043	.0326	.0002	-322.7813	-812.618	-.0008	.0024	1.8129	778.1557
arg(+)	.5206	-.4895	.9655	.0023	-527.4783	-1017.218	.0007	.9966	10.6361	1284.2054
lys(+)	.5164	-.4913	.9682	.0024	-418.4498	-908.252	-.0012	.9933	11.8803	1127.3169
his(+)	.5381	-.4723	1.0035	.0042	-959.8117	-959.810	.0011	1.0035	12.2654	2075.1119
Ave	.5250	-.4844	.9791	.0030			.0002	.9978		
asp(−)	.4510	-.5202	-.9207	.0062	-433.1112	-922.839	.0028	-.9899	7.2315	818.7943
glu(−)	.4670	-.5244	-.9394	.0031	-472.0724	-961.864	.0007	-.9969	8.1575	994.4368
tyr(−)	.4710	-.5309	-.9981	.0031	-1039.606	-1039.606	.0010	-.9998	10.9421	2417.828
Ave	.4630	-.5252	-.9528	.0042			.0009	-.9955		
lys(+)*Cl*(−)	.5046	-.5085	.0014	.0018	-878.267	-1367.989	.0004	-.0025	5.9152	1420.7070
asp(−)*Na*(+)	.4839	-.5116	.0023	.0037	-707.337	-1084.704	.0112	-.0017	4.2060	1031.7180

Table 4.33: Common Back-Bone Atomic Integration: Atom H

Aa	q	E	K	$\|\mu\|$	Q_1	Q_2	Q_3	$Vol.$	$\|L\|$
ala	0.4390	-0.4453	0.4450	0.1638	-0.0796	-0.0646	0.0790	25.9590	3.00E-05
arg	0.4480	-0.4400	0.4396	0.1614	-0.0618	-0.0216	0.0832	25.2380	1.00E-04
asn	0.4754	-0.4266	0.4264	0.1492	-0.0679	0.0029	0.0650	23.1630	1.00E-04
asp	0.4617	-0.4332	0.4330	0.1552	-0.0642	0.0022	0.0619	24.6400	7.00E-05
cys	0.4486	-0.4400	0.4399	0.1609	-0.0616	-0.0081	0.0697	25.6380	1.00E-04
glu	0.4407	-0.4444	0.4441	0.1632	-0.0689	-0.0158	0.0798	25.8480	1.80E-04
gln	0.4430	-0.4430	0.4428	0.1625	-0.0632	-0.0131	0.0763	25.6760	8.00E-05
gly	0.4439	-0.4439	0.4436	0.1613	-0.2356	-0.0596	0.0832	25.4680	9.00E-05
his	0.4408	-0.4445	0.4442	0.1631	-0.0641	-0.0162	0.0757	25.7810	1.20E-04
ile	0.4384	-0.4472	0.4470	0.1611	-0.0789	0.0055	0.0734	24.7770	9.00E-05
leu	0.4389	-0.4464	0.4461	0.1627	-0.0658	-0.0057	0.0715	25.8820	1.60E-04
lys	0.4383	-0.4457	0.4455	0.1641	-0.0648	-0.0135	0.0783	25.9680	1.10E-04
met	0.4410	-0.4441	0.4440	0.1632	-0.0640	-0.0143	0.0782	25.8250	8.00E-05
phe	0.4386	-0.4455	0.4453	0.1639	-0.0652	-0.0138	0.0790	25.8100	6.00E-05
ser	0.4496	-0.4406	0.4404	0.1591	-0.0608	0.0035	0.0573	25.6540	1.20E-04
thr	0.4462	-0.4323	0.4321	0.1518	-0.0718	0.0132	0.0585	22.8950	1.40E-04
trp	0.4387	-0.4458	0.4454	0.1633	-0.0650	-0.0135	0.0785	25.7800	1.20E-04
tyr	0.4386	-0.4459	0.4454	0.1634	-0.0651	-0.0135	0.0786	25.7850	1.00E-04
val	0.4386	-0.4468	0.4466	0.1612	-0.0786	0.0050	0.0736	24.9000	3.00E-05
Ave	0.4457	-0.4422	0.4419	0.1608	-0.0762	-0.0127	0.0737	25.4046	9.89E-05
AveDev	7.94E-03	4.24E-03	4.22E-03	2.92E-03	1.77E-02	1.26E-02	6.20E-03	6.37E-01	2.75E-05
σ^2	1.15E-04	3.19E-05	3.16E-05	1.76E-05	1.52E-03	3.95E-04	6.16E-05	8.63E-01	1.45E-09
σ	1.07E-02	5.65E-03	5.63E-03	4.19E-03	3.90E-02	1.99E-02	7.85E-03	9.29E-01	3.81E-05
Max	0.4754	-0.4266	0.4470	0.1641	-0.0608	0.0132	0.0832	25.9680	1.80E-04
Min	0.4383	-0.4472	0.4264	0.1492	-0.2356	-0.0646	0.0573	22.8950	3.00E-05
pro	...	...	...	...	...	...	...	...	...
arg(+)	0.4538	-0.4361	0.4357	0.1600	-0.0603	-0.0292	0.0899	25.0340	4.00E-05
lys(+)	0.4425	-0.4426	0.4424	0.1631	-0.0633	-0.0234	0.0867	25.7100	4.00E-05
his(+)	0.4482	-0.4392	0.4389	0.1614	-0.0620	-0.0279	0.0900	25.3060	2.60E-04
Ave	0.4482	-0.4393	0.4390	0.1615	-0.0619	-0.0268	0.0889	25.3500	1.13E-04
asp(−)	0.5568	-0.3801	0.3799	0.1159	-0.0912	-0.0710	1.6220	14.5190	7.30E-04
glu(−)	0.4401	-0.4453	0.4451	0.1634	-0.0658	0.0015	0.0643	25.7740	1.20E-04
tyr(−)	0.4315	-0.4498	0.4496	0.1670	-0.0689	0.0005	0.0683	26.1814	1.30E-04
Ave	0.4992	-0.4117	0.4115	0.1421	-0.0563	-0.0468	0.5891	20.3940	3.27E-04
lys(+)*Cl*(−)	0.4411	-0.4441	0.4439	0.1631	-0.0633	-0.0153	0.0786	25.8970	1.10E-04
asp(−)*Na*(+)	0.4938	-0.4178	0.4177	0.1407	-0.0732	-0.0298	0.1009	21.0075	2.00E-04

Table 4.34: Common Back-Bone Atomic Integration: Atom N

Aa	q	E	K	$\|\mu\|$	Q_1	Q_2	Q_3	$Vol.$	$\|L\|$
ala	-1.509	-55.325	55.300	0.230	-2.130	0.759	1.371	96.199	4.51E-03
arg	-1.503	-55.348	55.299	0.227	-2.103	0.632	1.471	92.524	2.90E-04
asn	-1.519	-55.347	55.325	0.224	-2.078	0.669	1.409	92.323	8.40E-04
asp	-1.514	-55.337	55.315	0.216	-2.126	0.652	1.474	92.799	7.20E-04
cys	-1.509	-55.325	55.309	0.224	-2.106	0.651	1.455	93.153	2.60E-04
glu	-1.503	-55.326	55.301	0.225	-2.188	0.698	1.490	94.180	1.00E-05
gln	-1.506	-55.333	55.309	0.257	-2.114	0.751	1.365	95.252	2.45E-03
gly	-1.501	-55.314	55.288	0.218	-2.191	0.652	1.538	97.676	4.00E-04
his	-1.501	-55.330	55.304	0.232	-2.166	0.699	1.467	93.628	5.30E-04
ile	-1.506	-55.330	55.304	0.227	-2.184	0.669	1.515	93.552	3.10E-04
leu	-1.504	-55.335	55.309	0.236	-2.010	0.578	1.432	90.943	9.00E-05
lys	-1.503	-55.328	55.302	0.228	-2.182	0.700	1.482	94.085	5.10E-04
met	-1.504	-55.318	55.301	0.225	-2.183	0.703	1.480	94.426	3.50E-04
phe	-1.508	-55.329	55.300	0.231	-2.119	0.762	1.357	94.975	4.02E-03
ser	-1.508	-55.334	55.309	0.258	-1.986	0.744	1.241	93.955	7.36E-03
thr	-1.510	-55.337	55.312	0.237	-2.131	0.771	1.361	93.266	4.00E-05
trp	-1.508	-55.327	55.299	0.235	-2.131	0.770	1.361	95.057	8.10E-03
tyr	-1.508	-55.327	55.300	0.234	-2.130	0.770	1.360	95.061	4.80E-03
val	-1.506	-55.329	55.302	0.226	-2.190	0.680	1.510	93.941	7.00E-05
Ave	-1.507	-55.330	55.305	0.231	-2.129	0.700	1.428	94.052	1.88E-03
AveDev	3.26E-03	6.13E-03	5.89E-03	7.55E-03	4.09E-02	4.48E-02	6.34E-02	1.10E+00	2.10E-03
σ^2	1.98E-05	7.05E-05	6.18E-05	1.19E-04	3.33E-03	3.09E-03	5.74E-03	2.26E+00	6.71E-06
σ	4.45E-03	8.40E-03	7.86E-03	1.09E-02	5.77E-02	5.56E-02	7.57E-02	1.50E+00	2.59E-03
Max	-1.501	-55.314	55.325	0.258	-1.986	0.771	1.538	97.676	8.10E-03
Min	-1.519	-55.348	55.288	0.216	-2.191	0.578	1.241	90.943	1.00E-05
pro	-1.545	-55.363	55.338	0.208	-1.660	0.748	0.911	78.038	2.50E-04
arg(+)	-1.503	-55.348	55.299	0.227	-2.125	0.577	1.549	93.852	2.90E-04
lys(+)	-1.500	-55.320	55.293	0.207	-2.209	0.658	1.552	94.895	6.10E-04
his(+)	-1.491	-55.312	55.285	0.198	-2.207	0.633	1.574	94.734	8.60E-04
Ave	-1.498	-55.327	55.292	0.211	-2.181	0.622	1.558	94.494	5.87E-04
asp(−)	-1.582	-55.418	55.398	0.222	-1.842	0.805	1.037	91.979	8.90E-04
glu(−)	-1.506	-55.333	55.309	0.257	-2.128	0.780	1.348	93.877	2.45E-03
tyr(−)	-1.506	-55.333	55.306	0.258	-2.117	0.797	1.320	93.920	2.03E-03
Ave	-1.531	-55.361	55.338	0.246	-2.029	0.794	1.235	93.259	1.79E-03
lys(+)*Cl*(−)	-1.503	-55.320	55.303	0.224	-2.170	0.682	1.488	94.169	1.20E-04
asp(−)*Na*(+)	-1.528	-55.364	55.341	0.233	-2.003	0.701	1.302	91.862	1.30E-04

Table 4.35: Common Back-Bone Atomic Integration: Atom C_α

Aa	q	E	K	$\|\mu\|$	Q_1	Q_2	Q_3	$Vol.$	$\|L\|$
ala	0.551	-37.492	37.475	0.580	-0.771	-0.402	1.173	41.145	5.30E-04
arg	0.544	-37.523	37.489	0.584	-0.751	-0.385	1.136	40.257	2.85E-03
asn	0.568	-37.483	37.468	0.568	-0.660	-0.471	1.296	40.238	4.10E-04
asp	0.571	-37.481	37.466	0.571	-0.757	-0.356	1.112	40.601	3.33E-03
cys	0.567	-37.486	37.475	0.562	-0.754	-0.385	1.139	39.445	2.88E-03
glu	0.546	-37.497	37.481	0.582	-0.762	-0.384	1.146	40.388	8.10E-04
gln	0.547	-37.501	37.483	0.594	-0.784	-0.377	1.161	40.121	5.18E-03
gly	0.590	-37.483	37.466	0.561	-0.737	-0.416	1.154	41.222	7.80E-03
his	0.541	-37.495	37.477	0.596	-0.756	-0.374	1.130	41.108	1.02E-03
ile	0.524	-37.510	37.492	0.603	-0.901	-0.314	1.215	40.133	1.70E-03
leu	0.534	-37.505	37.487	0.587	-0.779	-0.425	1.204	40.417	2.52E-03
lys	0.538	-37.503	37.485	0.591	-0.803	-0.367	1.170	40.500	4.10E-04
met	0.545	-37.493	37.482	0.589	-0.782	-0.375	1.157	40.357	1.95E-03
phe	0.541	-37.499	37.480	0.588	-0.750	-0.405	1.155	40.797	6.40E-04
ser	0.562	-37.511	37.495	0.574	-0.761	-0.361	1.219	39.914	2.00E-03
thr	0.556	-37.519	37.503	0.567	-0.721	-0.393	1.113	40.297	3.37E-03
trp	0.542	-37.500	37.481	0.590	-0.771	-0.375	1.146	40.600	7.90E-04
tyr	0.541	-37.501	37.482	0.591	-0.771	-0.375	1.146	40.646	1.51E-03
val	0.525	-37.509	37.491	0.602	-0.913	-0.293	1.207	40.219	1.35E-03
Ave	0.549	-37.500	37.482	0.583	-0.773	-0.381	1.167	40.442	2.16E-03
AveDev	1.29E-02	9.02E-03	7.26E-03	1.06E-02	3.43E-02	2.52E-02	3.29E-02	3.24E-01	1.35E-03
σ^2	2.76E-04	1.39E-04	9.36E-05	1.67E-04	3.13E-03	1.45E-03	1.97E-03	1.89E-01	3.46E-06
σ	1.66E-02	1.18E-02	9.68E-03	1.29E-02	5.59E-02	3.80E-02	4.44E-02	4.34E-01	1.86E-03
Max	0.590	-37.481	37.503	0.603	-0.660	-0.293	1.296	41.222	7.80E-03
Min	0.524	-37.523	37.466	0.561	-0.913	-0.471	1.112	39.445	4.10E-04
pro	0.492	-37.545	37.528	0.567	-0.782	-0.429	1.211	43.587	2.52E-03
arg(+)	0.564	-37.513	37.479	0.567	-0.674	-0.436	1.110	39.671	3.10E-04
lys(+)	0.563	-37.493	37.475	0.572	-0.752	-0.377	1.129	39.765	7.60E-04
his(+)	0.580	-37.479	37.461	0.564	-0.765	-0.260	1.026	40.048	1.24E-03
Ave	0.569	-37.495	37.472	0.568	-0.731	-0.358	1.088	39.828	7.70E-04
asp(−)	0.543	-37.496	37.482	0.594	-0.827	-0.278	1.105	41.456	6.00E-04
glu(−)	0.508	-37.514	37.497	0.628	-0.980	-0.265	1.244	42.311	3.91E-03
tyr(−)	0.507	-37.518	37.499	0.616	-0.923	-0.330	1.246	41.540	2.63E-03
Ave	0.519	-37.509	37.493	0.613	-0.910	-0.291	1.199	41.769	2.38E-03
lys(+)*Cl*(−)	0.548	-37.494	37.482	0.588	-0.794	-0.374	1.168	40.261	3.63E-03
asp(−)*Na*(+)	0.553	-37.494	37.479	0.570	-0.720	-0.420	1.140	40.451	2.33E-03

Table 4.36: Common Back-Bone Atomic Integration: Atom H_β

Aa	q	E	K	$\|\mu\|$	Q_1	Q_2	Q_3	$Vol.$	$\|L\|$
ala	0.042	-0.623	0.623	0.125	-0.182	-0.158	0.340	43.102	1.00E-04
arg	0.042	-0.622	0.622	0.125	-0.190	-0.148	0.338	43.137	1.10E-04
asn	0.035	-0.626	0.625	0.129	-0.179	-0.156	0.335	43.836	1.50E-04
asp	0.065	-0.614	0.613	0.121	-0.182	-0.143	0.329	41.385	1.50E-04
cys	0.057	-0.616	0.616	0.126	-0.172	-0.154	0.326	42.574	1.30E-04
glu	0.043	-0.623	0.622	0.125	-0.177	-0.156	0.333	42.895	1.90E-04
gln	0.035	-0.626	0.626	0.126	-0.178	-0.156	0.338	43.198	1.50E-04
gly	0.052	-0.612	0.612	0.127	-0.218	-0.148	0.366	43.640	1.10E-04
his	0.041	-0.623	0.623	0.126	-0.182	-0.152	0.333	43.214	1.50E-04
ile	0.039	-0.628	0.628	0.124	-0.186	-0.131	0.316	41.603	1.70E-04
leu	0.041	-0.626	0.626	0.118	-0.207	-0.139	0.346	41.319	6.00E-05
lys	0.039	-0.624	0.624	0.125	-0.177	-0.157	0.338	43.233	1.60E-04
met	0.040	-0.624	0.624	0.125	-0.177	-0.156	0.333	43.111	1.50E-04
phe	0.043	-0.623	0.623	0.124	-0.180	-0.157	0.337	43.148	1.70E-04
ser	0.055	-0.616	0.616	0.125	-0.175	-0.154	0.329	42.896	1.30E-04
thr	0.047	-0.621	0.621	0.124	-0.191	-0.137	0.328	42.181	1.40E-04
trp	0.040	-0.624	0.624	0.125	-0.180	-0.157	0.337	43.299	1.60E-04
tyr	0.040	-0.624	0.624	0.125	-0.180	-0.158	0.337	43.278	1.10E-04
val	0.040	-0.627	0.627	0.124	-0.181	-0.138	0.319	41.943	1.20E-04
Ave	0.044	-0.622	0.622	0.125	-0.184	-0.150	0.335	42.789	1.37E-04
AveDev	5.92E-03	3.43E-03	3.43E-03	1.40E-03	7.70E-03	7.14E-03	6.91E-03	6.03E-01	2.41E-05
σ^2	6.27E-05	2.11E-05	2.10E-05	5.07E-06	1.29E-04	7.29E-05	1.09E-04	5.56E-01	9.32E-10
σ	7.92E-03	4.59E-03	4.58E-03	2.25E-03	1.13E-02	8.54E-03	1.05E-02	7.46E-01	3.05E-05
Max	0.065	-0.612	0.628	0.129	-0.172	-0.131	0.366	43.836	1.90E-04
Min	0.035	-0.628	0.612	0.118	-0.218	-0.158	0.316	41.319	6.00E-05
pro	0.061	-0.614	0.614	0.121	-0.360	-0.204	-0.156	40.954	1.20E-04
arg(+)	0.044	-0.623	0.622	0.125	-0.194	-0.139	0.332	42.344	1.00E-04
lys(+)	0.043	-0.624	0.623	0.125	-0.182	-0.147	0.328	42.250	7.60E-04
his(+)	0.053	-0.620	0.619	0.125	-0.186	-0.143	0.329	41.564	1.80E-04
Ave	0.047	-0.622	0.622	0.125	-0.187	-0.143	0.330	42.053	3.47E-04
asp(−)	0.017	-0.632	0.631	0.129	-0.177	-0.170	0.347	45.494	1.10E-04
glu(−)	0.026	-0.629	0.628	0.127	-0.174	-0.167	0.341	44.800	1.50E-04
tyr(−)	0.037	-0.624	0.624	0.117	-0.174	-0.168	0.341	44.519	1.50E-04
Ave	0.026	-0.628	0.628	0.124	-0.175	-0.168	0.343	44.938	1.37E-04
lys(+)*Cl*(−)	0.041	-0.623	0.623	0.125	-0.179	-0.154	0.333	43.063	3.63E-03
asp(−)*Na*(+)	0.039	-0.623	0.623	0.127	-0.175	-0.164	0.339	43.755	1.10E-04

Table 4.37: Common Back-Bone Atomic Integration: Atom C

Aa	q	E	K	$\|\mu\|$	Q_1	Q_2	Q_3	$Vol.$	$\|L\|$
ala	1.734	-36.761	36.744	0.862	-0.945	-0.455	1.400	30.392	5.10E-04
arg	1.710	-36.793	36.760	0.854	-0.930	-0.462	1.391	29.055	1.53E-03
asn	1.751	-36.752	36.737	0.835	-0.968	-0.409	1.377	29.770	1.04E-03
asp	1.744	-36.755	36.741	0.859	-0.917	-0.500	1.417	29.417	1.24E-03
cys	1.744	-36.747	36.736	0.844	-0.984	-0.407	1.391	30.406	1.97E-03
glu	1.737	-36.761	36.745	0.857	-0.956	-0.452	1.408	29.578	2.13E-03
gln	1.730	-36.765	36.747	0.866	-0.944	-0.475	1.419	29.636	1.17E-03
gly	1.747	-36.744	36.727	0.843	-1.029	-0.374	1.403	31.943	3.30E-03
his	1.736	-36.756	36.738	0.867	-0.957	-0.450	1.411	30.405	5.40E-04
ile	1.724	-36.759	36.742	0.866	-0.989	-0.382	1.371	30.339	5.75E-03
leu	1.733	-36.763	36.746	0.865	-0.929	-0.458	1.387	29.864	6.00E-05
lys	1.727	-36.762	36.745	0.861	-0.993	-0.423	1.415	30.638	4.14E-03
met	1.734	-36.756	36.745	0.860	-0.943	-0.463	1.405	29.403	7.90E-04
phe	1.736	-36.757	36.738	0.849	-0.998	-0.409	1.407	30.582	3.59E-03
ser	1.739	-36.754	36.738	0.850	-0.978	-0.424	1.401	30.378	2.02E-03
thr	1.736	-36.744	36.727	0.850	-0.959	-0.436	1.395	29.542	9.60E-04
trp	1.737	-36.756	36.738	0.853	-1.007	-0.417	1.424	29.001	1.41E-03
tyr	1.737	-36.757	36.738	0.852	-1.002	-0.417	1.419	30.428	1.82E-03
val	1.732	-36.760	36.743	0.868	-0.952	-0.468	1.420	29.654	3.22E-03
Ave	1.736	-36.758	36.741	0.856	-0.967	-0.436	1.403	30.023	1.96E-03
AveDev	6.94E-03	6.37E-03	5.12E-03	7.78E-03	2.54E-02	2.75E-02	1.18E-02	5.59E-01	1.10E-03
σ^2	1.04E-04	1.08E-04	5.19E-05	8.84E-05	9.22E-04	1.08E-03	2.19E-04	4.85E-01	2.10E-06
σ	1.02E-02	1.04E-02	7.20E-03	9.40E-03	3.04E-02	3.28E-02	1.48E-02	6.97E-01	1.45E-03
Max	1.751	-36.744	36.760	0.868	-0.917	-0.374	1.424	31.943	5.75E-03
Min	1.710	-36.793	36.727	0.835	-1.029	-0.500	1.371	29.001	6.00E-05
pro	1.737	-36.757	36.741	0.869	-0.974	-0.452	1.426	31.382	2.90E-04
arg(+)	1.711	-36.794	36.761	0.841	-0.936	-0.454	1.390	29.216	5.30E-04
lys(+)	1.732	-36.764	36.747	0.851	-0.958	-0.439	1.397	29.949	2.72E-03
his(+)	1.750	-36.753	36.735	0.827	-0.982	-0.419	1.401	30.152	1.03E-03
Ave	1.731	-36.770	36.747	0.840	-0.958	-0.438	1.396	29.772	1.43E-03
asp(−)	1.702	-36.786	36.773	0.907	-0.924	-0.519	1.443	29.247	1.82E-03
glu(−)	1.725	-36.759	36.743	0.898	-0.904	-0.497	1.401	30.961	7.50E-04
tyr(−)	1.735	-36.753	36.735	0.872	-0.969	-0.437	1.406	30.280	6.60E-04
Ave	1.721	-36.766	36.750	0.892	-0.932	-0.484	1.417	30.163	1.08E-03
lys(+)Cl(−)	1.729	-36.760	36.748	0.861	-0.939	-0.482	1.420	29.733	7.00E-05
asp(−)Na(+)	1.736	-36.768	36.753	0.857	-0.965	-0.449	1.409	28.421	1.54E-03

Table 4.38: Common Back-Bone Atomic Integration: Atom O

Aa	q	E	K	$\|\mu\|$	Q_1	Q_2	Q_3	$Vol.$	$\|L\|$
ala	-1.337	-75.653	75.618	0.537	-0.532	0.069	0.463	133.790	1.50E-04
arg	-1.325	-75.620	75.551	0.528	-0.462	0.045	0.417	130.675	1.80E-04
asn	-1.338	-75.655	75.624	0.538	-0.526	0.061	0.465	133.483	1.60E-04
asp	-1.353	-75.661	75.632	0.517	-0.563	0.140	0.423	126.322	4.00E-05
cys	-1.333	-75.642	75.620	0.543	-0.521	0.064	0.457	133.331	1.40E-04
glu	-1.339	-75.653	75.619	0.534	-0.523	0.068	0.455	131.786	2.76E-03
gln	-1.342	-75.656	75.620	0.532	-0.460	0.042	0.418	131.867	2.00E-05
gly	-1.334	-75.652	75.618	0.540	-0.548	0.072	0.476	133.938	1.60E-04
his	-1.345	-75.660	75.625	0.531	-0.494	0.092	0.402	129.530	1.70E-04
ile	-1.342	-75.654	75.618	0.527	-0.631	0.141	0.490	131.580	4.85E-03
leu	-1.340	-75.654	75.618	0.535	-0.527	0.064	0.463	133.245	1.30E-04
lys	-1.338	-75.655	75.619	0.535	-0.484	0.065	0.419	131.868	7.10E-04
met	-1.339	-75.645	75.621	0.535	-0.463	0.051	0.413	131.876	1.20E-04
phe	-1.332	-75.662	75.623	0.544	-0.494	0.098	0.396	132.668	8.70E-04
ser	-1.334	-75.651	75.618	0.540	-0.530	0.065	0.465	133.614	1.40E-04
thr	-1.343	-75.642	75.608	0.523	-0.550	0.050	0.500	134.180	1.40E-04
trp	-1.333	-75.670	75.625	0.550	-0.477	0.039	0.438	133.041	6.30E-04
tyr	-1.332	-75.661	75.623	0.545	-0.476	0.040	0.786	132.682	5.30E-04
val	-1.342	-75.655	75.620	0.526	-0.584	0.098	0.485	131.701	4.51E-03
Ave	-1.338	-75.653	75.617	0.535	-0.518	0.072	0.465	132.167	8.64E-04
AveDev	4.59E-03	6.72E-03	7.76E-03	6.26E-03	3.52E-02	2.21E-02	4.35E-02	1.29E+00	1.00E-03
σ^2	3.82E-05	1.11E-04	2.72E-04	6.74E-05	2.04E-03	8.96E-04	7.01E-03	3.43E+00	2.19E-06
σ	6.18E-03	1.05E-02	1.65E-02	8.21E-03	4.52E-02	2.99E-02	8.37E-02	1.85E+00	1.48E-03
Max	-1.325	-75.620	75.632	0.550	-0.460	0.141	0.786	134.180	4.85E-03
Min	-1.353	-75.670	75.551	0.517	-0.631	0.039	0.396	126.322	2.00E-05
pro	-1.351	-75.644	75.610	0.516	-0.565	0.070	0.495	133.178	1.90E-04
arg(+)	-1.329	-75.619	75.551	0.518	-0.471	0.074	0.397	129.464	1.70E-04
lys(+)	-1.338	-75.657	75.621	0.537	-0.441	0.044	0.397	130.528	1.20E-04
his(+)	-1.346	-75.667	75.631	0.532	-0.517	0.041	0.114	127.165	1.40E-04
Ave	-1.337	-75.648	75.601	0.529	-0.477	0.053	0.303	129.052	1.43E-04
asp(−)	-1.353	-75.646	75.618	0.518	-0.522	0.469	0.532	133.298	1.40E-04
glu(−)	-1.346	-75.648	75.614	0.522	-0.478	0.041	0.437	133.751	1.20E-04
tyr(−)	-1.333	-75.657	75.620	0.541	-0.478	0.073	0.404	133.518	1.60E-04
Ave	-1.344	-75.650	75.617	0.527	-0.493	0.194	0.458	133.522	1.40E-04
lys(+)*Cl*(−)	-1.339	-75.643	75.619	0.536	-0.457	0.147	0.410	131.603	1.70E-04
asp(−)*Na*(+)	-1.339	-75.657	75.626	0.539	-0.491	0.036	0.455	131.241	1.70E-04

Table 4.39: Side Chain Integrated Properties: *Glycine*

Atom	N	q	K	L	E
$C^{\alpha} - H$	0.99222	0.00778	0.62660	0.00008	−0.62688

Table 4.40: Side Chain Integrated Properties: *Glycine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
$C^{\alpha} - H$	0.01321	−0.07826	−0.09421	2.48871	47.22827
Shifted to Origin	0.00012	−0.03904	−0.08071	0.08966	

Table 4.41: Side Chain Integrated Properties: *Alanine*

Atom	N	q	K	L	E
C^{β}	5.84667	0.15333	37.70214	0.00289	−37.71962
$C^{\beta} - H_1$	1.00290	−0.00290	0.62343	0.00008	−0.62372
$C^{\beta} - H_2$	1.02463	−0.02463	0.63223	0.00008	−0.63252
$C^{\beta} - H_3$	1.04059	−0.04059	0.63910	0.00009	−0.63940
Sum	8.91479	0.08521	39.59691	0.00314	−39.61526

Table 4.42: Side Chain Integrated Properties: *Alanine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.00682	−0.05520	−0.02338	2.51990	62.80101
$C^{\beta} - H_1$	−0.07562	−0.09498	0.02732	0.06033	48.54262
$C^{\beta} - H_2$	0.11971	−0.04552	0.01152	0.12444	49.95125
$C^{\beta} - H_3$	−0.00553	0.02258	−0.12653	0.12859	50.62395
Sum	0.04538	−0.17312	−0.11108	2.83326	211.91882
Shifted to Origin	−0.10776	0.28529	−0.10141	0.32138	

Table 4.43: Side Chain Integrated Properties: *Serine*

Atom	N	q	K	L	E
C^{β}	5.28584	0.71416	37.35527	0.00040	−37.37180
$C^{\beta} - H_1$	1.04104	−0.04104	0.65476	0.00011	−0.65505
$C^{\beta} - H_2$	1.01006	−0.01006	0.64159	0.00010	−0.64187
O^{γ}	9.24602	−1.24602	75.44286	0.00005	−75.47625
$O^{\gamma} - H$	0.38654	0.61346	0.35273	0.00003	−0.35289
Sum	16.96951	0.03049	114.44721	0.00069	−114.49786

Table 4.44: Side Chain Integrated Properties: *Serine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	−0.61574	0.00120	0.10950	3.09529	45.72282
$C^{\beta} - H_1$	−0.02726	−0.00989	0.12285	0.62541	49.01147
$C^{\beta} - H_2$	−0.05512	0.10586	−0.02742	0.12623	47.70170
O^{γ}	−0.05378	0.17393	0.11721	0.12246	119.46833
$O^{\gamma} - H$	0.05503	0.13018	0.06564	0.21652	20.11053
Sum	−0.69688	0.40128	0.38779	4.18592	282.01484
Shifted to Origin	0.42297	−0.57872	−0.24426	0.75729	

Table 4.45: Side Chain Integrated Properties: *Cysteine*

Atom	N	q	K	L	E
C^{β}	5.98581	0.01419	37.72644	0.00312	−37.73763
$C^{\beta} - H_1$	1.00804	−0.00804	0.63582	0.00010	−0.63601
$C^{\beta} - H_2$	0.97305	0.02695	0.62066	0.00009	−0.62085
S^{γ}	15.65601	0.34399	397.38309	0.00024	−397.50101
$S^{\gamma} - H$	1.36115	−0.36115	0.77863	0.00009	−0.77886
Sum	24.98405	0.01595	437.14464	0.00364	−437.27436

Table 4.46: Side Chain Integrated Properties: *Cysteine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.12775	0.01609	0.01031	3.09174	58.73351
$C^{\beta} - H_1$	−0.03469	−0.00536	0.12428	0.12917	47.54549
$C^{\beta} - H_2$	−0.04177	0.11487	−0.02639	0.12914	45.53701
S^{γ}	0.29524	−0.98546	−0.58437	0.12504	198.24582
$S^{\gamma} - H$	−0.06670	−0.16306	−0.08917	1.18313	59.76872
Sum	0.27984	−1.02292	−0.56535	4.65822	409.83055
Shifted to Origin	0.56476	−0.38470	−0.12421	0.69453	

Table 4.47: Side Chain Integrated Properties: *Methionine*

Atom	N	q	K	L	E
C^{β}	5.84912	0.15088	37.71239	0.00050	−37.72427
$C^{\beta} - H_1$	1.02310	−0.02310	0.63957	0.00012	−0.63977
$C^{\beta} - H_2$	1.03933	−0.03933	0.64523	0.00010	−0.64543
C^{γ}	5.98252	0.01748	37.73040	0.00466	−37.74228
$C^{\gamma} - H_1$	1.00903	−0.00903	0.63460	0.00011	−0.63480
$C^{\gamma} - H_2$	0.99892	0.00108	0.63326	0.00017	−0.63346
S^{δ}	16.01941	−0.01941	397.58646	0.00161	−397.71164
C^{ϵ}	5.98208	0.01792	37.71032	0.00183	−37.72219
$C^{\epsilon} - H_1$	1.00597	−0.00597	0.62640	0.00008	−0.62659
$C^{\epsilon} - H_2$	1.00579	−0.00579	0.62630	0.00007	−0.62650
$C^{\epsilon} - H_3$	0.99462	0.00538	0.61975	0.00006	−0.61994
Sum	40.90989	0.09011	515.16468	0.00931	−515.32689

Table 4.48: Side Chain Integrated Properties: *Methionine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	−0.03262	0.00215	0.02461	2.59333	52.63179
$C^{\beta} - H_1$	−0.03643	0.12405	−0.00956	0.04092	48.55746
$C^{\beta} - H_2$	0.02097	−0.00160	0.12859	0.12964	49.74839
C^{γ}	−0.10794	0.01353	0.05283	0.13029	56.23112
$C^{\gamma} - H_1$	−0.00383	0.00402	−0.13167	0.12093	48.52923
$C^{\gamma} - H_2$	0.03065	−0.11967	0.02033	0.13179	45.74576
S^{δ}	−0.18111	0.64156	0.59725	0.12520	199.22533
C^{ϵ}	0.06912	0.07540	0.04154	0.89505	69.15650
$C^{\epsilon} - H_1$	−0.01627	−0.00472	−0.13289	0.11041	49.38813
$C^{\epsilon} - H_2$	0.03048	−0.12952	0.01629	0.13397	49.22324
$C^{\epsilon} - H_3$	−0.12120	0.00623	0.04334	0.13405	48.12024
Sum	−0.34818	0.61144	0.65065	4.54558	716.55719
Shifted to Origin	−0.04363	0.51534	0.68949	0.86191	

Table 4.49: Side Chain Integrated Properties: *Lysine*

Atom	N	q	K	L	E
C^{β}	5.85945	0.14055	37.71595	0.00084	−37.73373
$C^{\beta} - H_1$	1.04475	−0.04475	0.64853	0.00012	−0.64883
$C^{\beta} - H_2$	1.06052	−0.06052	0.65375	0.00013	−0.65406
C^{γ}	5.85244	0.14756	37.70691	0.00170	−37.72468
$C^{\gamma} - H_1$	1.04826	−0.04826	0.64859	0.00011	−0.64890
$C^{\gamma} - H_2$	1.03833	−0.03833	0.64731	0.00030	−0.64762
C^{δ}	5.86754	0.13246	37.72711	0.00060	−37.74490
$C^{\delta} - H_1$	1.07692	−0.07692	0.65874	0.00012	−0.65905
$C^{\delta} - H_2$	1.07731	−0.07731	0.65869	0.00012	−0.65900
C^{ϵ}	5.39936	0.60064	37.43794	0.00191	−37.45559
$C^{\epsilon} - H_1$	1.04598	−0.04598	0.65318	0.00011	−0.65349
$C^{\epsilon} - H_2$	1.04548	−0.04548	0.65300	0.00009	−0.65331
N^{ζ}	8.16409	−1.16409	54.87001	0.00006	−54.89588
$N^{\zeta} - H_1$	0.65607	0.34393	0.49169	0.00005	−0.49192
$N^{\zeta} - H_2$	0.65582	0.34418	0.49158	0.00006	−0.49181
Sum	40.89230	0.10770	211.66297	0.00421	−211.76275

Table 4.50: Side Chain Integrated Properties: *Lysine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.08803	−0.04586	0.00819	2.35032	52.54568
$C^{\beta} - H_1$	0.03261	0.12633	0.00660	0.09959	49.47449
$C^{\beta} - H_2$	−0.01902	−0.00093	−0.12963	0.13064	50.57125
C^{γ}	0.03952	−0.01132	−0.01361	0.13103	52.73505
$C^{\gamma} - H_1$	0.01090	0.01392	0.12800	0.04330	50.54995
$C^{\gamma} - H_2$	−0.02164	−0.11940	−0.02843	0.12922	47.94098
C^{δ}	0.02253	−0.06045	0.04728	0.12464	53.46873
$C^{\delta} - H_1$	0.03280	0.13019	0.01281	0.07998	51.86732
$C^{\delta} - H_2$	−0.00864	−0.00223	−0.13470	0.13487	52.06196
C^{ϵ}	−0.46362	−0.07037	0.18851	0.13500	46.91669
$C^{\epsilon} - H_1$	0.00528	0.01116	0.12764	0.50540	49.93475
$C^{\epsilon} - H_2$	−0.03506	−0.12153	−0.02053	0.12824	49.90626
N^{ζ}	0.08527	−0.14552	0.10659	0.12814	117.13279
$N^{\zeta} - H_1$	−0.00933	−0.00797	−0.19288	0.19952	33.07322
$N^{\zeta} - H_2$	0.04973	0.18531	0.02252	0.19327	32.77018
Sum	−0.19064	−0.11868	0.12836	4.51317	790.94930
Shifted to Origin	0.22460	−0.38692	−0.00379	0.44740	

Table 4.51: Side Chain Integrated Properties: *Lysine(+)*

Atom	N	q	K	L	E
C^{β}	5.85647	0.14353	37.71240	0.00026	−37.73048
$C^{\beta} - H_1$	1.03191	−0.03191	0.64269	0.00012	−0.64300
$C^{\beta} - H_2$	1.04473	−0.04473	0.64681	0.00013	−0.64712
C^{γ}	5.85099	0.14901	37.70612	0.00064	−37.72420
$C^{\gamma} - H_1$	1.02505	−0.02505	0.64017	0.00013	−0.64048
$C^{\gamma} - H_2$	1.00634	−0.00634	0.63598	0.00006	−0.63628
C^{δ}	5.86066	0.13934	37.71550	0.00284	−37.73357
$C^{\delta} - H_1$	1.03582	−0.03582	0.64239	0.00011	−0.64270
$C^{\delta} - H_2$	1.03571	−0.03571	0.64227	0.00012	−0.64258
C^{ϵ}	5.63650	0.36350	37.57594	0.00100	−37.59394
$C^{\epsilon} - H_1$	0.95082	0.04918	0.61245	0.00009	−0.61274
$C^{\epsilon} - H_2$	0.94957	0.05043	0.61190	0.00008	−0.61219
N^{ζ}	8.14942	−1.14942	55.01640	0.00037	−55.04277
$N^{\zeta} - H_1$	0.51441	0.48559	0.41606	0.00007	−0.41626
$N^{\zeta} - H_2$	0.51428	0.48572	0.41599	0.00005	−0.41618
$N^{\zeta} - H_3$	0.51066	0.48934	0.41441	0.00004	−0.41461
Sum	40.97337	1.02663	212.04746	0.00611	−212.14911

Table 4.52: Side Chain Integrated Properties: *Lysine(+)* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.00524	0.00098	−0.01093	11.88038	52.74459
$C^{\beta} - H_1$	0.04299	0.12111	0.00388	0.01216	49.06171
$C^{\beta} - H_2$	−0.00451	−0.01069	−0.12727	0.12857	50.23537
C^{γ}	−0.10299	−0.01701	0.02400	0.12780	53.47160
$C^{\gamma} - H_1$	0.02570	0.01778	0.12258	0.10711	48.53879
$C^{\gamma} - H_2$	0.00131	−0.11675	−0.02845	0.12650	45.02530
C^{δ}	−0.19374	0.06016	−0.01336	0.12018	54.66976
$C^{\delta} - H_1$	0.05680	0.11433	0.01081	0.20330	49.67367
$C^{\delta} - H_2$	0.02160	−0.01743	−0.12507	0.12812	49.83191
C^{ϵ}	−0.54312	−0.01836	0.15806	0.12811	53.65308
$C^{\epsilon} - H_1$	0.02892	0.01666	0.11354	0.56595	44.80380
$C^{\epsilon} - H_2$	−0.00593	−0.11584	−0.02329	0.11834	44.53657
N^{ζ}	−0.05200	−0.01137	0.02465	0.11831	91.68984
$N^{\zeta} - H_1$	0.00187	−0.02015	−0.14426	0.05866	25.70881
$N^{\zeta} - H_2$	0.04378	0.13767	0.01847	0.14567	25.48021
$N^{\zeta} - H_3$	0.10518	−0.08197	0.05239	0.14564	25.09372
Sum	−0.56889	0.05910	0.05576	14.11479	764.21873
Shifted to Origin	−10.95978	0.29386	0.23267	10.96619	

Table 4.53: Side Chain Integrated Properties: *Valine*

Atom	N	q	K	L	E
C^{β}	5.85509	0.14491	37.69020	0.00066	−37.70808
$C^{\beta} - H$	1.05169	−0.05169	0.65902	0.00013	−0.65934
$C^{\gamma 1}$	5.86281	0.13719	37.70539	0.00236	−37.72328
$C^{\gamma 1} - H_1$	1.03168	−0.03168	0.63836	0.00001	−0.63866
$C^{\gamma 1} - H_2$	1.07344	−0.07344	0.65469	0.00002	−0.65500
$C^{\gamma 1} - H_3$	1.03912	−0.03912	0.63874	0.00008	−0.63905
$C^{\gamma 2}$	5.84881	0.15119	37.70205	0.00391	−37.71994
$C^{\gamma 2} - H_1$	1.03295	−0.03295	0.63660	0.00008	−0.63691
$C^{\gamma 2} - H_2$	1.02640	−0.02640	0.63392	0.00178	−0.63422
$C^{\gamma 2} - H_3$	1.05045	−0.05045	0.64332	0.00008	−0.64362
Sum	24.87242	0.12758	117.60231	0.00911	−117.65809

Table 4.54: Side Chain Integrated Properties: *Valine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.01135	0.07664	−0.03405	2.50743	46.06129
$C^{\beta} - H$	−0.12397	−0.02763	0.02557	0.08463	48.73511
$C^{\gamma 1}$	0.01861	0.00286	0.03082	0.12956	61.51612
$C^{\gamma 1} - H_1$	0.10603	0.04343	−0.05041	0.03612	48.34639
$C^{\gamma 1} - H_2$	−0.02796	−0.10374	−0.06043	0.12518	50.49705
$C^{\gamma 1} - H_3$	−0.08345	0.08696	−0.04876	0.12327	50.58952
$C^{\gamma 2}$	0.02502	0.00106	−0.00047	0.13001	60.84130
$C^{\gamma 2} - H_1$	−0.01681	−0.00114	0.12709	0.02505	50.16563
$C^{\gamma 2} - H_2$	0.11800	0.05069	−0.03847	0.12820	48.41219
$C^{\gamma 2} - H_3$	−0.07884	0.09650	−0.03661	0.13406	51.09420
Sum	−0.05200	0.22562	−0.08571	3.42351	516.25881
Shifted to Origin	0.24581	−0.33522	−0.22221	0.47135	

Table 4.55: Side Chain Integrated Properties: *Isoleucine*

Atom	N	q	K	L	E
C^{β}	5.87604	0.12396	37.69751	0.00821	−37.71542
$C^{\beta} - H$	1.05431	−0.05431	0.65940	0.00013	−0.65971
$C^{\gamma 2}$	5.86438	0.13562	37.70668	0.00224	−37.72459
$C^{\gamma 2} - H_1$	1.04203	−0.04203	0.63986	0.00436	−0.64016
$C^{\gamma 2} - H_2$	1.07396	−0.07396	0.65595	0.00004	−0.65626
$C^{\gamma 2} - H_3$	1.03311	−0.03311	0.63863	0.00006	−0.63893
$C^{\gamma 1}$	5.84700	0.15300	37.70050	0.00008	−37.71841
$C^{\gamma 1} - H_1$	1.04098	−0.04098	0.64744	0.00091	−0.64775
$C^{\gamma 1} - H_2$	1.04547	−0.04547	0.65093	0.00012	−0.65124
$C^{\delta 1}$	5.84925	0.15075	37.68819	0.00325	−37.70609
$C^{\delta 1} - H_1$	1.04163	−0.04163	0.63926	0.00007	−0.63956
$C^{\delta 1} - H_2$	1.05808	−0.05808	0.64519	0.00009	−0.64549
$C^{\delta 1} - H_3$	1.06053	−0.06053	0.64721	0.00378	−0.64752
Sum	32.88676	0.11324	156.61676	0.02334	−156.69115

Table 4.56: Side Chain Integrated Properties: *Isoleucine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	−0.06038	0.06870	−0.04549	2.47337	47.44941
$C^{\beta} - H$	−0.07189	−0.10691	0.02283	0.10215	48.65470
$C^{\gamma 2}$	0.00476	0.01845	0.03641	0.13084	61.07893
$C^{\gamma 2} - H_1$	−0.12635	0.00222	−0.05967	0.04109	48.99931
$C^{\gamma 2} - H_2$	0.05659	−0.09211	−0.05717	0.13975	50.07943
$C^{\gamma 2} - H_3$	0.04450	0.10714	−0.04891	0.12229	48.41045
$C^{\gamma 1}$	−0.01632	0.01241	−0.01015	0.12590	52.57513
$C^{\gamma 1} - H_1$	0.05532	0.11003	−0.03941	0.02287	48.29732
$C^{\gamma 1} - H_2$	0.00427	−0.01241	0.12596	0.12930	48.91998
$C^{\delta 1}$	0.02815	0.00042	0.00637	0.12664	63.20571
$C^{\delta 1} - H_1$	−0.05729	0.09617	0.06581	0.02886	51.00016
$C^{\delta 1} - H_2$	−0.07372	−0.10758	0.02300	0.12986	51.89114
$C^{\delta 1} - H_3$	−0.02860	0.02328	−0.13314	0.13243	50.31489
Sum	−0.24095	0.11980	−0.11355	3.70536	670.87656
Shifted to Origin	0.22511	−0.10199	−0.27830	0.37219	

Table 4.57: Side Chain Integrated Properties: *Threonine*

Atom	N	q	K	L	E
C^{β}	5.34069	0.65931	37.37550	0.00152	−37.39231
$C^{\beta} - H$	1.02207	−0.02207	0.65235	0.00011	−0.65264
$C^{\gamma 1}$	9.26238	−1.26238	75.45863	0.00053	−75.49255
$C^{\gamma 1} - H$	0.38570	0.61430	0.35135	0.00003	−0.35151
$C^{\gamma 2}$	5.85003	0.14997	37.72348	0.00780	−37.74044
$C^{\gamma 2} - H_1$	1.02582	−0.02582	0.63171	0.00008	−0.63200
$C^{\gamma 2} - H_2$	1.04974	−0.04974	0.64129	0.00008	−0.64158
$C^{\gamma 2} - H_3$	1.02192	−0.02192	0.63180	0.00006	−0.63208
Sum	24.95832	0.04168	153.46610	0.01021	−153.53511

Table 4.58: Side Chain Integrated Properties: *Threonine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	−0.05516	0.11025	0.61610	2.60187	39.87987
$C^{\beta} - H$	0.05330	0.10756	0.02812	0.62831	47.60450
$O^{\gamma 1}$	−0.09775	0.18046	0.01049	0.12329	114.95290
$O^{\gamma 1} - H$	−0.06758	0.12105	−0.07186	0.20550	20.28605
$C^{\gamma 2}$	−0.00843	−0.02051	0.00950	0.15615	61.28170
$C^{\gamma 2} - H_1$	−0.00248	0.02406	0.12378	0.02412	49.79371
$C^{\gamma 2} - H_2$	−0.06263	0.09585	−0.05925	0.12612	51.39198
$C^{\gamma 2} - H_3$	−0.07792	−0.09955	−0.01961	0.12892	48.86692
Sum	−0.31865	0.51917	0.63728	3.99429	434.05764
Shifted to Origin	0.39903	−0.67060	−0.58271	0.97390	

Table 4.59: Side Chain Integrated Properties: *Leucine*

Atom	N	q	K	L	E
C^{β}	5.87061	0.12939	37.71733	0.00450	−37.73529
$C^{\beta} - H_1$	1.06250	−0.06250	0.65459	0.00012	−0.65490
$C^{\beta} - H_2$	1.02542	−0.02542	0.64291	0.00012	−0.64321
C^{γ}	5.85006	0.14994	37.67956	0.00468	−37.69751
$C^{\gamma} - H$	1.06770	−0.06770	0.66604	0.00010	−0.66636
$C^{\delta 1}$	5.85914	0.14086	37.70454	0.00292	−37.72250
$C^{\delta 1} - H_1$	1.04786	−0.04786	0.64195	0.00007	−0.64226
$C^{\delta 1} - H_2$	1.04620	−0.04620	0.64491	0.00002	−0.64522
$C^{\delta 1} - H_3$	1.06667	−0.06667	0.64888	0.00006	−0.64919
$C^{\delta 2}$	5.85835	0.14165	37.70351	0.00251	−37.72147
$C^{\delta 2} - H_1$	1.05184	−0.05184	0.64372	0.00212	−0.64403
$C^{\delta 2} - H_2$	1.04916	−0.04916	0.64228	0.00008	−0.64259
$C^{\delta 2} - H_3$	1.04223	−0.04223	0.63931	0.00008	−0.63961
Sum	32.89774	0.10226	156.62953	0.01738	−156.70414

Table 4.60: Side Chain Integrated Properties: *Leucine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.10844	0.02290	−0.02225	2.48195	53.01717
$C^{\beta} - H_1$	−0.02117	−0.04192	0.12150	0.11304	50.33493
$C^{\beta} - H_2$	−0.04370	0.11535	0.00432	0.13026	47.48927
C^{γ}	−0.00048	−0.01976	−0.05089	0.12342	47.60434
$C^{\gamma} - H$	0.04651	0.08233	0.08863	0.05459	49.79784
$C^{\delta 1}$	−0.02509	0.02795	−0.01447	0.12960	61.31343
$C^{\delta 1} - H_1$	0.13045	0.00832	−0.00681	0.04025	51.13300
$C^{\delta 1} - H_2$	−0.01456	−0.09225	−0.08350	0.13090	48.90515
$C^{\delta 1} - H_3$	−0.02293	−0.05599	0.11687	0.12528	52.25670
$C^{\delta 2}$	−0.01355	0.00124	0.02295	0.13161	62.20544
$C^{\delta 2} - H_1$	−0.03317	−0.08732	−0.10332	0.02668	50.17487
$C^{\delta 1} - H_2$	0.12667	0.02477	−0.01996	0.13929	51.34277
$C^{\delta 1} - H_3$	−0.05771	0.11306	−0.02370	0.13060	50.82482
Sum	0.17970	0.09867	0.02937	3.75746	676.39971
Shifted to Origin	−0.21302	−0.20678	0.09290	0.31107	

Table 4.61: Side Chain Integrated Properties: *Aspartic Acid*

Atom	N	q	K	L	E
C^{β}	5.86289	0.13711	37.75784	0.00043	−37.77266
$C^{\beta} - H_1$	0.99305	0.00695	0.62686	0.00010	−0.62711
$C^{\beta} - H_2$	0.96788	0.03212	0.61336	0.00010	−0.61360
C^{γ}	4.17454	1.82546	36.64403	0.00073	−36.65841
$O^{\delta 1}$	9.29769	−1.29769	75.60368	0.00018	−75.63335
$O^{\delta 2}$	9.34379	−1.34379	75.65695	0.00011	−75.68665
$O^{\delta 2} - H$	0.34593	0.65407	0.32364	0.00004	−0.32377
Sum	30.98577	0.01423	227.22636	0.00169	−227.31555

Table 4.62: Side Chain Integrated Properties: *Aspartic Acid* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	−0.11294	0.05631	0.03575	2.45192	54.07806
$C^{\beta} - H_1$	−0.03521	−0.01185	0.11626	0.13117	46.46141
$C^{\beta} - H_2$	−0.01904	0.11654	−0.01996	0.12205	45.56810
C^{γ}	−0.88380	0.25171	0.05953	0.11976	30.35146
$O^{\delta 1}$	0.10252	−0.18540	0.25080	0.92087	119.76528
$O^{\delta 2}$	−0.30025	0.32823	−0.38186	0.32831	135.21165
$O^{\delta 2} - H$	0.13303	−0.04063	−0.00860	0.58626	18.09357
Sum	−1.11568	0.51492	0.05193	4.66035	449.52954
Shifted to Origin	0.38033	−0.43464	0.65441	0.87282	

Table 4.63: Side Chain Integrated Properties: *Aspartate*

Atom	N	q	K	L	E
C^{β}	5.91359	0.08641	37.76162	0.00204	−37.77554
$C^{\beta} - H_1$	1.05322	−0.05322	0.64991	0.00012	−0.65015
$C^{\beta} - H_2$	1.01898	−0.01898	0.63667	0.00011	−0.63691
C^{γ}	3.96966	2.03034	36.48660	0.00127	−36.50005
$O^{\delta 1}$	9.41366	−1.41366	75.58356	0.00012	−75.61142
$O^{\delta 2}$	9.43615	−1.43615	75.55149	0.00024	−75.57935
Sum	30.80526	−0.80526	226.66985	0.00390	−226.75341

Table 4.64: Side Chain Integrated Properties: *Aspartate* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.18389	0.00869	0.01795	7.23156	55.73962
$C^{\beta} - H_1$	−0.05001	0.02303	0.11885	0.18497	49.36733
$C^{\beta} - H_2$	−0.03718	0.11826	−0.05196	0.13099	48.57210
C^{γ}	−0.64804	0.15536	−0.09164	0.13442	28.54234
$O^{\delta 1}$	−0.30570	−0.27977	0.11553	0.67268	144.38095
$O^{\delta 2}$	−0.02561	0.30927	−0.11673	0.43020	136.19917
Sum	−0.88265	0.33484	−0.00801	8.78481	462.80152
Shifted to Origin	6.84910	1.24983	0.89891	7.01999	

Table 4.65: Side Chain Integrated Properties: *Asparagine*

Atom	N	q	K	L	E
C^{β}	5.88538	0.11462	37.76846	0.00356	−37.78375
$C^{\beta} - H_1$	0.98869	0.01131	0.62469	0.00010	−0.62494
$C^{\beta} - H_2$	1.00837	−0.00837	0.63202	0.00010	−0.63228
C^{γ}	4.23405	1.76595	36.70306	0.00051	−36.71792
$O^{\delta 1}$	9.35296	−1.35296	75.61621	0.00016	−75.64684
$N^{\delta 2}$	8.39193	−1.39193	55.15593	0.00013	−55.17827
$N^{\delta 2} - H_1$	0.56291	0.43709	0.44473	0.00006	−0.44491
$N^{\delta 2} - H_2$	0.55129	0.44871	0.43647	0.00005	−0.43665
Sum	30.97558	0.02442	207.38157	0.00467	−207.46555

Table 4.66: Side Chain Integrated Properties: *Asparagine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	−0.04085	0.03581	0.02410	2.88255	53.81752
$C^{\beta} - H_1$	−0.03378	−0.00337	0.11899	0.05943	46.14419
$C^{\beta} - H_2$	−0.01271	0.11603	−0.03115	0.12374	47.29735
C^{γ}	−0.78993	0.29306	−0.01070	0.12081	31.05196
$O^{\delta 1}$	−0.21256	0.37321	−0.28994	0.84260	134.62234
$N^{\delta 2}$	−0.15011	−0.11980	0.19506	0.51820	121.23672
$N^{\delta 2} - H_1$	−0.06929	0.11045	−0.09918	0.27374	27.31728
$N^{\delta 2} - H_2$	0.15413	−0.03821	−0.03102	0.16382	26.97855
Sum	−1.15510	0.76717	−0.12383	4.98489	488.46643
Shifted to Origin	0.49977	−1.23583	1.27875	1.84723	

Table 4.67: Side Chain Integrated Properties: *Glutamic Acid*

Atom	N	q	K	L	E
C^{β}	5.85797	0.14203	37.70906	0.00363	−37.72608
$C^{\beta} - H_1$	1.02463	−0.02463	0.64170	0.00013	−0.64199
$C^{\beta} - H_2$	1.02751	−0.02751	0.64153	0.00012	−0.64182
C^{γ}	5.84769	0.15231	37.74027	0.00386	−37.75730
$C^{\gamma} - H_1$	0.98402	0.01598	0.62307	0.00011	−0.62335
$C^{\gamma} - H_2$	0.98340	0.01660	0.61949	0.00230	−0.61977
C^{δ}	4.21178	1.78822	36.66409	0.00270	−36.68063
$O^{\epsilon 1}$	9.30228	−1.30228	75.59434	0.00036	−75.62845
$O^{\epsilon 1} - H$	0.35098	0.64902	0.32682	0.00004	−0.32697
$O^{\epsilon 2}$	9.34138	−1.34138	75.64767	0.00014	−75.68181
Sum	38.93163	0.06837	266.20805	0.01339	−266.32817

Table 4.68: Side Chain Integrated Properties: *Glutamic Acid* (continued)

Atom	no	μ_x	μ_y	μ_z	$\|\mu\|$	Vol(0.001)
C^{β}		−0.03110	0.00953	−0.02861	2.83751	53.54271
$C^{\beta} - H_1$		0.02288	0.00179	−0.12161	0.04332	47.31700
$C^{\beta} - H_2$		−0.05515	−0.11443	0.00604	0.12376	48.72392
C^{γ}		0.15097	0.01879	0.06814	0.12716	54.22485
$C^{\gamma} - H_1$		−0.00292	−0.01418	0.12212	0.16670	46.46534
$C^{\gamma} - H_2$		0.03526	0.12137	−0.02536	0.12297	44.96373
C^{δ}		0.90563	0.06539	0.33078	0.12891	31.85746
$O^{\epsilon 1}$		−0.21029	−0.08105	0.20488	0.96636	120.08419
$O^{\epsilon 1} - H$		−0.13339	−0.01026	−0.04722	0.30457	18.34734
$O^{\epsilon 2}$		0.48542	0.14130	−0.30560	0.14187	137.03629
Sum		1.16732	0.13825	0.20357	4.96314	602.56281
Shifted to Origin		−0.61065	−0.14678	0.20912	0.66195	

Table 4.69: Side Chain Integrated Properties: *Glutamate*

Atom	N	q	K	L	E
C^{β}	5.85242	0.14758	37.72287	0.00349	−37.73941
$C^{\beta} - H_1$	1.02399	−0.02399	0.64146	0.00023	−0.64174
$C^{\beta} - H_2$	1.04871	−0.04871	0.65057	0.00013	−0.65085
C^{γ}	5.90281	0.09719	37.75117	0.00183	−37.76772
$C^{\gamma} - H_1$	1.05361	−0.05361	0.64793	0.00010	−0.64821
$C^{\gamma} - H_2$	1.05598	−0.05598	0.64863	0.00001	−0.64891
C^{δ}	3.96865	2.03135	36.48574	0.00180	−36.50173
$O^{\epsilon 1}$	9.44168	−1.44168	75.53796	0.00006	−75.57106
$O^{\epsilon 2}$	9.43779	−1.43778	75.54181	0.00011	−75.57492
Sum	38.78564	−0.78564	265.62814	0.00776	−265.74456

Table 4.70: Side Chain Integrated Properties: *Glutamate* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	−0.13214	0.09949	−0.00994	8.15753	50.77288
$C^{\beta} - H_1$	0.06294	−0.00873	−0.12511	0.16571	46.17417
$C^{\beta} - H_2$	−0.03186	−0.12837	0.02593	0.14032	49.59639
C^{γ}	−0.17412	0.00783	−0.04710	0.13478	56.39891
$C^{\gamma} - H_1$	0.00951	−0.01282	0.14090	0.18054	51.11124
$C^{\gamma} - H_2$	0.05627	0.11405	−0.02257	0.14180	50.20013
C^{δ}	0.57074	0.01054	0.31526	0.12917	28.69472
$O^{\epsilon 1}$	−0.01327	0.17024	0.30810	0.65210	144.08356
$O^{\epsilon 2}$	0.28690	−0.16041	−0.16756	0.35226	147.79620
Sum	0.63497	0.09182	0.41790	10.05420	624.82820
Shifted to Origin	−8.62073	0.61667	−1.32274	8.74340	

Table 4.71: Side Chain Integrated Properties: *Glutamine*

Atom	N	q	K	L	E
C^{β}	5.84360	0.15640	37.71185	0.00317	−37.72966
$C^{\beta} - H_1$	1.00681	−0.00681	0.63449	0.00011	−0.63479
$C^{\beta} - H_2$	1.02733	−0.02733	0.64173	0.00013	−0.64203
C^{γ}	5.89061	0.10939	37.75688	0.00309	−37.77471
$C^{\gamma} - H_1$	1.01360	−0.01360	0.63085	0.00010	−0.63115
$C^{\gamma} - H_2$	1.00237	−0.00237	0.62990	0.00013	−0.63019
C^{δ}	4.24786	1.75214	36.70480	0.00073	−36.72215
$O^{\epsilon 1}$	9.34981	−1.34981	75.60874	0.00013	−75.64446
$N^{\epsilon 2}$	8.39229	−1.39229	55.14730	0.00049	−55.17336
$N^{\epsilon 2} - H_1$	0.56969	0.43031	0.44759	0.00006	−0.44780
$N^{\epsilon 2} - H_2$	0.55469	0.44531	0.43776	0.00005	−0.43797
Sum	38.89866	0.10134	246.35189	0.00819	−246.46827

Table 4.72: Side Chain Integrated Properties: *Glutamine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	−0.02510	−0.00147	−0.02184	3.02661	52.36726
$C^{\beta} - H_1$	−0.03443	−0.11919	0.02363	0.03330	46.56615
$C^{\beta} - H_2$	0.02921	0.00928	−0.12398	0.12629	48.24518
C^{γ}	0.07277	0.00860	0.01915	0.12771	54.83214
$C^{\gamma} - H_1$	−0.01958	−0.00904	0.12526	0.07574	48.76725
$C^{\gamma} - H_2$	0.02289	0.11414	−0.02722	0.12710	45.73573
C^{δ}	0.73814	0.16354	0.40895	0.11955	32.23165
$O^{\epsilon 1}$	0.10544	0.37778	0.34550	0.85956	135.75742
$N^{\epsilon 2}$	0.18513	−0.17187	−0.07153	0.52269	121.43739
$N^{\epsilon 2} - H_1$	0.03430	0.11857	0.11090	0.26254	27.79670
$N^{\epsilon 2} - H_2$	−0.15262	−0.00464	−0.05751	0.16593	27.12263
Sum	0.95616	0.48569	0.73131	5.44703	640.85950
Shifted to Origin	0.02423	−1.20595	−1.26207	1.74577	

Table 4.73: Side Chain Integrated Properties: *Arginine*

Atom	N	q	K	L	E
C^{β}	5.86481	0.13519	37.71217	0.00037	−37.74619
$C^{\beta} - H_1$	1.04157	−0.04157	0.64680	0.00011	−0.64738
$C^{\beta} - H_2$	1.05037	−0.05037	0.65220	0.00014	−0.65279
C^{γ}	5.85463	0.14537	37.71002	0.00157	−37.74405
$C^{\gamma} - H_1$	1.01737	−0.01737	0.64213	0.00126	−0.64271
$C^{\gamma} - H_2$	1.03592	−0.03592	0.64504	0.00011	−0.64562
C^{δ}	5.42862	0.57138	37.43121	0.00494	−37.46498
$C^{\delta} - H_1$	1.01409	−0.01409	0.64542	0.00004	−0.64601
$C^{\delta} - H_2$	1.06168	−0.06168	0.66112	0.00010	−0.66172
N^{ϵ}	8.44627	−1.44627	55.17152	0.00035	−55.22129
$N^{\epsilon} - H$	0.58782	0.41218	0.45783	0.00010	−0.45824
C^{ζ}	3.95129	2.04871	36.56269	0.00006	−36.59568
$N^{\eta 1}$	8.41371	−1.41371	55.04329	0.00010	−55.09295
$N^{\eta 1} - H$	0.65639	0.34361	0.48455	0.00006	−0.48498
$N^{\eta 2}$	8.33095	−1.33095	55.05528	0.00026	−55.10496
$N^{\eta 2} - H_1$	0.59440	0.40560	0.45970	0.00009	−0.46011
$N^{\eta 2} - H_2$	0.56291	0.43709	0.44276	0.00006	−0.44316
Sum	54.91278	0.08722	320.42372	0.00664	−320.71282

Table 4.74: Side Chain Integrated Properties: *Arginine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.08064	−0.02566	0.00014	2.80394	52.88359
$C^{\beta} - H_1$	0.02847	0.12537	−0.01183	0.08463	49.70020
$C^{\beta} - H_2$	−0.01541	−0.01274	−0.12651	0.12910	48.96030
C^{γ}	−0.00136	−0.02320	0.01154	0.12808	53.77267
$C^{\gamma} - H_1$	−0.00966	−0.11076	0.00767	0.02595	44.86029
$C^{\gamma} - H_2$	0.02099	0.04099	0.12016	0.11145	49.57491
C^{δ}	−0.43119	0.30232	−0.18841	0.12868	47.68958
$C^{\delta} - H_1$	0.01907	0.11896	−0.02227	0.55930	45.28804
$C^{\delta} - H_2$	−0.02647	−0.03014	−0.12639	0.12252	50.69821
N^{ϵ}	0.16072	0.07487	−0.05864	0.13261	100.76532
$N^{\epsilon} - H$	−0.04277	−0.13357	0.09999	0.18675	28.80814
C^{ζ}	−0.02962	−0.02580	0.00152	0.17225	28.21632
$N^{\eta 1}$	−0.02715	0.18276	−0.26862	0.03931	148.26195
$N^{\eta 1} - H$	−0.08358	−0.10963	0.13845	0.32603	33.21681
$N^{\eta 2}$	−0.06494	−0.03897	0.15085	0.19539	116.96671
$N^{\eta 2} - H_1$	−0.07226	−0.01470	−0.15731	0.16880	29.12375
$N^{\eta 2} - H_2$	0.15885	0.04437	−0.00997	0.17373	27.56747
Sum	−0.33567	0.36446	−0.43962	5.48851	956.35426
Shifted to Origin	1.50057	−0.03877	0.37474	1.54714	

Table 4.75: Side Chain Integrated Properties: *Arginine(+)*

Atom	N	q	K	L	E
C^{β}	5.86298	0.13702	37.70570	0.00040	−37.73977
$C^{\beta} - H_1$	1.02576	−0.02576	0.63971	0.00010	−0.64029
$C^{\beta} - H_2$	1.03690	−0.03690	0.64630	0.00013	−0.64689
C^{γ}	5.86261	0.13739	37.71469	0.00198	−37.74878
$C^{\gamma} - H_1$	1.02759	−0.02759	0.64175	0.00011	−0.64233
$C^{\gamma} - H_2$	0.97159	0.02841	0.62430	0.00015	−0.62487
C^{δ}	5.53479	0.46521	37.48882	0.00151	−37.52270
$C^{\delta} - H_1$	1.00301	−0.00301	0.63694	0.00011	−0.63752
$C^{\delta} - H_2$	1.00769	−0.00769	0.63890	0.00011	−0.63948
N^{ϵ}	8.45418	−1.45418	55.24308	0.00069	−55.29301
$N^{\epsilon} - H$	0.53359	0.46641	0.42853	0.00011	−0.42891
C^{ζ}	3.82346	2.17654	36.48560	0.00071	−36.51857
$N^{\eta 1}$	8.38921	−1.38921	55.15551	0.00028	−55.20536
$N^{\eta 1} - H_1$	0.51721	0.48279	0.41796	0.00005	−0.41834
$N^{\eta 1} - H_2$	0.51178	0.48822	0.41608	0.00010	−0.41645
$N^{\eta 2}$	8.38022	−1.38022	55.14344	0.00002	−55.19328
$N^{\eta 2} - H_1$	0.51802	0.48198	0.41834	0.00005	−0.41872
$N^{\eta 1} - H_2$	0.51498	0.48502	0.41654	0.00003	−0.41692
Sum	54.97558	1.02442	320.86220	0.00972	−321.15219

Table 4.76: Side Chain Integrated Properties: *Arginine(+)* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.01045	0.01590	−0.02423	10.63615	53.55622
$C^{\beta} - H_1$	0.02376	0.12405	−0.01555	0.03081	49.06328
$C^{\beta} - H_2$	0.00474	−0.02395	−0.12326	0.12726	48.36555
C^{γ}	−0.11776	−0.03903	0.05183	0.12565	54.25783
$C^{\gamma} - H_1$	0.03203	0.05007	0.11074	0.13445	48.72585
$C^{\gamma} - H_2$	0.01561	−0.10852	0.02072	0.12568	40.79338
C^{δ}	−0.49831	0.25303	−0.21166	0.11158	49.89725
$C^{\delta} - H_1$	−0.00133	−0.04570	−0.11563	0.59761	47.16404
$C^{\delta} - H_2$	0.04433	0.10966	−0.03334	0.12434	47.11844
N^{ϵ}	0.21257	0.01460	−0.11232	0.12289	98.72123
$N^{\epsilon} - H$	−0.03472	−0.07087	0.13239	0.24086	25.79910
C^{ζ}	−0.01131	−0.00744	0.00634	0.15412	24.31148
$N^{\eta 1}$	−0.07622	−0.10057	0.21292	0.01495	118.22791
$N^{\eta 1} - H_1$	0.14127	−0.01136	−0.04208	0.24750	25.15772
$N^{\eta 1} - H_2$	−0.09153	0.07183	−0.08798	0.14784	23.84019
$N^{\eta 2}$	−0.15807	0.12123	−0.13641	0.14587	119.20230
$N^{\eta 2} - H_1$	0.14086	−0.01429	−0.04375	0.24144	25.26595
$N^{\eta 2} - H_2$	−0.04232	−0.06037	0.12735	0.14818	25.15521
Sum	−0.40596	0.27826	−0.28391	13.47719	924.62294
Shifted to Origin	−9.56206	0.05229	−0.13440	9.56315	

Table 4.77: Side Chain Integrated Properties: *Proline*

Atom	N	q	K	L	E
C^{β}	5.85798	0.14203	37.72953	0.00056	−37.74671
$C^{\beta} - H_1$	1.00004	−0.00004	0.62791	0.00006	−0.62820
$C^{\beta} - H_2$	1.04328	−0.04328	0.64727	0.00009	−0.64756
C^{γ}	5.86140	0.13860	37.73006	0.00085	−37.74723
$C^{\gamma} - H_1$	1.04450	−0.04450	0.64837	0.00006	−0.64867
$C^{\gamma} - H_2$	1.03802	−0.03802	0.64139	0.00006	−0.64169
C^{δ}	5.47924	0.52076	37.48928	0.00158	−37.50634
$C^{\delta} - H_1$	1.05426	−0.05426	0.65793	0.00012	−0.65823
$C^{\delta} - H_2$	0.98293	0.01707	0.63304	0.00003	−0.63333
Sum	23.36164	0.63836	116.80480	0.00343	−116.85796

Table 4.78: Side Chain Integrated Properties: *Proline* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.00642	−0.09540	0.02646	1.81294	55.90071
$C^{\beta} - H_1$	0.07136	−0.09838	0.00164	0.09921	46.86706
$C^{\beta} - H_2$	−0.09017	−0.03226	−0.08359	0.12155	50.17295
C^{γ}	0.00973	−0.05531	−0.00335	0.12712	55.99612
$C^{\gamma} - H_1$	0.08285	−0.00709	0.09357	0.05626	50.16348
$C^{\gamma} - H_2$	−0.08619	−0.08459	0.03422	0.12518	50.29285
C^{δ}	−0.28697	−0.30633	0.39014	0.12552	48.54397
$C^{\delta} - H_1$	−0.12017	−0.00549	−0.04254	0.57306	49.86121
$C^{\delta} - H_2$	−0.01013	0.02738	0.11721	0.12759	43.21871
Sum	−0.42328	−0.65747	0.53377	3.16843	451.01707
Shifted to Origin	2.06174	0.72853	−0.24291	2.20012	

Table 4.79: Side Chain Integrated Properties: *Histidine*

Atom	N	q	K	L	E
C^{β}	5.84089	0.15911	37.71233	0.00239	−37.72997
$C^{\beta} - H_1$	1.00815	−0.00815	0.63423	0.00011	−0.63453
$C^{\beta} - H_2$	1.02542	−0.02542	0.63919	0.00012	−0.63949
C^{γ}	5.44498	0.55502	37.50310	0.00059	−37.52065
$N^{\delta 1}$	8.48351	−1.48351	55.22158	0.00062	−55.24742
$C^{\delta 2}$	5.55798	0.44202	37.53083	0.00016	−37.54839
$C^{\delta 2} - H$	0.90405	0.09595	0.58416	0.00001	−0.58444
$C^{\epsilon 1}$	4.62086	1.37914	36.97513	0.00053	−36.99243
$C^{\epsilon 1} - H$	0.95714	0.04286	0.60784	0.00004	−0.60812
$N^{\epsilon 2}$	8.51270	−1.51270	55.34549	0.00017	−55.37139
$N^{\epsilon 2} - H$	0.54793	0.45207	0.43621	0.00004	−0.43642
Sum	42.90360	0.09640	263.19009	0.00478	−263.31325

Table 4.80: Side Chain Integrated Properties: *Histidine* (continued)

Atom	μ_x	μ_y	μ_z	$\|\mu\|$	Vol(0.001)
C^{β}	−0.00053	0.01067	0.01993	2.93364	52.80525
$C^{\beta} - H_1$	−0.01186	0.12861	−0.00207	0.02261	47.39185
$C^{\beta} - H_2$	0.04262	−0.00901	0.12264	0.12917	48.78703
C^{γ}	0.57200	−0.14852	−0.31596	0.13015	61.38902
$N^{\delta 1}$	−0.14451	−0.06332	−0.02377	0.67013	127.61726
$C^{\delta 2}$	0.79632	0.16177	−0.11472	0.15955	77.97672
$C^{\delta 2} - H$	0.03790	−0.05347	−0.09033	0.82065	40.94832
$C^{\epsilon 1}$	−0.76441	0.00542	0.24002	0.11160	55.87919
$C^{\epsilon 1} - H$	−0.11517	−0.00929	0.02448	0.80123	46.46434
$N^{\epsilon 2}$	0.05715	0.14113	0.13119	0.11811	106.42746
$N^{\epsilon 2} - H$	−0.08739	−0.10432	−0.08363	0.20098	26.79386
Sum	0.38212	0.05967	−0.09222	6.09781	692.48030
Shifted to Origin	0.47476	1.03326	1.17254	1.63336	

Table 4.81: Side Chain Integrated Properties: *Histidine(+)*

Atom	N	q	K	L	E
C^{β}	5.83533	0.16467	37.69869	0.00012	−37.71695
$C^{\beta} - H_1$	0.98323	0.01677	0.62199	0.00011	−0.62229
$C^{\beta} - H_2$	0.99138	0.00862	0.62343	0.00011	−0.62373
C^{γ}	5.57718	0.42282	37.56279	0.00047	−37.58099
$N^{\delta 1}$	8.49999	−1.49999	55.37506	0.00015	−55.40188
$N^{\delta 1} - H$	0.49245	0.50755	0.40479	0.00006	−0.40499
$C^{\delta 2}$	5.48637	0.51363	37.50911	0.00011	−37.52728
$C^{\delta 2} - H$	0.79615	0.20385	0.53749	0.00001	−0.53775
$C^{\epsilon 1}$	4.58198	1.41802	36.94176	0.00329	−36.95965
$C^{\epsilon 1} - H$	0.83009	0.16991	0.55038	0.00003	−0.55065
$N^{\epsilon 2}$	8.50239	−1.50239	55.38134	0.00034	−55.40816
$N^{\epsilon 2} - H$	0.47985	0.52015	0.39790	0.00004	−0.39810
Sum	43.05640	0.94360	263.60473	0.00484	−263.73240

Table 4.82: Side Chain Integrated Properties: *Histidine(+)* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.17031	0.09216	0.06425	8.15449	54.78478
$C^{\beta} - H_1$	−0.03903	0.11932	0.00730	0.20403	46.14089
$C^{\beta} - H_2$	0.01260	−0.01818	0.12194	0.12575	47.27279
C^{γ}	0.74063	−0.13369	−0.28981	0.12393	62.97771
$N^{\delta 1}$	−0.08243	−0.16031	−0.14737	0.80647	101.96702
$N^{\delta 1} - H$	−0.02296	0.09528	0.10267	0.23284	24.18898
$C^{\delta 2}$	0.77192	0.02733	−0.17670	0.14194	73.99357
$C^{\delta 2} - H$	0.03962	−0.05199	−0.08016	0.79236	34.07112
$C^{\epsilon 1}$	−0.90910	0.02705	0.22510	0.10351	53.84720
$C^{\epsilon 1} - H$	−0.10532	0.00340	0.02582	0.93694	39.87252
$N^{\epsilon 2}$	0.00364	0.16778	0.17033	0.10849	103.06972
$N^{\epsilon 2} - H$	−0.07006	−0.09029	−0.07770	0.23911	23.49739
Sum	0.51004	0.07787	−0.05434	11.96985	665.68369
Shifted to Origin	7.43240	0.14864	0.08156	7.43433	

Table 4.83: Side Chain Integrated Properties: *Phenylalanine*

Atom	no	N	q	K	L	E
C^{β}		5.85269	0.14731	37.71566	0.00326	−37.73512
$C^{\beta} - H_1$		1.02491	−0.02491	0.64061	0.00012	−0.64094
$C^{\beta} - H_2$		1.04389	−0.04389	0.64895	0.00011	−0.64928
C^{γ}		6.01476	−0.01476	37.83723	0.00128	−37.85675
$C^{\delta 2}$		5.98417	0.01583	37.81030	0.00038	−37.82981
$C^{\delta 2} - H$		1.01975	−0.01975	0.63311	0.00001	−0.63344
$C^{\delta 1}$		5.98649	0.01351	37.81214	0.00013	−37.83165
$C^{\delta 1} - H$		1.01775	−0.01775	0.63208	0.00009	−0.63241
$C^{\epsilon 2}$		5.98560	0.01440	37.80693	0.00054	−37.82544
$C^{\epsilon 2} - H$		1.00798	−0.00798	0.62739	0.00006	−0.62771
$C^{\epsilon 1}$		5.98283	0.01717	37.80544	0.00049	−37.82495
$C^{\epsilon 1} - H$		1.00628	−0.00628	0.62676	0.00006	−0.62708
C^{ζ}		5.98260	0.01740	37.80359	0.00060	−37.82310
$C^{\zeta} - H$		1.00711	−0.00711	0.62674	0.00006	−0.62706
Sum		48.91681	0.08319	269.02592	0.00719	−269.16473

Table 4.84: Side Chain Integrated Properties: *Phenylalanine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.04229	0.00747	−0.01560	2.49994	52.38684
$C^{\beta} - H_1$	0.00548	0.12901	−0.00228	0.04569	48.62448
$C^{\beta} - H_2$	−0.03988	−0.01335	−0.12017	0.12914	48.56497
C^{γ}	0.01505	0.01643	−0.02472	0.12732	68.40670
$C^{\delta 2}$	−0.00088	−0.01156	−0.06809	0.03328	52.05956
$C^{\delta 2} - H$	−0.06116	−0.04023	−0.09841	0.06907	50.07356
$C^{\delta 1}$	0.01049	0.05604	0.02887	0.12265	81.86828
$C^{\delta 1} - H$	−0.04322	0.09131	0.06757	0.06391	49.72083
$C^{\epsilon 2}$	0.04162	−0.04361	−0.03967	0.12154	84.01131
$C^{\epsilon 2} - H$	0.04398	−0.08959	−0.07119	0.07217	49.46137
$C^{\epsilon 1}$	0.04965	0.02306	0.04426	0.12259	83.15158
$C^{\epsilon 1} - H$	0.06457	0.04176	0.09522	0.07040	49.33004
C^{ζ}	0.07111	−0.02405	0.01053	0.12239	84.37603
$C^{\zeta} - H$	0.10908	−0.04929	0.02589	0.07580	49.42157
Sum	0.30819	0.09340	−0.16779	3.67590	881.45712
Shifted to Origin	−0.02655	−0.05551	−0.18305	0.19312	

Table 4.85: Side Chain Integrated Properties: *Tyrosine*

Atom	N	q	K	L	E
C^{β}	5.85037	0.14963	37.70919	0.00037	−37.72799
$C^{\beta} - H_1$	1.02768	−0.02768	0.64174	0.00011	−0.64206
$C^{\beta} - H_2$	1.04610	−0.04610	0.64990	0.00011	−0.65023
C^{γ}	5.99308	0.00692	37.81941	0.00189	−37.83826
$C^{\delta 1}$	5.97881	0.02119	37.81198	0.00059	−37.83082
$C^{\delta 1} - H$	1.01219	−0.01219	0.63035	0.00009	−0.63066
$C^{\delta 2}$	5.98165	0.01835	37.81362	0.00026	−37.83247
$C^{\delta 2} - H$	1.01429	−0.01429	0.63152	0.00004	−0.63183
$C^{\epsilon 1}$	5.95940	0.04060	37.79354	0.00033	−37.81238
$C^{\epsilon 1} - H$	1.02028	−0.02028	0.62990	0.00008	−0.63021
$C^{\epsilon 2}$	5.93238	0.06762	37.78739	0.00041	−37.80622
$C^{\epsilon 2} - H$	0.98284	0.01716	0.61517	0.00007	−0.61548
C^{ζ}	5.46659	0.53341	37.49707	0.00093	−37.51577
O^{η}	9.27351	−1.27351	75.50903	0.00004	−75.54666
$O^{\eta} - H$	0.37392	0.62608	0.34430	0.00001	−0.34447
Sum	56.91306	0.08694	343.88411	0.00533	−344.05551

Table 4.86: Side Chain Integrated Properties: *Tyrosine* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.04713	0.00791	−0.01522	1.97843	52.91628
$C^{\beta} - H_1$	−0.00202	0.12972	0.00129	0.05015	48.83319
$C^{\beta} - H_2$	−0.04173	−0.01308	−0.11986	0.12974	48.62931
C^{γ}	−0.01566	0.03122	−0.03792	0.12759	68.66173
$C^{\delta 1}$	−0.03792	0.05384	−0.00767	0.05156	80.29434
$C^{\delta 1} - H$	−0.04467	0.08510	0.07302	0.06630	49.44167
$C^{\delta 2}$	−0.02094	0.01374	−0.04594	0.12070	80.56672
$C^{\delta 2} - H$	−0.05987	−0.03837	−0.09886	0.05233	49.70152
$C^{\epsilon 1}$	−0.01869	0.10650	0.09892	0.12177	83.42162
$C^{\epsilon 1} - H$	0.07208	0.03335	0.09434	0.14655	49.81465
$C^{\epsilon 2}$	−0.04519	−0.08255	−0.16234	0.12332	82.55963
$C^{\epsilon 2} - H$	0.04314	−0.08338	−0.07872	0.18764	47.76152
C^{ζ}	−0.74915	0.31626	−0.14122	0.12252	61.04221
O^{η}	−0.11961	0.17299	0.15642	0.82534	121.35220
$O^{\eta} - H$	0.05825	0.05366	0.12840	0.26210	19.52394
Sum	−0.93486	0.78691	−0.15536	4.36606	944.52053
Shifted to Origin	0.06817	−0.39506	−0.64596	0.76026	

Table 4.87: Side Chain Integrated Properties: *Tyrosine (-)*

Atom	N	q	K	L	E
C^{β}	5.84190	0.15810	37.69087	0.00251	−37.70924
$C^{\beta} - H_1$	1.06329	−0.06329	0.65560	0.00012	−0.65592
$C^{\beta} - H_2$	1.07951	−0.07951	0.66274	0.00009	−0.66306
C^{γ}	6.01656	−0.01656	37.79855	0.00027	−37.81697
$C^{\delta 1}$	6.04553	−0.04553	37.83577	0.00040	−37.85421
$C^{\delta 1} - H$	1.07885	−0.07885	0.65460	0.00008	−0.65492
$C^{\delta 2}$	6.04232	−0.04232	37.83470	0.00068	−37.85314
$C^{\delta 2} - H$	1.06226	−0.06226	0.65637	0.00015	−0.65669
$C^{\epsilon 1}$	6.04812	−0.04812	37.80811	0.00008	−37.82653
$C^{\epsilon 1} - H$	1.05041	−0.05041	0.63907	0.00007	−0.63938
$C^{\epsilon 2}$	6.05003	−0.05003	37.80615	0.00057	−37.82457
$C^{\epsilon 2} - H$	1.05370	−0.05370	0.63996	0.00008	−0.64027
$C^{\zeta 1}$	4.94141	1.05859	37.14437	0.00204	−37.16247
O^{η}	9.41727	−1.41727	75.47013	0.00013	−75.50691
Sum	56.81116	−0.81116	343.29699	0.00727	−343.46431

Table 4.88: Side Chain Integrated Properties: *Tyrosine (-)* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.18805	−0.03771	−0.00358	8.10523	51.74808
$C^{\beta} - H_1$	−0.01561	0.14076	0.00075	0.19183	50.43015
$C^{\beta} - H_2$	−0.05806	−0.00375	−0.12496	0.14163	49.86126
C^{γ}	0.03006	−0.01276	−0.00603	0.13784	74.16425
$C^{\delta 1}$	0.07113	−0.01474	−0.00348	0.03321	81.86024
$C^{\delta 1} - H$	−0.04974	0.09402	0.07714	0.07272	53.91000
$C^{\delta 2}$	0.07620	−0.01436	0.00357	0.13140	81.68871
$C^{\delta 2} - H$	−0.07434	−0.04156	−0.10114	0.07763	54.14970
$C^{\epsilon 1}$	0.10881	−0.03586	0.00128	0.13222	90.81956
$C^{\epsilon 1} - H$	0.06267	0.06390	0.11061	0.11457	52.63742
$C^{\epsilon 2}$	0.11021	−0.03169	0.01200	0.14228	91.18401
$C^{\epsilon 1} - H$	0.02941	−0.09750	−0.10101	0.11530	52.82257
C^{ζ}	−0.73723	0.28082	−0.10764	0.14343	48.88412
O^{η}	−0.27909	0.10850	−0.04098	0.79622	148.58076
Sum	−0.53753	0.39806	−0.28346	10.33551	982.74086
Shifted to Origin	8.90344	−0.44873	−0.09434	8.91524	

Table 4.89: Side Chain Integrated Properties: *Triptophan*

Atom	N	q	K	L	E
C^{β}	5.78290	0.21710	37.66209	0.00094	−37.72268
$C^{\beta} - H_1$	1.05328	−0.05328	0.65694	0.00011	−0.65800
$C^{\beta} - H_2$	1.06945	−0.06945	0.66339	0.00012	−0.66446
C^{γ}	5.96732	0.03268	37.79603	0.00024	−37.85683
$C^{\delta 1}$	5.44874	0.55126	37.45019	0.00060	−37.51044
$C^{\delta 1} - H$	1.00674	−0.00674	0.63127	0.00004	−0.63228
$N^{\epsilon 1}$	8.59004	−1.59004	55.27102	0.00031	−55.35993
$N^{\epsilon 1} - H$	0.52540	0.47460	0.42157	0.00004	−0.42225
$C^{\epsilon 2}$	5.49616	0.50384	37.54401	0.00006	−37.60441
$C^{\delta 2}$	5.99863	0.00137	37.85811	0.00018	−37.91902
$C^{\epsilon 3}$	5.96068	0.03932	37.76216	0.00018	−37.82291
$C^{\epsilon 3} - H$	1.04372	−0.04372	0.64782	0.00003	−0.64886
$C^{\zeta 3}$	5.95199	0.04801	37.74684	0.00053	−37.80756
$C^{\zeta 3} - H$	1.04297	−0.04297	0.64760	0.00008	−0.64864
$C^{\eta 2}$	5.95034	0.04966	37.75042	0.00009	−37.81115
$C^{\eta 2} - H$	1.03942	−0.03942	0.64663	0.00006	−0.64767
$C^{\zeta 2}$	5.94090	0.05910	37.74780	0.00028	−37.80853
$C^{\zeta 2} - H$	1.03944	−0.03944	0.64342	0.00007	−0.64445
Sum	68.90815	0.09185	399.54729	0.00396	−400.19006

Table 4.90: Side Chain Integrated Properties: *Triptophan* (continued)

Atom	μ_x	μ_y	μ_z	$\lvert\mu\rvert$	Vol(0.001)
C^{β}	0.02451	−0.01875	0.00897	3.47732	52.07722
$C^{\beta} - H_1$	0.02184	0.11126	−0.02292	0.03214	49.09657
$C^{\beta} - H_2$	−0.05317	−0.01805	−0.09901	0.11568	49.79939
C^{γ}	−0.06393	0.15595	0.00690	0.11382	69.11108
$C^{\delta 1}$	−0.71025	0.38648	−0.20938	0.16869	73.31921
$C^{\delta 1} - H$	−0.04855	−0.05735	−0.07040	0.83526	48.70837
$N^{\epsilon 1}$	−0.07988	0.14676	0.03504	0.10296	107.38309
$N^{\epsilon 1} - H$	0.07083	−0.14401	−0.04350	0.17073	25.28745
$C^{\epsilon 2}$	0.27654	0.52192	0.51255	0.16628	59.83183
$C^{\delta 2}$	−0.09915	0.14316	0.01804	0.78204	71.07494
$C^{\epsilon 3}$	0.02824	0.04795	0.02749	0.17508	82.38512
$C^{\epsilon 3} - H$	−0.06635	0.08054	0.00804	0.06207	50.76215
$C^{\zeta 3}$	0.02004	0.04553	0.02483	0.10466	84.19315
$C^{\zeta 3} - H$	0.02375	0.08118	0.06632	0.05560	49.92343
$C^{\eta 2}$	0.02017	0.04081	0.03074	0.10748	82.87055
$C^{\eta 2} - H$	0.09143	0.00140	0.05537	0.05493	49.85361
$C^{\zeta 2}$	0.12299	0.00939	0.06903	0.10690	83.52064
$C^{\zeta 2} - H$	0.06489	−0.08313	−0.01614	0.14135	50.08513
Sum	−0.35606	1.45103	0.40199	6.77296	1139.28294
Shifted to Origin	−0.03994	0.69881	0.26988	0.75018	

4.5 Transferability of Amidic Surface Properties and Form

Chang and Bader were the first to draw attention to the complementarity of amidic surfaces. In their study of the transferability of amino acid residues of small β-sheet polypeptides (Chang and Bader 1992), it was observed that the shape of amidic surfaces of zero-flux were virtually superimposable. This observation was made by means of plotting two-dimensional maps of the gradient paths which, upon originating at the bond critical point of an amide bond and delineating their respective gradient paths, delimit the atoms involved in the bonding interaction from one another. When these plots were superimposed over contour plots of the electron density at the same planes and scale they noticed that the distribution of the electron density was equally superimposable as well. In essence, Chang and Bader's study showed that amidic surfaces, just like the distribution of the electron density for amino acid residues, exhibited transferability. These conclusions are qualitative, i.e. drawn by the simple visual inspection of the plotted systems. In this section we provide means of quantifying the transferability of surfaces, and applied them to the systems casted with the glycine mold. This quantification is done by studying

- The transfer in three dimensions of the geometries of these surfaces.
- The transfer in the integrated properties of these surfaces.

Chang and Bader's observation of the complementarity of amidic surfaces has its physical foundation in Gauss' flux conservation Law. Such law states that the charge of a object is the result of the net flux in the electric field through this object's surface. Because a residue in a protein not participating in hydrogen bonding, disulfide bridges or salt bridges, is bounded by two amidic surfaces, the net charge of the residue is determined by the fluxes across these surfaces, the N-Surface and C-Surface. This law can be written as

$$q_{|\mathcal{A}\omega|} = \sum_{i}^{2} q_i(S_i, \Omega) \quad , \tag{4.6}$$

where

$$q(S_i, \Omega) = -\left(\frac{1}{4\pi}\right) \oint dS(r_{S_i})\ \mathbf{n} \cdot \mathbf{E}(r_{S_i}) \quad . \tag{4.7}$$

It is the direct consequence of 4.6 that if a residue bears no charge, the flux in the electric field through the N-Surface should be equal in magnitude and opposite in direction to that through the C-Surface. This inter-relationship between the amidic surfaces has already been introduced in the preceding section as *surface complementarity*, and they are the consequence of the zero net charge

exhibited by the $|Aa|$. Thus, because the $Gly'|$ exhibited a net charge of +.5 e, there is a flux of the electric field through the $Gly'|Aa$ surface into the middle amino acid residue whose integrated value amounts to the mentioned net charge. Similarly, because the net charge of the $|Gly''$ group is -.5 e, there is a flux of the electric field into this group from its neighboring $|Aa|$ by the same integrated amount the later fragment received from Gly'.

4.5.1 Transferability of Surface Form

Suppose we were to devise means of measuring how similar two surfaces of zero-flux are in terms of their respective geometric distributions. Such similarity could be computed by superimposing one surface onto another ensuring that their respective origins, their bond critical points, are overlapped. The surfaces are then aligned with respect to one another to ensure that both share the same orientation. After these geometrical manipulations of both surfaces, one could compute the volume of non-overlap between the surfaces. The volume thus calculated is a measure of the mismatch between the surfaces; the larger the volume, the greater the dissimilarity.

After the suitable alignment of the surfaces is performed (a process described in greater detail in section 5.2) the smallest rectangular box which fully encloses the overlapped surfaces is computed. Then, by implementing a Quasi-Random Monte Carlo procedure one can estimate the volume of non-overlap by doing a thorough sampling of points within the rectangular box and tabulating how many of these points lie between the surfaces or not. Sub-random sequences are used to ensure a more homogeneous sampling of the box, which in turn provides faster and more accurate integration results over normal Monte Carlo. This is so since sub-random sequences converge according to N^{-1} rather than $N^{-\frac{1}{2}}$, where N is the number of samples.

The volume scoring function computed corresponds to

$$V = V_{box} \frac{1}{N} \sum_{i}^{N} \delta(r_i) \quad , \tag{4.8}$$

where

$$\delta(r_i) = \begin{cases} 1 & \text{if } r_i \in [S_A, S_B] \\ 0 & \text{otherwise} \end{cases} \tag{4.9}$$

This function has an associated uncertainty given by

$$U = V_{box} \sqrt{\sigma^2} \quad , \tag{4.10}$$

where

$$\sigma^2 = \frac{1}{N}\sum_i^N \delta(r_i) - \frac{1}{N^2}\left(\sum_i^N \delta(r_i)\right)^2 \quad , \tag{4.11}$$

is the variance of the statistical sampling. $N = 10^6$ observations were performed in all cases. To further distinguish between the two possible forms of non-overlap we proposed to subdivide the volume function into two contribution forms. If the sampled points where found in the interval $[S_A, S_B]$ they would be registered by Volume (+) whereas if found in $[S_B, S_A]$ they would contribute to Volume (-). In addition we further cataloged the non-overlap regions by subdividing the box into spherical sectors whose radii are referenced to the point where the surfaces' bond critical points are superimposed. Sub-classifying the volume function in this manner provides a quantitative description of the contribution to the non-overlap volume in terms of the radial distribution from the centroid of the surfaces.

4.5.2 Transferability of the $|Aa|$ Surface Geometries

Having proposed a methodology for the computation of the similarity between surfaces of zero-flux, we now proceed to study the similarity between the amidic surfaces in the set of $|Aa|$. This entails comparing how similar are the N-Surface and C-Surface within the same residue (*intra-residue* surface similarity), or among different residues (*inter-residual* surface similarity). Computing the later would produce a measurement of the error incurred in carrying out a theoretical synthesis of a two-residue protein.

Because of the transferability of the net charge of the $|Aa|$ demonstrated in the preceding section, computing all the "intra-residue" and "inter-residue" similarity of their respective amidic surfaces would a lot of redundant information. Instead, we decided to compute the surface overlaps which take place in a naturally-occurring protein which will be synthesized in the next chapter. Briefly described, this protein consists of 37 residues, 11 of which are non-repeating. This protein have been found to inhibit ice formation, and thus it is commonly referred to as "AFP", or anti-freeze-protein. Structurally, AFP adopts a near-perfect $3_{3.6}$ α-helical secondary structure conformation, as diagnosed by X-ray studies (Sicheri and Yang 1995). For a more thorough description, the reader is referred to section 5.3.

Two sets of non-overlap volume computations were performed on the residues encountered in AFP. The first corresponds to the intra-residue set of N-Surface/C-Surface computations and is geared at providing a measure of the residue's transferability based on its compliance with Gauss'

Law. The second corresponds to the inter-residue set of amidic surface overlaps providing relative measure of the geometrical error produced upon synthesizing the AFP molecule from the $|Aa|$ which constitutes it, the subject of chapter 5. Table 4.91 displays the results obtained for the first set, whereas table 4.93 displays the results obtained for the second.

Table 4.91: Computation of the non-overlap volume of the intra-residue amidic surfaces.

	Box Volume	IAS inter-volume		Volume (+)		Volume (-)		Uncertainty	
	au^3	au^3	%	au^3	%	au^3	%	au^3	%
$A\|A$	330.215	13.143	3.98	3.987	1.21	9.156	2.77	1.07E-2	3.25E-3
$D^-\|D^-$	462.429	34.040	7.36	11.947	2.58	22.091	4.78	1.69E-2	3.65E-3
$E^-\|E^-$	338.882	14.900	4.40	1.736	0.51	13.169	3.89	1.14E-2	3.36E-3
$K^+\|K^+$	324.215	14.511	4.48	3.833	1.18	10.680	3.29	1.12E-2	3.45E-3
$L\|L$	333.719	12.880	3.86	3.114	0.93	9.769	2.93	1.07E-2	3.20E-3
$N\|N$	333.060	15.909	4.78	4.600	1.38	11.311	3.40	1.17E-2	3.50E-3
$S\|S$	330.186	15.861	4.80	4.489	1.36	11.373	3.44	1.16E-2	3.53E-3
$T\|T$	327.927	8.300	2.53	3.518	1.07	4.782	1.46	8.75E-2	2.67E-3
Ave	347.580	16.193	4.52	4.653	1.28	11.541	3.24	1.16E-2	3.33E-3
AveDev	28.712	4.461	0.84	1.823	0.37	3.0443	0.64	1.34E-3	2.15E-4
σ^2	2172.166	57.969	1.84	9.502	0.35	24.184	0.91	5.46E-6	9.33E-8
σ	46.605	7.614	1.36	3.083	0.60	4.918	0.95	2.34E-3	3.05E-4
Min	324.215	8.300	2.53	1.736	0.51	4.782	1.46	8.75E-3	2.67E-3
Max	462.429	34.040	7.36	11.947	2.58	22.091	4.78	1.69E-2	3.66E-3

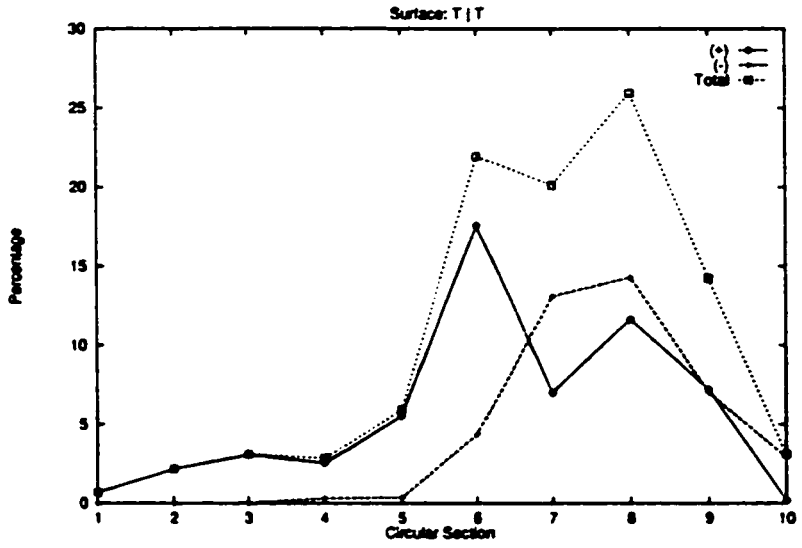


Figure 4.12: Intra-Residue "best" amidic surface overlap.

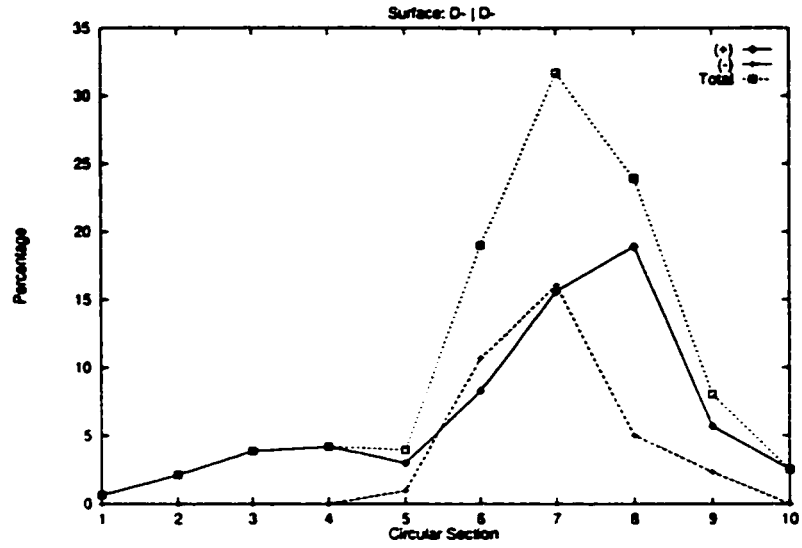


Figure 4.13: Intra-Residue "worst" amidic surface overlap.

The examination of table 4.91 reveals that the N-Surface and C-Surface in our study are complementary, with the exception of aspartate. The most transferable residue in the intra-residue series is threonine, with a non-overlap volume of 8.3 au^3. This value represents 2.53% of the volume of the minimal integration box which contains both N-Surface and C-Surface in threonine. This is a very good result if one considers that most of the non-overlap volume is produced in the outer limits of the box which encloses the overlapped surfaces. Figure 4.12 depicts the contributions to the non-overlap error in terms of 10 cylindrical sectors. As expected, the error increases as one moves away from the origin of the overlapped surfaces, becoming more pronounced at circular sector number 4 and reaching a maximum value at cylindrical sector number 8. Table 4.92 provides a description of the ranges in the electron density of the amidic surfaces in this study as per circular sector. As applied to threonine, the greatest contributions to the non-overlap error take place in the regions with density values of 0.04 au or less. The error distribution shows a decrease after the cylindrical section 8. This does not mean that the surfaces become more similar in the remaining cylindrical sectors, but is the result that the points sampled in this region by the Monte Carlo procedure cease to be bound by any one of the surfaces, or none at all, i.e. they are either bound between a surface and the walls of the box, or simply by the walls of the box itself.

Table 4.92: Electron density ranges per cylindrical sector, in atomic units.

Section	ρ	Section	ρ
1	ρ_b - .2	6	.02 - .008
2	.2 - .08	7	.02 -.008
3	.08 - .04	8	.008 - .004
4	.04 - .02	9	.004 - .002
5	.04 - .02	10	.002 -.001

The amino acid residue which showed the worst overlaps in the intra-residue series was found to be aspartate. This comes as no surprise as this residue exhibited an internal hydrogen bonding involving its N-terminus. Figure 4.13 depicts the distribution of the non-overlap error. there is a sharp increase in the region where the hydrogen bond takes place, around circular disk 7.

The rest of the residues in the series showed a great degree of constancy in their overlap values, averaging 13.64 au^3 (discounting aspartate), or 4.12% inter-surface volume, as computed with respect to their respective integration boxes.

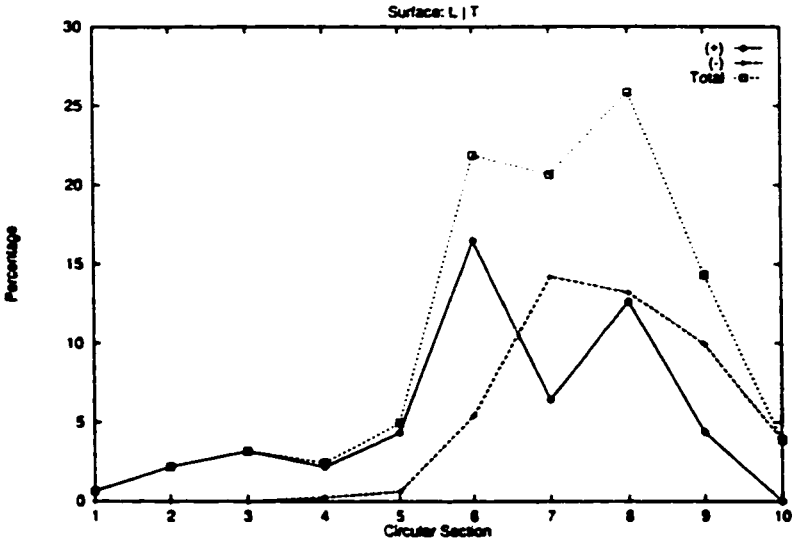


Figure 4.14: Inter-Residue "best" overlap.

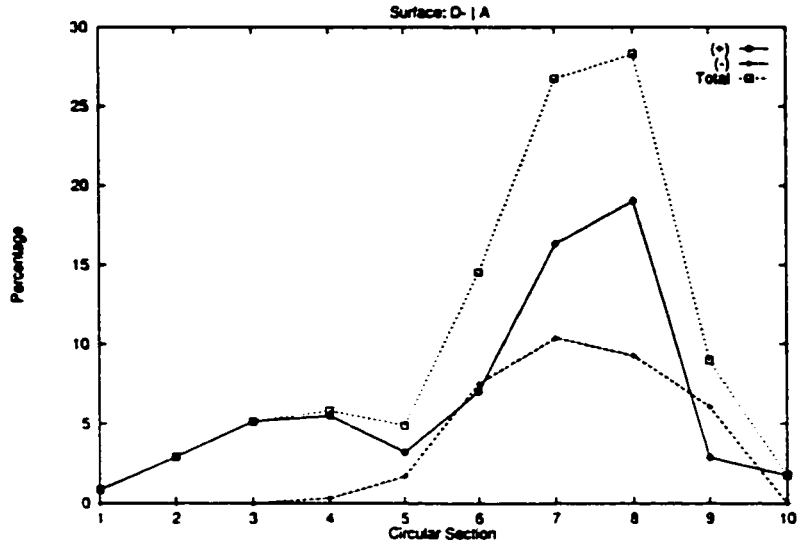


Figure 4.15: Inter-Residue "worst" overlap.

Table 4.93: Computation of the non-overlap volume of the inter-residue amidic surfaces.

	Box Volume	IAS inter-volume		Volume (+)		Volume (-)		Uncertainty	
	au^3	au^3	%	au^3	%	au^3	%	au^3	%
$A\|E^-$	338.043	13.879	4.11	1.611	0.48	1.226	3.63	3.71E-3	1.10E-3
$A\|K^+$	329.608	13.876	4.21	4.671	1.42	9.204	2.79	3.70E-3	1.12E-3
$A\|L$	332.474	12.933	3.89	3.512	1.06	9.420	2.83	3.58E-3	1.08E-3
$A\|N$	330.543	14.669	4.44	4.101	1.24	10.568	3.20	3.80E-3	1.15E-3
$A\|S$	329.373	14.692	4.46	3.891	1.18	10.800	3.28	3.81E-3	1.16E-3
$D^-\|A$	368.532	22.546	6.12	7.960	2.16	14.586	3.96	4.71E-3	1.28E-3
$E^-\|L$	333.618	13.730	4.12	3.515	1.05	10.215	1.05	3.68E-3	1.10E-3
$K^+\|A$	328.111	14.116	4.30	3.255	0.99	10.861	3.31	3.73E-3	1.14E-3
$L\|T$	327.854	8.304	2.53	3.947	1.20	4.357	1.33	2.87E-3	8.57E-4
$N\|A$	330.423	14.579	4.41	4.549	1.38	10.003	3.44	3.79E-3	1.15E-3
$S\|D^-$	429.802	23.119	5.38	7.288	1.70	15.830	3.68	4.78E-3	1.11E-3
$T\|A$	328.270	12.635	3.85	3.042	0.93	9.593	2.92	3.53E-3	1.08E-3
Ave	342.221	14.923	4.32	4.279	1.23	9.722	2.95	3.81E-3	1.11E-3
AveDev	18.982	2.636	0.54	1.226	0.29	2.468	0.64	3.12E-4	5.46E-5
σ^2	886.276	16.617	0.74	3.090	0.17	15.073	0.80	2.55E-7	8.37E-9
σ	29.770	4.076	0.86	1.758	0.42	3.882	0.90	5.05E-4	9.15E-5
Min	327.854	8.304	2.53	1.611	0.48	1.226	1.05	2.87E-3	8.75E-4
Max	429.802	23.119	6.12	7.960	2.16	15.830	3.96	4.78E-3	1.28E-3

The inter-residue series also exhibits an average error in the neighborhood of 4 % of the volume of their respective boxes or 13 au^3. By having the most similar N-terminus and C-terminus surfaces, threonine is also the most transferable residue in terms of surface geometries. The two di-residue sequences occurring in AFP which involve this residue are $L|T$ and $T|A$ and the computation of their inter-surface volumes produced the smaller values of the series tabulated in 4.93.

The most dissimilar surfaces in the series are the N-Surface and C-Surface of aspartate. Two di-residue sequences involving aspartate are found in AFP, namely $D^-|A$ and $S|D^-$. The volume computation of the first surface overlap yielded 22.5 au^3 and the second 23.1 au^3 . The N-surface of aspartate was found to be distorted by the presence of a near bcp resulting from an internal hydrogen bonding involving the back-bone's H and an oxygen in the R-group. There is no hydrogen bond involving the C-terminus, so one would think that there is no reason for the residue to exhibit a distorted C-Surface. But the C-Surface in aspartate is in fact different from the rest of the C-surfaces in the series as an indirect result of enforcing a Le Chatelier-like preservation of the residue's properties upon the change introduced by the formation of the hydrogen bond. This preservation process entails the re-distribution of the electron distribution within the residue so as to offset the change. Because the N-Surface experiences a different flux in the electric field it exhibits a different shape as well. The C-Surface tries to change as well in order to match the flux introduced through the N-Surface. The C-Surface does not completely match the change experienced by the N-Surface as the residue yielded -0.921 e as its charge. This observation does not imply that the residue is not transferable since it is unlikely that the residue will remain uncoupled to a counter-ion in vivo. In fact upon coupling aspartate with a sodium cation the residue charge becomes 0.002 e. One could argue that the need for aspartate to stabilize its charge by coupling with the sodium cation is a means used by the residue to regain transferability.

4.5.3 Transferability of Surface Properties

The program SURPROS enables one to determine the properties of an interatomic surface by integrating the corresponding property density over this surface numerically. The surface properties determined in this manner are summarized in tables 4.94 and 4.95. If the two amidic surfaces bounding the residue $|Aa|$, previously labeled the C-surface and the N-surface were perfectly complementary they would exhibit the same property values, excluding their surface virials, which as discussed below, entailed an averaging of a quantity weighted by position vectors referenced to the amidic C and N nuclei respectively. The integration of the quantity $\nabla\rho\cdot$n should yield zero because

of the zero-flux surface condition, eqn. [2.7], a condition that is closely met in nearly all cases. The integrated values for this condition are tabulated under the heading $\mathcal{ZF}$.

The surface averaging of the electron density and of the positive definite kinetic energy density G eqn.[2.9], yield respectively, $\mathcal{N}$, the number of electrons per unit length and $\mathcal{G}$, the electronic kinetic energy per unit length. Both sets of data exhibit relatively small deviations from the mean for each surface and average values for the two surfaces that differ by 1.6% for $\mathcal{N}$ and by 1.2% for $\mathcal{G}$. The surface virial reproduced here as referenced to the C atom

$$\mathcal{V}_s(\Omega) = \oint dS \mathbf{r}_C \cdot \overleftrightarrow{\sigma} \cdot \mathbf{n}_C \tag{4.12}$$

has the dimensions of energy and it is a measure of the contribution of each surface to the potential energy of the total system. The surface virials are referenced to the nuclei of the atoms sharing the interatomic surface through the presence of the position vector that links each nucleus to every point in the interatomic surface, the vectors, $\mathbf{r}_C$ and $\mathbf{r}_N$. Thus, while the C- and the N-surface integrals necessarily exhibit different values, one notes that the variation in their values through the series is small with few exceptions. The average values of the surface virial $\mathcal{V}_{SC}$ for the C-surface integrations differ by 6% while those for the N-surface, $\mathcal{V}_{SN}$, differ by 0.6%.

A more interesting quantity is obtained by summing the surface virials for the C and N atoms sharing an amidic surface. This sum can be expressed as a single surface integral involving the vector $\mathbf{R}_{C-N} = \mathbf{r}_C - \mathbf{r}_N$ linking the two nuclei as

$$\begin{aligned} \mathcal{V}_S(C-N) &= \oint dS \mathbf{r}_C \cdot \overleftrightarrow{\sigma} \cdot \mathbf{n}_C + \oint dS \mathbf{r}_N \cdot \overleftrightarrow{\sigma} \cdot \mathbf{n}_N \\ &= \mathbf{R}_{C-N} \cdot \oint dS \overleftrightarrow{\sigma} \cdot \mathbf{n}_C \quad . \end{aligned} \tag{4.13}$$

Equation [4.13] is a consequence of the unit surface vector perpendicular to the C-surface,$\mathbf{n}_C$, being the negative of that for the N-surface. It equates the surface virial for an amidic surface to a term that is independent of the two electronic position vectors and is instead proportional to the separation between the amidic C and N nuclei. Since a surface virial has the dimensions of energy the addition of two surface virials, as is done in eqn.[4.13] to obtain $\mathcal{V}_{SC-N}$, provides a measure of the contribution of the formation of the amidic surface to the total energy of the molecule. The average values of $\mathcal{V}_{SC-N}$ for the C- and N-surfaces differ by 12 kcal/mol out of a surface energy of the order of 600 kcal/mol, or 2 %. The magnitude of the energy change, relative to the isolated atom energies, undergone by the amidic carbon and nitrogen atoms in forming the amino acid residues are both of the order of 580 kcal/mol. This value, once again, far exceeds the average difference in the contributions to the energy change calculated for the two amidic surfaces.

When an $|A\alpha|$ is built into a polypeptide, the surface virials that are involved are $\mathcal{V}_{SC}$ of the C-surface and $\mathcal{V}_{SN}$ of the N-surface. Thus, the important measure of the transferability of the surfaces in the construction of the polypeptide is the difference in the separate values of $\mathcal{V}_{SC}$ and $\mathcal{V}_{SN}$ between the C- and N-surface. These are 9 kcal/mol for $\mathcal{V}_{SC}$ and 4 kcal/mol for $\mathcal{V}_{SN}$. The surface properties provide stringent measures of the constancy in the amidic surfaces and the resulting differences recorded here are in all cases small relative to the values of the quantities involved, pointing to an acceptable degree of their transferability.

Table 4.94: Summary of N-Surface integrations

Aa	$\mathcal{N}$	$\mathcal{G}$	$\mathcal{ZF}$	$\mathcal{V}_{SC}$	$\mathcal{V}_{SN}$	$\mathcal{V}_{SC-N}$
ala	1.5329	1.4924	0.0005	-0.2251	-1.0344	-1.260
arg	1.6170	1.5768	0.0048	-0.2818	-1.0010	-1.283
asn	1.5397	1.5060	0.0007	-0.2273	-1.0424	-1.270
asp	1.5272	1.4920	0.0001	-0.2254	-1.0319	-1.257
cys	1.5347	1.4980	0.0004	-0.2240	-1.0366	-1.261
glu	1.5328	1.4932	0.0005	-0.2246	-1.0319	-1.257
gln	1.5392	1.4987	0.0004	-0.2278	-1.0276	-1.255
gly	1.5290	1.4913	0.0006	-0.2240	-1.0368	-1.261
his	1.5387	1.4994	0.0007	-0.2257	-1.0387	-1.264
ile	1.5421	1.4999	0.0004	-0.2299	-1.0209	-1.251
leu	1.5418	1.4999	0.0007	-0.2262	-1.0371	-1.263
lys	1.5373	1.4961	0.0007	-0.2248	-1.0364	-1.261
met	1.5348	1.4946	0.0002	-0.2266	-1.0260	-1.253
phe	1.5293	1.4896	0.0005	-0.2283	-1.0166	-1.245
ser	1.5364	1.5037	0.0004	-0.2276	-1.0429	-1.271
thr	1.5395	1.4984	0.0011	-0.2257	-1.0328	-1.259
trp	1.5380	1.4949	0.0006	-0.2278	-1.0189	-1.247
tyr	1.5377	1.4952	0.0004	-0.2277	-1.0190	-1.247
val	1.5414	1.4989	0.0004	-0.2298	-1.0208	-1.251
Ave	1.5405	1.5010	0.0007	-0.2295	-1.0291	-1.2586
AveDev	0.0085	0.0088	0.0005	0.0056	0.0086	0.0069
σ^2	3.62E-4	3.55E-4	1.01E-6	1.64E-4	1.14E-4	8.77E-5
σ	1.90E-2	1.88E-2	1.01E-3	1.28E-2	1.07E-2	9.36E-3
Max	1.6170	1.5768	0.0048	-0.2240	-1.0010	-1.2450
Min	1.5272	1.4896	0.0001	-0.2818	-1.0429	-1.2829
pro	1.4538	1.5170	0.0001	-0.2616	-0.9452	-1.207
arg(+)	1.6626	1.6244	0.0025	-0.2922	-0.9800	-1.272
lys(+)	1.5652	1.5195	0.0074	-0.2738	-0.9778	-1.252
his(+)	1.6000	1.5573	0.0086	-0.3106	-0.9562	-1.267
Ave	1.6092	1.5671	0.0062	-0.2922	-0.9713	-1.264
asp(−)	1.5962	1.5481	0.0002	-0.2528	-1.0404	-1.293
glu(−)	1.5530	1.5081	0.0009	-0.2311	-1.0492	-1.280
tyr(−)	1.5499	1.5031	0.0005	-0.2295	-1.0431	-1.272
Ave	1.5664	1.5198	0.0005	-0.2378	-1.0442	-1.282
lys(+)*Cl*(−)	1.5353	1.4960	0.0007	-0.2242	-1.0366	-1.261
asp(−)*Na*(+)	1.5704	1.5299	0.0009	-0.2469	-1.0332	-1.280

Table 4.95: Summary of C-Surface integrations

Aa	$\mathcal{N}$	$\mathcal{G}$	$\mathcal{ZF}$	$\mathcal{V}_{SC}$	$\mathcal{V}_{SN}$	$\mathcal{V}_{SC-N}$
ala	1.5407	1.4913	0.0010	-0.2086	-1.0207	-1.229
arg	1.6191	1.5706	0.0016	-0.2503	-1.0040	-1.254
asn	1.5569	1.5049	0.0003	-0.2311	-1.0026	-1.234
asp	1.6793	1.6676	0.0015	-0.1346	-1.1680	-1.303
cys	1.5465	1.4931	0.0001	-0.2261	-1.0030	-1.229
glu	1.5481	1.4996	0.0011	-0.2103	-1.0253	-1.236
gln	1.5495	1.5030	0.0007	-0.2363	-1.0042	-1.241
gly	1.5194	1.4725	0.0009	-0.2185	-1.0090	-1.227
his	1.5560	1.5136	0.0008	-0.2149	-1.0309	-1.246
ile	1.5974	1.5500	0.0017	-0.1929	-1.0055	-1.198
leu	1.5438	1.4962	0.0008	-0.2137	-1.0189	-1.232
lys	1.5467	1.4985	0.0006	-0.2162	-1.0161	-1.232
met	1.5482	1.5004	0.0006	-0.2171	-1.0190	-1.236
phe	1.5509	1.4973	0.0010	-0.2076	-1.0165	-1.224
ser	1.5507	1.4983	0.0010	-0.2114	-1.0205	-1.232
thr	1.6233	1.5969	0.0017	-0.2838	-1.0420	-1.326
trp	1.5508	1.4981	0.0010	-0.2099	-1.0159	-1.226
tyr	1.5505	1.4974	0.0008	-0.2092	-1.0157	-1.225
val	1.5769	1.5308	0.0015	-0.2060	-1.0072	-1.213
Ave	1.5660	1.5200	0.0010	-0.2157	-1.0234	-1.2391
AveDev	0.0280	0.0332	0.0003	0.0162	0.0182	0.0183
σ^2	1.45E-3	2.19E-3	2.02E-7	7.81E-4	1.33E-3	8.49E-4
σ	3.81E-2	4.68E-2	4.50E-4	2.79E-2	3.65E-2	2.91E-2
Max	1.6793	1.6676	0.0017	-0.1346	-1.0026	-1.1984
Min	1.5194	1.4725	0.0001	-0.2838	-1.1680	-1.3258
pro	1.5790	1.5638	0.0005	-0.3108	-0.9936	-1.304
arg(+)	1.6190	1.5650	0.0014	-0.2111	-1.0437	-1.255
lys(+)	1.5440	1.4879	0.0005	-0.2244	-1.0059	-1.230
his(+)	1.5510	1.4972	0.0004	-0.2296	-1.0144	-1.244
Ave	1.5713	1.5167	0.0008	-0.2217	-1.0213	-1.243
asp(−)	1.6190	1.5700	0.0003	-0.2196	-1.0460	-1.266
glu(−)	1.5537	1.5243	0.0006	-0.2122	-1.0424	-1.255
tyr(−)	1.5540	1.5127	0.0009	-0.2030	-1.0300	-1.233
Ave	1.5756	1.5357	0.0006	-0.2116	-1.0395	-1.251
lys(+)*Cl*(−)	1.5485	1.4983	0.0011	-0.2091	-1.0228	-1.232
asp(−)*Na*(+)	1.6190	1.5720	0.0008	-0.2277	-1.0453	-1.2731

Chapter 5

Theoretical Synthesis of Proteins

5.1 Current Approaches to Modeling Macromolecules

The observation that the distribution of the electron density of functional groups in biological molecules is transferable has led several groups to propose methodologies for the creation of libraries of building blocks to be used in the modeling of large systems. The creation of these libraries hinges on the the implementation of partitioning schemes resulting from the expansion of the electron density in terms of atomic-centered basis functions. We present a survey of these approaches.

Molecular Electron Density Lego Approach

The simplest approach to the modeling of large systems from a database of fragments has been proposed by Mezey and co-workers (Walker and Mezey 1993). This procedure aims at creating approximate electron density distributions of macro molecules. To achieve this the authors compute wavefunctions for the functional groups of interest, while in their usual chemical environments, and then partition the density matrix into group-specific sub-matrices for each of the functional groups. This partitioning is implemented by using a scheme analogous to the Mulliken's partition of the atomic overlap matrix for a molecule whose density is expressed in terms of linear combinations of atomic orbitals. The partitioning is arbitrary in that it divides the overlapping atomic basis functions equally between the two contributing atoms. The divided basis functions are then used to compute cubic grid maps of the electron density for the corresponding groups, producing the so-called "fuzzy densities". These fuzzy densities are then merged together at the appropriate orientations in order to represent the electron density distribution of the desired system in a process called "inter-penetration

of the density". The end result is an inexpensive computational procedure which approximates the electron density of the "lego-like" synthesized molecule and which can be used to for visualization purposes in molecular docking and similarity surveys. This approximation makes no provisions for the preservation of charge and only properties determined by the electron density can be estimated over the synthesized molecules.

Kernel Projector Matrices

A more rigorous approach to synthesis has been proposed by Massa and co-workers (Massa *et al.* 1995, Huang *et al.* 1996) who express the density of a large molecule as a sum of densities of arbitrarily chosen fragments. A fragment consists of a kernel, a grouping of atoms of interest, together with its neighborhood, the contributions of orbitals centered on atoms of neighboring kernels. The overlaps of basis functions centered on the kernel and its neighbor are equally divided between the two, as in a Mulliken-type population analysis. The orbitals of each fragment are used to construct a density matrix (a matrix whose diagonal elements determine the electron density) and the density matrix of the entire molecule is expressed as a sum of the fragment matrices and subjected to two constraints: that the resulting matrix for the entire molecule be idempotent and properly normalized. The fragment matrices can be obtained from either theoretical calculations on suitable small systems or from experimentally determined X-ray structure factors. In the latter instance, additional constraints are imposed by requiring that the predicted density matrix reproduce the measured structure factors.

Transferability of Experimental Density Distributions

The work in this thesis finds a close experimental parallel in the work of Lecomte and co-workers. This group (Pichon-Pesme *et al.* 1995) has produced high quality X-ray studies of polypeptides, whose structure factors are subsequently used in the parameterization of analytical expressions of the electron density for these systems. The expression for the electron density is that used in MOLLY (Hansen and Coppens 1978), a function expanded in terms of the contributions from "pseudoatoms"; atom-centered contributions to the molecular electron density. These atomic contributions are in turn partitioned into *core* and a *valence* terms. While the charge in the core term is spherically distributed, the contribution to the electron density by the valence term is asymmetric and expanded in terms of multipole parameters. Lecomte and co-workers (Jelsch *et al.* 1998) have demonstrated that the experimental electron density multipole parameters are transferable for atoms of the same

type in similar environments. Based on this observation the group is building a database of these multipole parameters as encountered in polypeptides to serve a two-fold purpose:

- To provide faster and more accurate refining procedures in the X-ray characterizations of peptides.
- To produce experimental electrostatic potential maps from the fitted multipole parameters.

The electron densities fitted in this manner have been shown to confirm the topological features predicted by theoretical calculations (Benabicha *et al.* 2000).

5.2 AIM's Synthetic Procedure

The previous chapter has shown that the amino acid residues found in proteins are transferable by demonstrating the following:

- The geometries of these residues - backbone bond distances and angles - showed small variations in the set.
- The bond critical point data of the $|Aa|$ backbone bonds also showed small variations.
- The net charge of the residues is close to zero.
- The amidic surfaces in the $|Aa|$ are similar geometrically, as well as in terms of their properties.
- The N-Surfaces and C-Surfaces in the $|Aa|$ are complementary with one another.

Thus, because of the complementarity of the amidic surfaces, these fragments are suitable as building blocks in the synthesis of polypeptides. Synthesis in this context entails detaching the $|Aa|$ from the mold and the subsequent manipulation of the distribution of its electron density - translations and rotations - to fit the locations and orientations required to model the molecule of interest. More specifically, this is done as follows:

- *Detachment of the Functional Group from the mold: Computation of the "Vector of Bounding Points".* This entails the computation of points which, upon triangulation, render a closed surface which completely encapsulates the fragment. For convenience, the set of points is kept in a vector whose elements need not be arranged in any particular order. This vector is henceforth labeled **P**.

1. First, the surfaces of zero-flux which delimit this fragment from the mold are computed. This is accomplished by integrating the trajectories of $\nabla\rho$, the paths in the gradient of the electronic density, away from the bond critical point of the surface under consideration. The path equations are represented by

$$\frac{dr(s)}{ds} = \nabla\rho[r(s)] \quad , \tag{5.1}$$

with $r(s_0) = r(0) = r(bcp)$ as its origin. The paths resulting from the integration of this first order differential equation are given by

$$r(s) = r(s_0) + \int_{s_0}^{s} \nabla\rho[r(t)]dt \quad , \tag{5.2}$$

where s_0 is assigned iteratively the coordinates of the previously found coordinates of s, a process repeated until a desired cut-off criterion is met. Gradient paths are thus expressible as collections of position vectors determined by systematically varying s over its entire range. The collection of computed position vectors (the set of points obtained from following each trajectory in $\nabla\rho$) is then added to the vector of bounding vertexes **P**.

2. Once all the interatomic surfaces have been mapped, the second phase requires the calculation of the isodensity envelope of the regions not delimited from other fragments by surfaces of zero-flux. This isodensity surface merges with the interatomic surfaces computed in (1) to fully enclose the fragment. The choice of iso-density value is usually made to be either 0.001 or 0.002 au. The first recovers the van der Waals volumes of atoms from gas phase kinetic studies, whereas the second recovers the volumes for crystallographic data. This thesis uses $\rho = 0.001$ *au* as its default value. The iso-density envelope is then computed by radially sampling the outer regions of the atoms which make the functional group, and which are not bounded by surfaces of zero-flux. The points thus obtained are entered into **P**, once again, without obeying any particular order.

- *Triangulation of the vector of bounding vertexes* **P**. Having computed **P** such that it provides a relatively homogeneous sampling of the fragment's boundary, one proceeds to triangulate the set. The purpose for the triangulation is to produce relationships among the elements in **P** which will allow the rendering of the fragment as a solid object using ray-tracing techniques.

  It was indicated before that the elements of **P** are not ordered in any particular arrangement. Because of this, we can make no pre-conceived assumptions of the "shape" of the set; the

shape of the fragments is to be reconstructed from the scattered data contained in P. Recent advances in computer graphics, (Edelsbrunner and Mücke 1994) and (Akkiraju *et al.* 1995), have tackled this problem by presenting an algorithm that constructs the entire family of shapes for a given set of size n in time $O(n^2)$ as its worst case. Each shape or member of the family is a well defined polyhedron, derived from the Delaunay triangulation of the set, and determined by a scalar value α which varies between the values 0 (the point set itself) and ∞ (the convex hull of the set). For the intermediate values of α the associated α-shape is a polyhedron that is neither necessarily convex nor necessarily connected. To gain more of an intuitive understanding of the relationship α and the family of α-shapes of P, one could consider this scalar to be the radius of a sphere made to pass through all the triangles made from combinations of the elements in P. Initially, all possible combinations of vertexes are performed rendering triangles that are indexed into an array of triangles. If the test sphere passes through the interior of a given triangle, the index of such triangle is removed from the array of triangles. As the value for α is reduced, so are the dimensions of the sphere and thus more triangles will be eliminated form the array of triangles. As $\lim_{\alpha \to 0}$ the sphere becomes infinitesimally small and thus passes trough every triangle in the array of triangles, returning P as the α-shape of the set.

In practical terms, one needs to choose the best of the infinite number of values that α can take in order to best recover the shape of the set. Edelsberg and Mücke demonstrated (Edelsbrunner and Mücke 1994) that α can be made a function of several other geometrical variables such as the volume of the polyhedron, its surface area, its number of triangular edges, etc. For our purposes we chose to optimize the value for α such that it rendered the maximum number of surface Delaunay triangles in the set. Depicting functional groups in this manner renders images similar to the already familiar CPK[1] models. An example of a residue whose vector of bounding vertexes has been computed and then triangulated in the manner described above is presented in figures 5.1, 5.2 and 5.3. These figures depict different viewing angles of the alanyl fragment after detachment from the glycine mold.

Figure 5.1 presents a view of the alanyl fragment from its side. The R-group ($-CH_3$) is directed towards the top of the image. The amidic interatomic surfaces are oriented towards the bottom of the image. The N-Surface faces the left margin while the C-Surface, located at the right of the image, faces the bottom. Notice the curvature of the interatomic surfaces as

[1] Corey-Pauling-Koltun

well as the sharp edges made as these surfaces join with the isodensity envelope.

Figure 5.2 provides a view of the residue with the amidic surfaces in the forefront. The N-Surface is located to the right of the image whereas the C-Surface is positioned to the left. There are few patches of darker color in the areas of the junction between the surfaces at the center of the image. The appearance of these regions reflect the difficulty associated with triangulating the residues in the regions where sharp edges appear. In the particular case of the alanyl residue, the triangulation algorithm miss-assigned the direction in which to point the normal vector at the darkened triangles. The computation of the normal vector to each triangle is needed for the purpose of ray-tracing the object.

at their amidic surfaces such that the N-Surface of a residue is to matched with the C-Surface of the previously added residue to the strand, repeating the process from then onwards with each new residue. This procedure requires the following information:

(a) The coordinates of the bond critical points of each amidic surface in the $|Aa|$,

(b) the eigenvectors and eigenvalues of the Hessian of ρ computed at these points and

(c) the coordinates of the heavy atoms in the backbone.

The coordinates of the bond critical points are needed in order to translate the $|Aa|$ from their respective coordinate systems to the new coordinate system in which we are to carry out the synthesis. The eigenvalues are used to identify which eigenvector is perpendicular to the bonding axis or pointing along its direction. The eigenvector with associated positive eigenvalue defines the bonding axis. The remaining two eigenvectors are perpendicular to the bonding axis and define a local plane at the bond critical point which expands into the surface of zero-flux locally. One synthesizes an amidic bond by superimposing the local surfaces for both the C-Surface and N-Surface along the bonding axis. If further adjustment of the matching is needed for these surfaces, it is to be performed by rotating one surface with respect to the other along the bonding axis. In numerical terms, the synthesis process is represented by the following algorithm which works on the elements of the vector P as:

$$
\begin{aligned}
&\textit{for } i \ldots \textit{ Residues}: \\
&\qquad \textit{for } j \ldots N: \\
&\qquad\qquad p^{\circ}_{i,j} \;=\; [p_{i,j} - r_{bcp1_i}]\, R^{-1}_{bcp1_{i,(\alpha_1\alpha_2)}}\, R_{bcp2_{i-1,(\alpha_1\alpha_2)}}\, R_{\omega} + r_{bcp2_i-1} \quad . \quad (5.3)
\end{aligned}
$$

The index i refers to the residue order in the synthesis and j indexes the elements of P. The bond critical point at the N-Surface is labeled as $bcp1$ whereas that of the C-Surface is referred as $bcp2$. Each of the R in algorithm [5.3] refers to a rotational transformation.

The sequence of these transformations per fragment i is started by translating the later to the newly defined global origin. Following this, the local plane at the bond critical point of the N-Surface is aligned with the xy plane of the new coordinate system. In order to align these surfaces one needs to reference the geometrical distribution of the fragment with respect to the heavy atoms in its backbone.

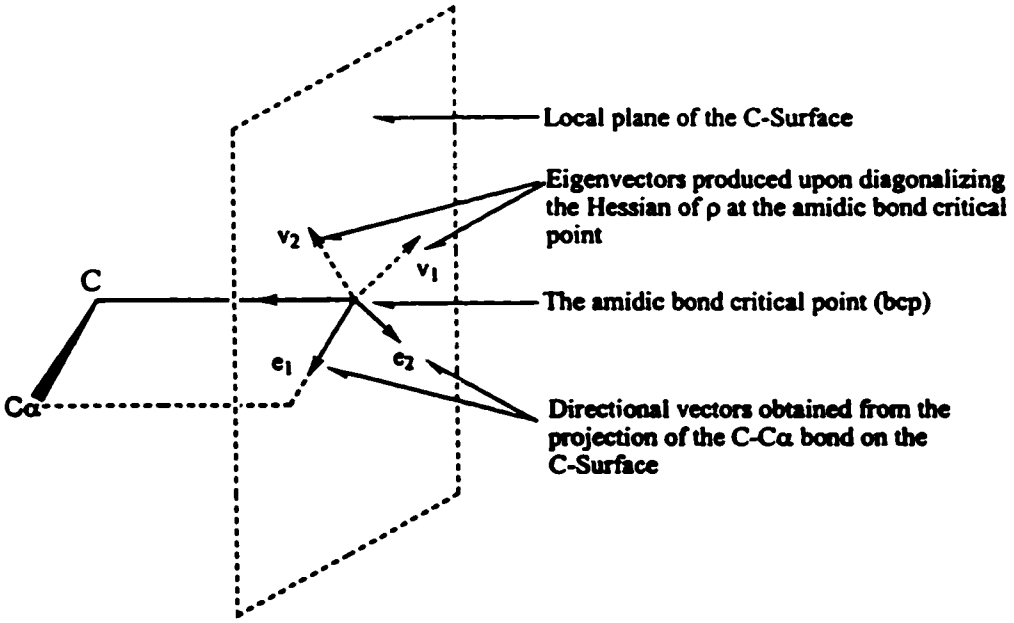


Figure 5.4: Definition of the local coordinate system at the N-Surface

Figure 5.4 illustrates how the redefinition of the surface geometrical distribution takes place. Because the eigenvectors $\mathbf{v}_1$ or $\mathbf{v}_2$ may have an infinite number of orientations along the local plane of the amidic surface (thus making it impossible to use these to align residues with respect to one another in the synthetic process) one needs to re-orient these vectors according to a common criterion inherent to all the residues. We chose to rotate $\mathbf{v}_1$ and $\mathbf{v}_2$ along the bond axis until $\mathbf{v}_1$ overlaps with the projection of the vector $\overrightarrow{CC_\alpha}$. $\mathbf{v}_2$ is rotated as well in order maintain its orthogonal relationship with $\mathbf{v}_1$. The rotated eigenvectors are re-labeled $\mathbf{e}_1$ and $\mathbf{e}_2$ as shown on figure 5.4. The same procedure is carried out in the N-Surfaces, only the projected bond vector is $\overrightarrow{NC_\alpha}$.

Having defined a common criterion for the three-dimensional manipulation of the $|A\alpha|$, and detached the fragment from the mold, the next fragment to be added to the protein strand is rotated as to match the orientation of the N-Surface of its predecessor. Having matched the surfaces one may have to rotate the fragment about the amidic bond axis to the required ω dihedral angle. This corresponds to making the C-Surface of the preceding residue complementary to the N-Surface of the residue to be linked. In our case, and since the protein to be synthesized in the ensuing section is to conform to an optimal $\alpha - R$ helix, the torsion angle

5.3 Synthesis of Winter Flounders Anti-Freeze Protein (AFP)

During our investigation of the possibility of synthesizing a polypeptide by the linking of predetermined fragments the existence of a family of proteins, recently discovered in arctic fish, was brought to our attention. These proteins are particularly interesting in that they have been shown to inhibit ice formation. Their structure is relatively simple since they are mono-stranded near-perfectly α-helical. We realized that modeling one of these proteins would be ideal for the purpose of testing our synthetic procedure in practical terms. For this purpose we chose the protein of the series mentioned above found in Winter Flounders, which from now onwards we will refer to as anti-freeze Protein (AFP). This protein has been characterized by Yang and co-workers (Sicheri and Yang 1995). The 1.5 Å resolution revealed a near perfect α-helical protein, save for the last peptide unit, which adopts a 3_{10}-helix conformation (figures 5.7 and 5.8). The protein is made of 37 residues 8 of which are non-repeating. The last residue is bonded to a $|NH_2$ group.

The sequence of residues in AFP is:

$$ASP^{1}\ THR^{2}\ ALA^{3}\ SER^{4}\ ASP^{5}\ ALA^{6}\ ALA^{7}\ ALA^{8}\ ALA^{9}\ ALA^{10}$$
$$ALA^{11}\ LEU^{12}\ THR^{13}\ ALA^{14}\ ALA^{15}\ ASN^{16}\ ALA^{17}\ LYS^{18}\ ALA^{19}\ ALA^{20}$$
$$ALA^{21}\ GLU^{22}\ LEU^{23}\ THR^{24}\ ALA^{25}\ ALA^{26}\ ASN^{27}\ ALA^{28}\ ALA^{29}\ ALA^{30}$$
$$ALA^{31}\ ALA^{32}\ ALA^{33}\ ALA^{34}\ THR^{35}\ ALA^{36}\ ARG^{37} - (NH2) \quad ,$$

where the superscripts index the $|Aa|$ for future discussions.

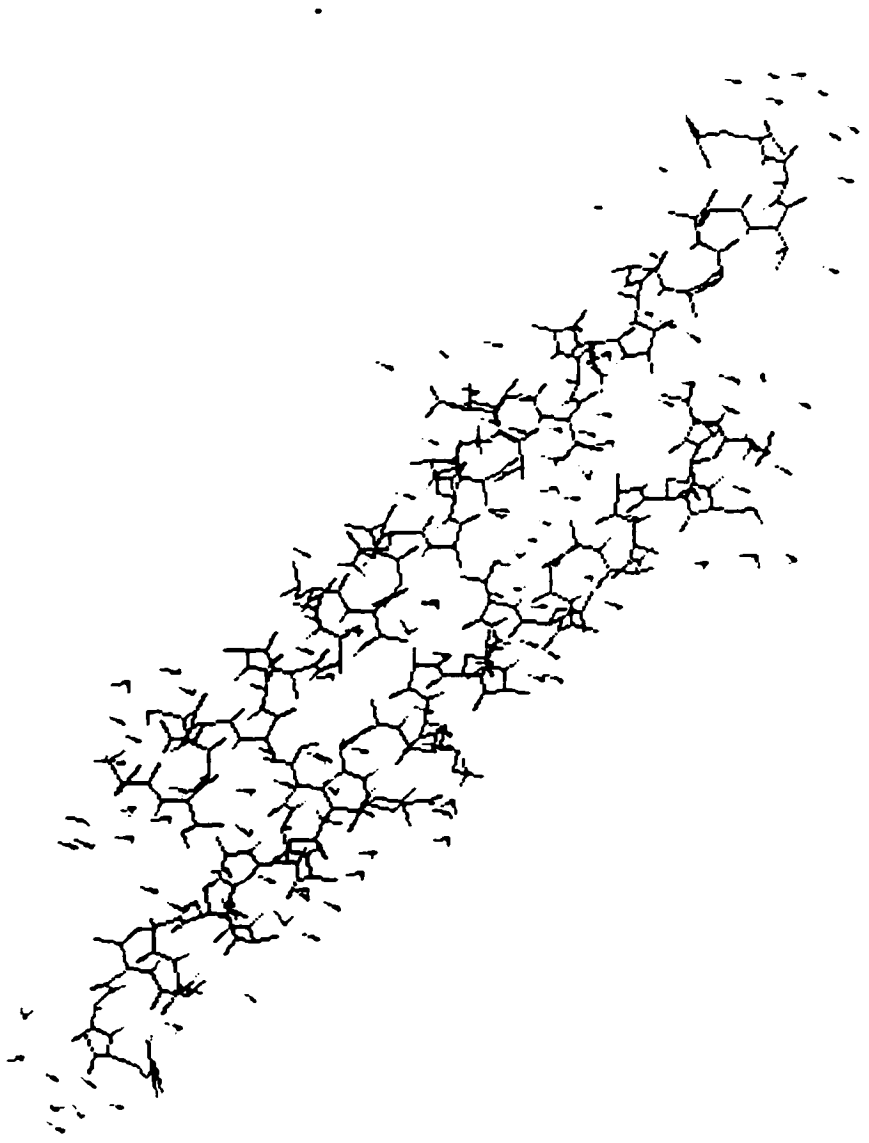


Figure 5.7: Winter Flounders Protein as crystallized in its dimeric form.

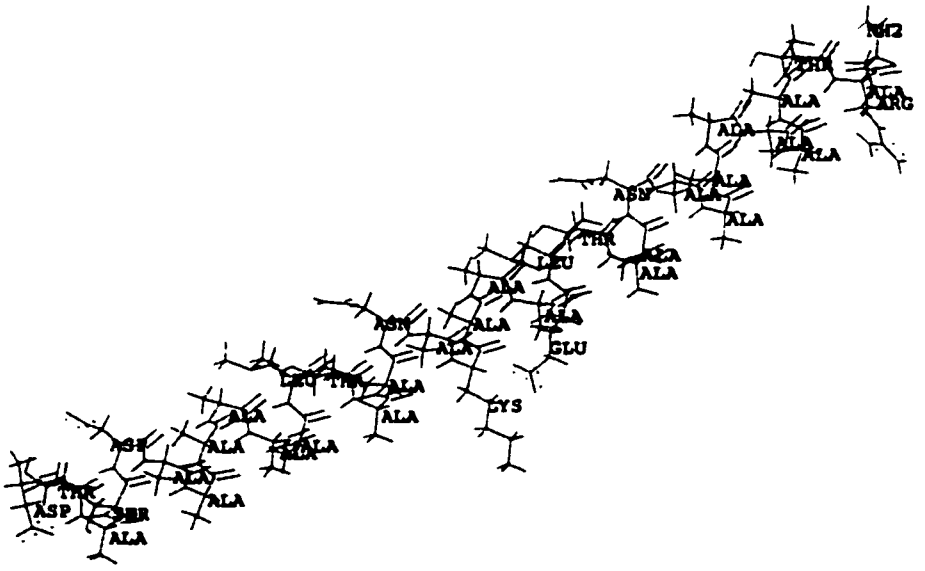


Figure 5.8: Single strand of Winter Flounders AFP.

We proceeded to synthesize AFP such that its secondary structure to be recovered would be an optimal α-helical configuration. To this end we extracted the $|A\alpha|$ from the glycine mold by following the steps indicated in section 5.2. First we removed the residues from the glycine mold by finding its bounding surfaces of zero-flux and iso-density envelopes. This was a relatively computationally expensive procedure as these system's wavefunctions contain a large number of basis functions and nuclei. This procedure generates a vector of bounding vertexes containing an average of 12000 elements. This vector is then subjected to the modified Delaunay triangulation from which one extracts the α-shape of the object. The values for α in the triangulations were chosen on an individual basis as the topological features of the different residues vary. Nonetheless, the criterion used in the choice of the α parameter is that it produced the maximum number of Delaunay triangles on the surface of the object, over internal or external triangles. Let us remember that the Delaunay algorithm used computed and indexed all the possible triangles that can be formed in the set. By choosing a value for α some of these triangles are removed from the set of triangles. The triangulating information of each of the residues in AFP was then stored in VRML format, a recent standard for graphic files largely based on OpenGL. Web-browser plugins and software for the visualization of these files are widely available as shareware for all computer platforms.

The residues, once detached from the mold and stored in their respective graphic files, where used in the synthesis of AFP. This entailed the application of algorithm [5.3] to each one of the $|A\alpha|$ with the creation of yet another graphic file for AFP. The synthetic process is extremely quick, as one only performs minor geometrical operations on the residues, i.e. translations and rotations, a task which any medium-size personal computer can accomplish. The visualization of the synthesized protein, however, requires a workstation with graphics accelerated hardware, as the synthesized AFP was stored in a 44 Mb file.

Although we didn't encounter any major problems triangulating the non-terminal residues in the AFP molecule, we experienced a great deal of difficulty in producing satisfactory triangulations for the termini. Such difficulty was founded on the lack of homogeneity in the distribution of the points expanding the iso-density envelope in certain regions of these fragments. The end result was that cavities developed in the triangulation procedure, thus misrepresenting the van der Waals surface of the fragment. Rather than halting the thread of our work in order to redesign the afflicted algorithms we opted not to include the $ASP^{1} - THR^{2}|$ and $ALA^{36} - ARG^{37} - (NH_2)$ moieties in the present work. In fact, as it is the case with AFP, one does not need to add these moieties in the synthesis. Since it is hypothesized that the biological activity of this protein takes place on the interior of the protein, modeling its terminal groups would add little value to the study of this

protein's mode of action. Of course this is not a valid excuse for the non-refinement of the software used in the determination of the set P for a given fragment, a task which we will undertake in the near future, but it indirectly highlights the power of the theory of AIM in modeling any molecular subsystem; it removes the need to put "caps" on residues, or sequences of these, in order to model such subsystem within the framework of quantum mechanics. This is the direct consequence of the transferability of the residues in the AFP synthesized sequence. Because of this, one could continue linking other residues to the N-Surface of ALA^3 and/or to the C-Surface of THR^{35}, and continue adding groups to the strand indefinitely. In the limit that the $|Aa|$ exhibited near-perfect transferability, and it is our belief that this limit would be attainable if larger molds were to be used, this process would only be compromised by the error in the numerical computation of the properties and form of the individual residues.

Figures 5.9, 5.10 and 5.11 present different views of the synthesized AFP strand. To ease in the identification of the $|Aa|$ these have been color-coded according to the following scheme:

ALA	Green
SER	Orange
ASP	Grey
LEU	Olive Green
ASN	Blue
THR	Red
LYS	Purple
GLU	Pale Green

The first of these figures shows specifically how much of the AFP has been synthesized within the AIM formalism (as indicated in the preceding section the termini of this protein were not triangulated to our satisfaction). The closed-surface objects depict the various $|Aa|$ in AFP in sequential order from ALA^3 to THR^{35}. One immediately notices how well the resides fit together to form the protein strand - the direct consequence of the complementarity of the $|Aa|$ N-terminus and C-terminus amidic surfaces - producing virtually seamless joints. The

The first feature drawn from studying the composition of AFP is how abundant the alanyl residues are over any other, so much the case that the functionality of this family of proteins is thought to take place in the non-alanyl regions. This functionality is divided into two types, to bind to ice in specific lattice sites of the ice crystals, and to provide structural support to the protein. The ice-binding properties of the protein have been discovered to take place in a clustering of motifs occurring at a specific frequency along the strand. These motifs are found in AFP at THR^{2}/ASP^{6}, $THR^{13}/ASN^{16}/LEU^{12}$, $THR^{24}/ASN^{27}/LEU^{23}$ and THR^{35}. These clusterings can be seen clearly on figure 5.10. The arrangements of the motifs produce a near-perfect flat surface, recovering the structural features characterised by Yang and co-workers. In fact, the existence of "flat" surface areas expanded in terms of similar ice-binding motifs is now being used as the criterion for algorithms used in the identifications of other anti-freeze proteins (Yang *et al.* 1998). But the small number of possible hydrogen bonds made between AFP and the ice-surface does not explain how this protein would dock onto ice crystals, in particular when competing with an excess of water molecules in the serum of these fish. The consensus is that AFP binds onto the 2021 ice plane by establishing hydrogen-bond interactions with between the hydroxyl group in THR and oxygens in ice. Other evidence draws attention to the need for hydrophobic interactions as well between the THR residues to allow AFP to dock onto ice (Zhang and Laursen 1998).

Figure 5.11 presents AFP as viewed from the side opposite to the plane of the ice-binding motifs. This viewing angle highlights the motif formed by LYS_{18} and GLU_{22}. This motif can be considered a salt bridge and its functionality is that of preserving the helicity of the protein by reinforcing the strand at its middle part.

linked, as with the $|Aa|$, and that the hydrogen bond fragments be attached. However, Bader and Carroll performed a topological study of the hydrogen bond formed between a large number and variety of base molecules with the hydrogen in HF (Carroll and Bader 1988). They found, as is characteristic of such close shell interactions, that the value of the electron density at r_c, the location of the bond critical point, is identical to within 0.001 au to the sum of the contributions of the densities of the unperturbed B, the atom exhibiting the "base" behavior, and H atoms at the respective values of their bonded radii in the hydrogen bond. The additivity of the B and H electron densities at the critical point of the hydrogen bond demonstrates that such a bond results primarily from the interpenetration of the non-bonded densities on the H and B atoms with little accompanying polarization. Carroll and Bader showed that there is an almost complete lack of charge transfer between H and B (Carroll and Bader 1988). Thus, an acceptable model of the inter-residue hydrogen bonding interaction is obtained by simply adding the non-bonded electron densities, at the required separation, of the non-bonded H and O, provided they are properly aligned in the synthesized protein. A different study (Popelier and Bader 1994) determined the geometry of the atoms defining a typical α-helical hydrogen bond upon twisting a glycyl-based polypeptide. The calculated values are compared with the values predicted here by the linking of the $|Aa|$ residues, and summarized in figure 5.12.

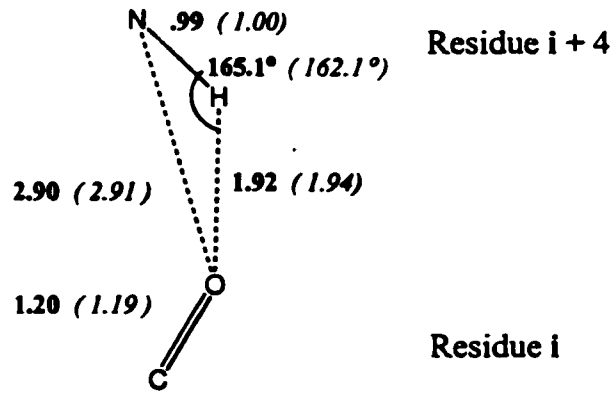


Figure 5.12: Comparison of the local geometry of the inter-residue hydrogen bond recovered by Popelier and Bader upon twisting a glycyl-based α-helical polypeptide (in bold font) and the corresponding values from the synthesized AFP strand (in italics). Bond distances are reported in angstroms, and the $N - H \cdots O$ angle in degrees.

The values obtained in the theoretical synthesis of the AFP strand are in excellent agreement with those reported by Popelier and Bader, demonstrating that the α-helical structure imposed on each

residue gives the geometry required for the formation of the hydrogen bonds, the very bonds that are responsible for stabilizing the α-helical secondary structure of proteins.

This reproducibility of the geometrical parameters demonstrates that the approximation of the hydrogen bond interaction in an α-helix in terms of the overlapping of the undisturbed non-bonded densities of H and O is an acceptable procedure.

It should be born in mind that we have not corrected for the additivity of the electron densities in our synthesis of AFP. This is because the computational implementation of the proposed idea is not trivial, albeit feasible, and we have not yet devised an efficient algorithm to tackle this problem. In practical terms, as it is in the modeling of docking problems, it seems that there is little to be gained by accounting for the hydrogen bond. Figure 5.13 shows a close-up view of where the a hydrogen bond between ALA^{7} and ALA^{11} should take place in the synthesized AFP strand. The effect of implementing the proposed correction of the electron density would have resulted in the slight smoothing of the ridge formed between O in ALA^{7} and H in ALA^{11}.

As already indicated, an important requirement for the use of the $|Aa|$, in the proposed synthetic procedure is that each fragment bears zero net charge. To the extent to which this is true, the linking of the fragments to form the polypeptide will result in charge conservation, and the resulting structure will bear no net charge. Indeed, using neutral residues, the net charge rendered in synthesizing the 33 residue strand in figure 5.3 is 0.087 e. This value is calculated by adding the charges of the residues tabulated in 4.32. This low value for the charge is a physically acceptable result, and a computational achievement, as 383 atomic integrations were required. Such amount is remarkably low and highlights that the building blocks of proteins *are* and *need to be* transferable in order to allow for the many combinations of residues possible found in proteins. It also confirms, if indirectly, the validity of the theory of AIM in recovering the exhibition of transfer of these pieces. It should be noted that we used the charges of the non-ionized arg instead of $arg(+)$, and asp instead of $asp(-)$ towards the calculation of the total charge of the synthesized strand. As already indicated, ionized residues are unlikely to exist unstabilized in an aqueous medium. This stabilization will be provided by hydration effects as well as ionic interactions with solutes. Even then, as the work on section 4.4 has shown, when stabilized by a counter-ion, the contribution to the charge of the strand is close to zero.

Because the charge of the AFP strand was zero (within computational errors), the direct consequence of the transferability of the residues, one can make sound predictions for any property

of the strand. These predictive values are obtained by adding the $|Aa|$ contributions to the corresponding properties. For example, the estimated total energy of the strand amounts to 9688.31 au or 6.08 $\times 10^6$ kcal/mol. Similarly, the distribution of the Laplacian of the electron density for the strand will be the added contribution of the linked pieces. In the ensuing chapter we discuss the link between the topology of the Laplacian of ρ and molecular recognition and reactivity. Among the atomic properties evaluated are the atomic multipole moments. Their use enables one to determine the electrostatic field in the neighborhood of the peptide chain that is exerted by its charge distribution. A particular study (Bader *et al.* 1992*a*) has revealed that if the atomic moments are transferable for an amino acid reside, then these moments allow for the calculation of electrostatic potential fields which are virtually identical to those computed as defined by Tomassi *et al.* (Bonaccorsi *et al.* 1970). The electrostatic potential is an important and widely used tool in problems of molecular recognition and protein docking. We present a more detailed description for this methodology in the ensuing chapter.

Geometrical parameters can be estimated as well. Just like Chang and Bader predicted the amidic bond lengths on small polypeptides (Chang and Bader 1992), we can do the same to the α-helical proteins. To achieve this, one need only take the average of the r_B column in table 4.17, the distance between the bond critical point which gives origin to the N-Surface and N, and add it to the average for the column under the heading r_A in table 4.29 showing the distance between C and the bond critical point at the C-Surface. This sum equals to 2.564 au, or 1.357 Å . The predicted bond length is slightly shorter that the mean of the $N_0 - C$ and $C - N_1$ bond length averages shown on table 4.2, by 0.001 Å. As compared to the experimental values, the predicted amidic bond lengths are slightly longer than the amidic bond lengths reported in Pauling's study (figure 4.10), by 0.0369 Å.

Errors incurred in the synthesis of the AFP strand can be quantified in terms of the net volume of non-overlap produced upon superimposing the complementary amidic surfaces in the $|Aa|$, as described in section 4.5. The error is easily calculated by adding the volume contributions shown on table 4.93 for the corresponding C-Surface|N-Surface overlaps. In the synthesized AFP strand, this volume amounts to 412.53 au^3, or 2.02% of the total volume occupying the 33 residues (the individual $|Aa|$ volumes are presented in table 4.32). Because the match of the amidic surfaces averages to approximately 13.34 au^3, the propagation of the error in the synthesis is linearly related to the number of amidic surfaces matched considering them, as it currently stands, in the absence of side-bonding like hydrogen bonding, disulfide bridges, salt bridges, etc.

Of course, the synthesis described in this chapter may seem rigid in the modeling of macro-systems in that does not permit for corrections once the pieces have been assembled together, even though the pieces may not exhibit as much transferability as it would be hoped for. We should bear in mind that this thesis presents an initial survey of the issues arising from implementing the proposed theoretical synthesis procedure. After realizing that the $|Aa|$ made for transferable pieces, and the feasibility in their use in the modeling proteins, we realized that the progression of this work would ensue upon taking the next most obvious step; devising theoretical means of correcting for the miss-matches in the surfaces upon making amidic bonds, and to reflect these corrections on the properties of the $|Aa|$ used in the synthesis. This is a problem currently under investigation in our laboratory.

Chapter 6

Reactivity Models and Theoretical Proteins

Computer-based molecular modeling tools have been within the chemist's reach for most of the last two decades. These tools have evolved in complexity as advances in computer hardware have taken place. Medium size workstations can perform at the Giga-FLOP level. Such performance has permitted the tackling of difficult numerical problems as well as the real time handling of highly demanding computer graphics.

One of the disciplines in pharmaceutical science which has adopted the use of computers towards enhancing its research has been the field of *Rational Drug Design*. This field generally models large biological systems with the intent of gaining chemical insights to be used in the formulation and prediction of biological activity. As a subset of Rational Drug Design, the field of *Protein Docking* is of particular computational and mathematical interest as it focuses on sampling a large number of substrates from databases, computes the energetics of interaction between these substrates and an active site of a protein (after sampling various geometrical configurations), and tabulates a hierarchy of candidate substrates in terms of their scoring performance. Unfortunately, the criteria used in the scoring methodologies is limited in its scope as it uses molecular mechanics almost invariably. In addition, the graphical modeling of chemical systems is generally rendered via nuclei-centered intersecting spheres of unequal radii, depending on the atomic number of the nucleus they are centered at. These generalizations of intermolecular interactions and atomic shapes could be improved upon. It is our view that these improvements will be produced if one made use of the

electron density, an experimental observable.

The preceding chapters have proven that transferable functional groups comply with the following criteria:

1. They recover the contribution to the experimentally measured properties of their parent molecules.

2. Their properties are entirely determined by their distribution of electronic density; if a functional group is transferable then it contributes equally to the total properties of each one of the systems it is a part of.

3. Their contributions to the properties of the total system are additive and yield the average of this system's properties.

$$\langle M \rangle = \sum_{\Omega} M(\Omega) \quad , \tag{6.1}$$

where Ω references each of the atomic open systems.

4. Because they are transferable, they can be used as building blocks in the synthesis of macromolecules without incurring in the loss of chemical information.

This thesis has demonstrated that the synthesis of proteins from transferable amino acid residues is feasible. The immediate benefit of this are that very good depictions of the shape and form of the created macromolecules in terms of the contours of the electron density can be produced with a moderate computational cost. This is already an improvement over the commonly used CPK-based molecular visualization techniques. But the usefulness of the synthesis described in this thesis need not end there. In fact, because the fragments are defined and handled in terms of distributions of the electron density, one can further benefit from developing and deploying reactivity tools which operate on the density distributions of the newly created macrosystems. This chapter proposes some possible avenues for future research along this idea.

6.1 The *pharmacophore* hypothesis.

Experimental studies have shown that complementarity in the spatial distribution of the complex formed between a protein's active site and a specific substrate plays a key role in rational approaches for drug design (Rebek 1987). The successive phases thought to take place in the process of receptor/substrate interaction are thought to be:

- *Recognition:* The receptor identifies substrates (likely to exist in their minimal energy conformation) that show potential for binding.
- *Binding:* The substrate and receptor form an energetically stable and sterically allowed complex.
- *Specific Perturbation:* The receptor is perturbed and induces the biological response.

Of these, only the first stage can be modeled without incurring in excessive complexity. In order for a substrate to be recognized by a receptor, the former must exhibit specific structural and property features which complement those of the receptor. The set of such features is termed the *pharmacophore.* To date, most of the information gathered about a pharmacophore does not include the electron density, or manifestations of it, since its computation for large system is prohibitory. By using theoretical synthesis of the active site of a protein, or of substrates, "brute force" approaches to computing the quantum mechanical properties of a system can be avoided. These properties can in turn be used to extend a pharmacophore.

6.2 Molecular Similarity and Complementarity

Molecular recognition is generally understood in terms of two physical principles: *chemical similarity* and *molecular complementarity.* Similarity can be quantified by comparing selectedly shared chemical features among sets of molecules in a given study. Complementarity is quantified in terms of *scoring functions*, functions designed to measure the extent to which the ability of a substrate to fit geometrically into an active site, interact with it energetically, etc.

AIM quantifies similarity and complementarity in terms of global and local properties. Global properties encompass the averaging of atomic theorems averaged over an atomic basin. Local properties refer to the study of topological features of scalar and vector fields maps such as that of the electron density, its Laplacian, the virial. In fact, critical point data from the first two are generally used as descriptors or "finger prints" of a given molecule. Examples of this use have been provided in chapter 4, where we used bond critical point data to compare how "similar" the backbones of the various $|Aa|$ are in order to study how assess their transferability. Similarity is thus synonymous with transferability in this context. It remains to be shown how could one benefit from exploiting local topological information in the study of molecular complementarity.

Traditionally, the three-dimensional modeling of molecules is rendered by placing spheres of pre-set radii at the positions of the nuclei. By virtue of intersecting with one another, these

spheres end up producing a three-dimensional approximation of the shape and size of the molecule modeled. But by modeling atoms as overlapping spheres with fixed radii (Garezzotti 1983) one may be making crude approximations which could in turn lead to ill predictions of substrate-receptor interactions. Isodensity envelopes are likely to be better descriptors of molecular complementarity since the electron density is a physical observable which characterizes any system with shape and structure (Bader *et al.* 1967).

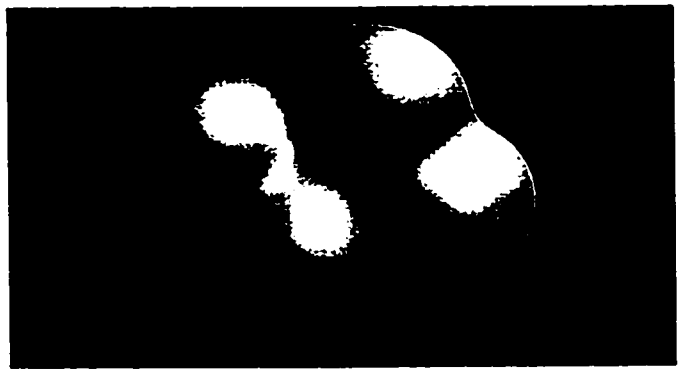


Figure 6.1: Glycine Iso-density Envelope ($\rho = 0.001$ au): The amino acid is positioned with the acid end to the right of the image, and the amide group at the left.

The 0.001 au density envelope yields molecular sizes that are in agreement with those determined from the analysis of kinetic theory data for gas phase atoms and molecules, whereas the smaller envelope of 0.002 au has been demonstrated to better recover atomic sizes for atoms in more closely packed environments (Bader *et al.* 1987, Bader and Preston 1970).

Figure 6.1 depicts 0.001 au isodensity envelope of Glycine. Figure 6.1 is to be compared with the CPK modeling of the same structure, produced with RASMOL (Bernstein 1999) and shown in figure 6.2. One notices the marked differences in the shape of the molecule. The use of spheres in figure 6.2 to denote atoms produces molecular shapes marked with sharp valleys, a function of the radii chosen to depict these atoms. These radii are generally not standardized as each software package will implement different values according to their criteria. In addition to the lack of agreement in the choice of the size of the atoms in CPK depictions, it commonly occurs that a particular atom type has no entry for its preferred radius in the database and a default value is assigned to it. For example, one can notice in figure 6.2 that all hydrogen atoms are rendered with spheres of equal radius, despite their being bonded to different atoms. Thus, the hydroxyl hydrogen depicted on the left of the image is equal in size and form as the CH_2 hydrogens depicted at the top

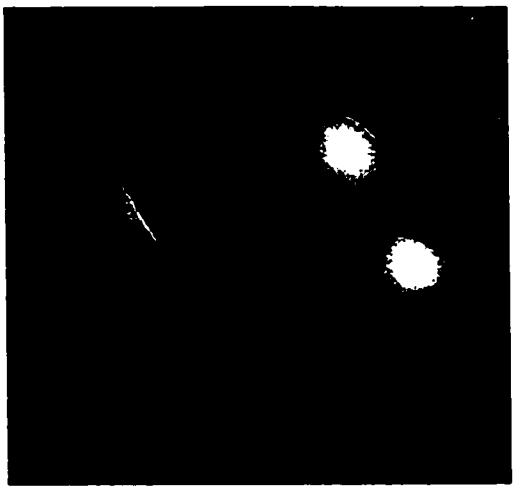


Figure 6.2: Glycine CPK rendering (to scale) at the same orientation as that of figure 6.1.

of the image. This is not the case when one models the same system using the electron density as the molecular shape descriptor, where each atom's size and shape is rendered by quantum mechanics.

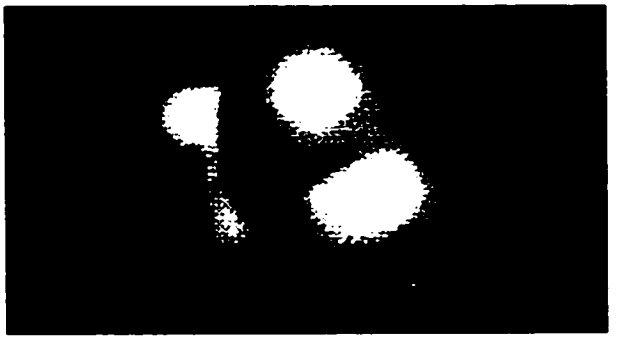


Figure 6.3: Glycine Iso-density Envelope: View of the CH_2 (top) and NH_2 (bottom) groups.

Because the theoretical synthesis has been proven feasible in the preceding chapter, one can now devise methodologies for the determination of structure complementarity to a protein's active site for the purposes of proposing pharmacophores. Using the *key* and *lock* analogy, this would entail the designing of algorithms which would analyze the local topology of the isodensity envelope of the protein's pocket, the lock, and from it propose different substrates, or the keys, to fit it. This process would be entirely driven by the structural information of the pocket, and would need to be complemented with reactivity information as well. Ideas in that regard are presented in the next

Figure 6.4: Glycine Iso-density Envelope: View of the NH_2 terminus.

section.

6.3 The Laplacian of the electron density.

The Laplacian of the electron density is a powerful tool for the prediction of reactivity of molecules. Like the electron density, it is a scalar field which topological features can be summarized in terms of the number and types of its critical points, i.e., the locations in real space where $\nabla\ \nabla^2\ \rho(r) = 0$. When $\nabla^2\rho(r) < 0$, the electronic density is concentrated at r, whereas when $\nabla^2\rho(r) > 0$, the electron density is locally depleted. Sites of electrophilic and nucleophilic attack in a molecule correlate with the sites of charge concentration and charge depletion, respectively. Local charge concentrations has been characterized as Lewis bases or a nucleophiles, whereas a local charge depletions corresponds to Lewis acids sites or electrophiles (Bader 1990). A chemical reaction occurs as sites of charge concentration interact with sites of charge depletion. For convenience, one classifies the topological features of the negative of the Laplacian function since it is more intuitive to associate positive values of a function with concentration and negative with depletion. Thus, the (3, +1) critical points in $-\ \nabla^2\rho$ correspond to regions where the electronic density is locally depleted. On the other hand, the (3, -3) critical points highlight regions of charge concentration which exhibiting "basicity" and are electrophilic. The pairing of (3, +1) regions of a substrate with (3, -3) regions of a receptor or vice versa result the modeling of chemical reactivity. This tool has been implemented successfully in several studies of which we provide two examples: the successful prediction of the protonation of the keto oxygen in formamide and substituted derivatives (Slee and Bader 1992), and the equally successful prediction and explanation for the Michael addition with acrylic acid and derivatives

(Carroll *et al.* 1989).

As demonstrated in section 2.8.1, the local implementation of the virial theorem yields the means of relating the total energy of an atom in a molecule as a function of the contributions of the kinetic energy and potential energies. This relationship is given as

$$2T(\Omega) = -\mathcal{V}(\Omega) \quad . \tag{6.2}$$

The variable T is the kinetic energy depicted in eqn.[2.87] and $\mathcal{V}$ the virial in eqn.[2.77]. It naturally follows that the electronic energy of a basin satisfies the identities:

$$\begin{aligned} E_e(\Omega) &= \int_\Omega E_e(r)d\tau \\ &= -T(\Omega) \\ &= \frac{1}{2}\mathcal{V}(\Omega) \quad . \end{aligned} \tag{6.3}$$

Since the potential energy is everywhere negative, as opposed to the kinetic which is everywhere positive, the sign of the Laplacian determines, via

$$\frac{\hbar^2}{4m}\nabla^2\rho(r) = 2G(r) = \mathcal{V}(r) \quad , \tag{6.4}$$

which of the contributors to the energy is predominant of the 2:1 equilibrium ratio. Thus, the energy of those regions where the Laplacian is *negative*, described as regions where the electronic density is locally concentrated, is dominated by contributions from the potential energy term. Conversely, those regions where the Laplacian is locally depleted, i.e. producing local *positive* values for the Laplacian, feel a greater contribution to their local energy from the kinetic energy term.

Figure 6.5: Glycine Iso-Laplacian Envelope: View of the reactive surface ($\nabla^2\rho = 0$). The molecule is oriented as in figure 6.1, with the acid end at the left of the image, the keto oxygen shown at the bottom of the image, and the amide group at the right.

The impact of the stated above, is that one can use the Laplacian as a predictive tool of Lewis acid-base reactions as the centering of a local molecular region of charge concentration with that of another molecule which exhibits charge depletion. This corresponds with producing an interaction as a region of low potential energy is matched with one of high kinetic energy, giving a result that satisfies the virial ratio.

As an illustration, figure 6.5 provides a view of the zero contour of the Laplacian, the so-called "reactive surface" for Glycine oriented in the same direction for the same molecule in figure 6.1. The "lumps" correspond to the areas of charge concentration which are particularly well formed at the keto oxygen. In contrast, the keto carbon is clearly electrophilic, as its corresponding region of space in figure 6.5 is depicted with a cavity, save for the core, shown as a suspended small sphere.

Much like maps of the electron density can be stored in a database and subsequently retrieved for the purposes of synthesizing macrosystems, Laplacian maps could be stored in a similar manner. When one is to examine the geometries at which a substrate could interact with a protein, one need only reconstruct the Laplacian map for the active site. In addition, algorithms could be designed to read the topological features of the Laplacian map of an active site and, translate this to specific complementary features which can be used towards more sensitive identifications of successful ligands.

6.4 Electrostatic Potential from Moment expansion.

The electrostatic potential is a real quantity which can be measured experimentally by X-ray diffraction (Doucet and Weber 1996). Its field can also be calculated as indicated in the important article by Tomassi *et al* (Bonaccorsi *et al.* 1970) who proposed it as a tool for the prediction of the reactivity between molecules as a function of the complementarity of their respective fields. The electrostatic potential is a scalar field which describes the energy of interaction between a molecular system and a positive test charge located at a reference point in space. The electrostatic potential is evaluated as

$$\phi(\mathbf{r}) = \sum_{i}^{N} \frac{Z_i}{|\mathbf{R}_i - \mathbf{r}|} - \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \quad , \tag{6.5}$$

where the summation term evaluates the nuclei contribution to the field, and the integral term that of the electron density. $\mathbf{R}_i$ references the location of the nuclei, and $\mathbf{r}'$ is the integration variable for the electron density over all space. When eqn.[6.5] is evaluated for a molecule at a desired isodensity envelope, one can then predict the likelihood of the interaction between positively charged region in

a substrate with regions where the electrostatic potential is strongly negative (electrophilic attack). The relative simplicity of this function, in conjunction with its long-range nature arising from its r^{-1} behavior, has made the analysis of the electrostatic potential an invaluable tool in the prediction and indexing the extent of reactivity in many substrate-receptor studies.

However, the caveat in using this reactivity index is readily drawn from examining eqn.[6.5]. The calculation of the potential at every point of the field requires an integration of the electron density over all space, a prohibitory venture when the chemical system of interest is large. In order to overcome this practical limitation we propose an alternative to computing the equation of Tomassi *et al.* This alternate procedure computes the field of the electrostatic potential in terms of atomic moments of the electron density defined within the AIM formalism (Messer 1993).

The multipole moment expansion method expresses the electrostatic potential as a summation of its moments as (Jackson 1962)

$$\phi(\mathbf{r}) = \sum_{i=0}^{\infty} \phi_i(\mathbf{r}) \quad , \tag{6.6}$$

where $\phi_0(r)$ relates to the first moment, the monopole, $\phi_1(r)$ to the dipole, and so on. Moment expansions centered at a single atom have been shown to converge slowly (Pack *et al.* 1972), requiring high order multipole moments in order to recover the potential field. However, the convergence is greatly improved if the expansion takes place over many centers.

The electrostatic potential maps are valid only at large distances from the molecular charge distribution. This is so because the positioning of a charged species in close proximity to a molecule would cause polarizations large enough to change the potential energy surface sampled to compute the wavefunction for such molecule. Thus, in the ensuing derivation we will assume that $\mathbf{r}$, the location of the test charge, will lie outside the isodensity contour $\rho(\mathbf{r}) = 0.002$ au. The encapsulated region of space bound by this isodensity contour contains 98-99% of the total charge of the system (Bader *et al.* 1971).

We proceed to derive the multipole expansion for $\phi(\mathbf{r})$ by expanding the $\frac{1}{|\mathbf{r}-\mathbf{r}'|}$ term in eqn.[6.5]. If we translate the problem from Cartesian to polar coordinates, then (r, θ, ϕ) represents the position of the test charge and (r', θ', ϕ') represents the integration variable. Using Legendre polynomials of the form $P_n^m(\cos\theta)$

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} \frac{(n-m)!}{(n+m)!} \frac{r'^n}{r^{n+1}} P_n^m(\cos\theta') P_n^m(\cos\theta) e^{im(\phi'-\phi)} \quad . \tag{6.7}$$

Upon substitution of eqn.[6.7] into eqn.[6.5] one obtains the expression for the electrostatic potential

as

$$\phi(\mathbf{r}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} \frac{(n-m)!}{(n+m)!} q_n^m \frac{P_n^m(\cos\theta)\, e^{im(\phi'-\phi)}}{r^{n+1}} \quad , \tag{6.8}$$

where

$$q_n^m = \int \rho(r) r^n P_n^m \cos\theta' e^{im\phi'} dr \quad . \tag{6.9}$$

The set of q_n^m values can be considered as coefficients in the expansion of eqn.[6.8] and correspond only to the properties of the charge distribution. They are directly related to the multipole moments of the system allowing eqn.[6.8] to be reformulated as

$$\phi(\mathbf{r}) = \phi_0(\mathbf{r}) + \phi_1(\mathbf{r}) + \dots \quad , \tag{6.10}$$

where $\phi_0(\mathbf{r})$ refers to the monopole contribution to the potential, $\phi_1(\mathbf{r})$ the dipole, and so on. This result is of great importance for it indicates that if a suitable set of moments are computed for a functional group in a molecule, and if these moments are transferable, a library of such moments would be be of great help as the field of the electrostatic potential could be easily calculated for the synthesized systems. As already indicated in chapter 2, AIM provides the means for computing the moments of the electron density distribution of an atom in a molecule. The monopole, or charge of an atom is computed according eqn.[2.86], the dipole eqn.[2.92] and quadrupole eqn.[2.94]. The explicit computation of the electrostatic potential for the one-center expansion in terms of the above defined moments becomes

$$\phi(\mathbf{r}) = \sum_{\Omega=1}^{N} \left[\frac{q(\Omega)}{r} + \frac{\mu(\Omega)\cdot\mathbf{r}}{r^3} + \frac{\sum_i \sum_j Q_{x_i x_j}(\Omega) x_i x_j}{2r^5} + \dots \right] \quad , \tag{6.11}$$

where the nuclear charges have taken into account within the charge term $q(\Omega)$.

In the past the modeling of the molecular force field was frequently restricted to a set of atom based point charges (monopoles). These consisted of charges derived by fitting them to an empirical force field that reproduces the crystal geometry (Hagler *et al.* 1977), of Mulliken atomic charges obtained from SCF calculations (Clementi *et al.* 1987, Sordo *et al.* 1987), of overlap-normalized partial charges obtained from CNDO/2 calculations (Momany *et al.* 1975) or of charges derived by fitting of an independently derived potential (Weimer *et al.* 1986). An interatomic transfer of electronic charge results in the 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 multipole moments. In one approach, cumulative atomic multipole moments (CAMMS) are derived from the corresponding set of molecular dipole 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 of Stone (Stone 1981, Stone 1985) in the determination of the multipole representations of the charge densities of blocked residues of the naturally occurring amino acids. These authors used standard force field bond lengths and angles to fix the molecular geometries.

With the exception of a few cases (Cooper and Stutchbury 1985, Popelier 1996), all of the previously defined charges and multipole moments have been expressed in terms of spatially overlapping contributions such as atomic centered basis functions, as opposed to the disjoint partitioning of space used to define the atomic multipole moments as in equations [2.86, 2.92, 2.94]. Stone's DMA can be implemented using moments defined for spatially disjoint regions or in terms of a set of many-centered basis functions, but Stone and Alderton (Stone 1985) argue that the latter procedure should in general produce faster convergence ti the true potential since atomic boundaries are generally non-spherical. This conclusion is, however, not borne out in practice as shown by Cooper and Stutchbury (Cooper and Stutchbury 1985) in the modeling of van der Waals complexes and more recently Popelier's (Popelier 1996) direct comparison of the electrostatic potential for a number of molecules using the multipole moments defined within the AIM framework. Popelier concludes that the atomic multipoles obtained from the theory of AIM reproduce the exact *ab initio* electrostatic potential to the required accuracy.

As important as is the ability of the 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.

The results presented in this thesis represent the first set of moments, complete to include quadrupoles, of all the amino acid residues which make *alpha*-helical proteins or portions thereof, allowing for the computation of any electrostatic potential fields of these systems in a quick and accurate manner.

Appendix A

Units

Table A.1: Units Conversion Table

Length	1 *au*	=	a_0	=	0.529177Å
				=	5.29177×10^{-11} m
Charge	1 *au*	=	e	=	1.60219×10^{-19} C
Electronic Density	1 *au*	=	e/a_0^3	=	6.7483 eÅ^{-3}
Laplacian Density	1 *au*	=	e/a_0^5	=	24.099 eÅ^{-5}
Dipole Moment	1 *au*	=	ea_0	=	2.542 Debyes
Energy	1 *au*	=	e^2/a_0	=	627.51 kcal/mol
				=	2.6255×10^3 kJ/mol

Appendix B

Post-Examination Corrections

In compliance with the defense committee requests we present in this appendix a comparison of some of the geometrical and topological properties of the $|Aa|$ with recently published crystallographic data.

Lecomte and co-workers have recently presented a study (Benabicha *et al.* 2000) in which glycyl-L-threonine (GlyThr) was crystallized from water by solvent evaporation. The experimental electron density distribution for this molecule was derived using a single-crystal X-ray diffraction data technique, at 100 K, and to a resolution of $(\sin\theta/\lambda) = 1.2$ Å^{-1}. The values obtained are presented in table B.1 and they are to be compared with those reported in tables 4.2 (bond lengths of the $|Aa|$) and tables 4.17 to 4.30 inclusive (bond critical point data of the $|Aa|$) which are included in table B.1 in *italic* font.

Table B.1: Experimentally derived bond critical point properties for a small polypeptide. Bond distance in Å, the rest of the values are in *au* (Benabicha *et al*, 2000). The corresponding values produced with the glycine mold are reported in *italics*

Bond	d(Å)	r_A	r_B	ρ_c	$\nabla^2\rho_c$
$C_0 - N$	1.339 *1.34*	1.06 *.870*	1.47 *1.69*	.352 *.327*	-.929 *-.819*
$N - H$	1.031 *.997*	1.44 *1.42*	.504 *.467*	.293 *.355*	-1.09 *-1.99*
$N - C_\alpha$	1.471 *1.44*	1.60 *1.82*	1.18 *.920*	.239 *.263*	-.282 *-.586*
$C_\alpha - H$	1.086 *1.08*	1.38 *1.32*	.673 *.722*	.256 *.299*	-.705 *-1.13*
$C_\alpha - C$	1.516 *1.53*	1.41 *1.38*	1.46 *1.51*	.256 *.267*	-.597 *-.769*
$C = O$	1.232 *1.20*	.930 *.756*	1.40 *1.51*	.405 *.426*	-1.16 *.037*

The bond lengths in table B.1 are in good agreement with Pauling's idealized α-helical geometry depicted in figure 4.10, with the exception of $C_\alpha - C$, which is 0.015 Å shorter in Lecomte's study. Our reported value for the same bond is in closer agreement with Pauling than with Lecomte. This is expected because our dihedral angles are the same as Pauling's and different than Lecomte's.

Lecomte's bond critical point data is also in reasonable agreement with ours. The small variations in ρ_b may be explained by the numerical inaccuracies introduced upon refining the crystallographic data. The experimental values of $\nabla^2\rho_b$ also incur errors by implementing the numerical differences analysis from the densities derived above. The largest discrepancy between our reported values and those of Lecomte is found for the $N - H$ values, where ρ_b is 0.062 au larger in the $|Aa|$. The rest of the values are in closer agreement.

References

Akkiraju, N., Edelsbrunner H., Facello M. P. Fu, E. P. Mücke and C. Varela (1995). Alpha shapes: Definition and software. In: *Proceedings of the 1st International Computational Geometry Software Workshop.* pp. 63–66.

Arfken, G. (1985). *Mathematical Methods for Physicists, 3rd. Ed.*. Academic Press.

Bader, R.F.W. (1990). *Atoms In Molecules: A Quantum Theory.* Oxford University Press.

Bader, R.F.W. and C. Matta (n.d.). Polarization density. Unpublished Results.

Bader, R.F.W. and F. Martín (1998). Interdeterminancy of basin and surface properties of an open system. *Can. J. Chem.* **76**, 284–291.

Bader, R.F.W. and M.J.T. Preston (1970). Determination of the Charge Distribution of Methane by a Method of Density Constraints. *The. Chim. Acta* **17**, 384–395.

Bader, R.F.W. and P.L.A. Popelier (1993). Atomic Theorems. *Int. J. Quant. Chem.* **45**, 189–207.

Bader, R.F.W., M.T. Caroll, J.R. Cheeseman and C. Chang (1987). Properties of Atoms in Molecules: Atomic Volumes. *J. Am. Chem. Soc.* **109**, 7968–7979.

Bader, R.F.W., P.L.A. Popelier and C. Chang (1992*a*). Similarity and Complementarity in Chemistry. *Theo. J. Mol. Struc.* **87**, 145–171.

Bader, R.F.W., P.M. Bedall and P.E. Cade (1971). Partitioning and Characterization of Molecular Charge Distributions. *J. Am. Chem. Soc.* **93**, 3095–3107.

Bader, R.F.W., T.A. Keith, K.M. Gough and K.E. Laidig (1992*b*). Properties of atoms in molecules: additivity and transferability of group polarizabilities. *Mol. Phys.* **75**(5), 1167–1189.

Bader, R.F.W., W.H. Henneker and P.E. Cade (1967). Molecular Charge Distribution and Chemical Binding. *J. Chem. Phys.* **46**, 3341–3363.

Barret, G.C. (1985). *Chemistry and Biochemistry of the Amino Acid.* Chapman and Hall.

Benabicha, F., V. Pichon-Pesme, C. Jelsch, Lecomte C. and A. Khmou (2000). Experimental charge density and electrostatic potential of glycyl-l-threonine dihydrate. *Acta Cryst. B* **56**, 155–165.

Benson, S.W., F.R. Cruichshank, D.M. Golden, G.R. Haugen, H.E. O'Neal, R.S. Rodgers and R. Walsh (1969). Additivity rules for the estimation of thermodynamical properties. *Chem. Rev.* **69**, 279–324.

Bernstein, H. J. (1999). *RASMOL 2.7.1: Molecular Graphics Visualization Tool.* Glaxo Wellcome Research and Development. Based on version 2.6 of R. Sayle.

Bonaccorsi, R., E. Scrocco and J. Tomassi (1970). Molecular SCF Calculations for the Ground State of Some Three-Membered Ring Molecules: $(CH_2)_3$, $(CH_2)_2NH$, $(CH_2)_2NH_2^+$, $(CH_2)_2O$, CH_2s, $(CH_2)_2CH_2$ and n_2ch_2. *J. Chem. Phys.* **52**, 5270–5284.

Carroll, M.T.C. and R.F.W. Bader (1988). An analysis of the hydrogen bond in BASE-HF complexes using the theory of atoms in molecules. *Mol. Phys.* **65**, 695–722.

Carroll, M.T.C., J.R. Cheeseman, R. Osman and H. Weinstein (1989). Nucleophilic addition to activated double bonds: prediction of reactivity from the laplacian of the electronic density. *J. Phys. Chem.* **93**, 5120–5123.

Chang, C. and R.F.W. Bader (1992). Theoretical Construction of a Polypeptide. *J. Phys. Chem.* **96**, 1654–1662.

Clementi, E., F. Cavallone and R. Scordamaglia (1987). Analytical potential from *Ab Initio* computations for the interaction between biomolecules. *J. Am. Chem. Soc.* **99**, 5531–5545.

Cooper, D.L. and N.C.J. Stutchbury (1985). Distributed muliple 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 Densitites and Chemical Bonding.* Oxford Science Publications.

Doucet, J.P. and J. Weber (1996). *Computer-Aided Molecular Design: Theory and Applications.* Academis Press.

Edelsbrunner, H. and E. P. Mücke (1994). Three-Dimensional Alpha Shapes. *ACM Transactions on Graphics* **13**, 43–72.

Feynman, R.P. (1964). *The Feynman lectures on Physics, Vol. 2, Lecture 19.*. Addison-Wesley, Reading , Mass.

Frisch, M.J., G.W. Trucks, M. Head-Gordon, P.M.W. Gill, M.W. Wong, J.B. Foresman, B.G. Johnson, H.B. Schlegel, M.A. Robb, E.S. Replogle, R. Gomperts, J.L. Andres, K. Raghavachari, Binkley J.S., C. Gonzalez, R.L. Martin, D.J. Fox, D.J. Defrees, J. Baker, J.J.P. Stewart and J.A. Pople (1994). *Gaussian 94, Revision E.* Gaussian Inc.

Garezzotti, A. (1983). The Calculation of Molecular Volumes and the use of Volume Analysis in the Investigation of Structured Media and of Solid-State Organic Reactivity. *J. Am. Chem. Soc.* **105**, 5220–5225.

Gelfand, I. M. and S.V. Fomin (2000). *Calculus Of Variations.* Dover.

Hagler, A.T., E. Huler and S. Lifson (1977). Energy functions for peptides and proteins. I. Derivation of a consisten force field including the hydrogen bond from amide crystals. *J. Am. Chem. Soc.* **96**, 5319–5327.

Hansen, N.K. and P. Coppens (1978). Testing aspherical atom refinement on small molecules data sets. *Acta Cryst.* **34**, 6242–6250.

Head-Gordon, T., M. Head-Gordon, M. Frisch, C. L. Brooks III and J. Pople (1991). Theoretical studies of blocked glycine and alanine peptide analogues. *J. Am. Chem. Soc.* **113**, 5989–5997.

Huang, L., L. Massa and J. Karle (1996). Kernel projector matrices for Leu1-zervamicin.. *Int. J. Quantum Chem., Quantum Chem. Symp.* **30**, 479–488.

Jackson, J.D. (1962). *Classical Electrodynamics.* John Wiley & Sons.

Jelsch, C., V. Pichon-Pesme, C. Lecomte and A. Aubry (1998). Transferability of multipole charge-density parameters: application to very high resolution olygopeptide and protein structures. *Acta Cryst.* **54**, 1306–1318.

Levine, Ira N. (1970). *Quantum Chemistry, 2nd. Ed.*. Allyn and Bacon.

Massa, L., L. Huang and J. Karle (1995). Quantum christallography and the use of kernel projector matrices. *Int. J. Quantum Chem., Quantum Chem. Symp.* **29**, 371–384.

Matta, C. and R.F.W. Bader (n.d.). Free amino acid parameters. Unpublished Results.

Messer, R.R. (1993). The Virial Partitioning Theory and Its Applications to Molecular Systems. PhD thesis. McMaster University, Hamilton, Canada.

Momany, F.A., R.F. McGuire, A.W. Burgess and H.A. Scheraga (1975). Energy parameters in polypeptides. VII. geometric parameters, partial atomic charges, non-bonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occuring amino acids. *J. Phys. Chem.* **79**, 2361–2381.

Murray, R.K., D.K Granner, P.A. Mayes and V.W. Rodwell (2000). *Harper's Biochemistry. 25th. Ed.* Appelton and Lange.

Pack, G.R., H. Wang and R. Rein (1972). A Quantitative Demonstration of the Domain Of Convergence Of Multipole Representations Of Molecular Potentials. *Chem. Phys. Lett.* **17**, 381–384.

Pauling, L., R. B. Corey and H. R. Branson (1951). The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Ntl. Ac. Sc. U.S.A.* **37**, 205–211.

Pichon-Pesme, V., C. Lecomte and H. Lachekar (1995). On building a databank of transferable experimental density parameters: applications to polypeptides. *J. Phys. Chem.* **99**, 6242–6250.

Popelier, P. (1996). Integration of atoms in molecules: a critical examination. *Molec. Phys.* **87**, 1169–1187.

Popelier, P.L.A. and R.F.W. Bader (1994). Effect of Twisting a Polypeptide on Its Geometry and Electron Distribution. *J. Phys. Chem.* **98**, 4473–4481.

Price, S.L., C.H. Faerman and C.W. Murray (1991). Toward accurate transferable electrostatic models for polypeptides: a distributed multipole study of blocked amino acid residue charge distributions. *J. Comp. Chem.* **12**, 1187–1197.

Ramakrishnan, C and G. N. Ramachandran (1965). Stereochemical Criteria for Polypeptide and Protein Chain Conformations II. Allowed conformations for a pair of peptide units. *Biophys. J.* **5**, 909–933.

Rebek, R. Jr. (1987). Model Studies in Molecular Recognition. *Science* **235**, 1478–1483.

Rossini, F.D. (1946). Heats of formation and combustion of the normal Alkylcyclopentanes and cyclohexanes and the increment per CH_2 group for several homologous series of Hydrocarbons. *J. Res. Nat. Bur. Stand.* **37**, 51–56.

Schulz, G.E. and R.H. Schirmer (1979). *Principles of Protein Structure.* Springer-Verlag.

Schwinger, J. (1951). Theory of Quantized Fields (I). *Phys. Rev.* **82**, 914–927.

Sicheri, F. and D.S.C. Yang (1995). Ice-binding structure and mechanism of an antifreeze protein from winter flounder. *Nature* **375**, 427–431.

Slee, T.S. and R.F.W Bader (1992). Properties of Atoms in Molecules: Protonation at Carbonyl Oxygen. *J. Mol. Struct. (Theochem)* **255**, 173–188.

Smith, D.R. (1974). *Variational Methods in Optimization.* Dover.

Sokalski, W.A. and R.A. Poirier (1983). Cumulative atomic multipole representation of the molecular charge distribution and its basis set dependence. *Chem. Phys. Lett.* **98**, 86–92.

Sokalski, W.A., K. Maruszewski, P.C. Hariharan and J.J. Kauffman (1989). Library of cumulative atomic multipole moments. II. neutral and charged amino acid residues.. *Int. J. Quant. Chem. Biol. Symp.* **16**, 119–164.

Sordo, J.A., M. Probst, G. Corongiu and E. Clementi (1987). *Ab Initio* pair potentials for the interactions between aliphatic amino acids. *J. Am. Chem. Soc.* **109**, 1702–1708.

Srebrenik, S. and R.F.W. Bader (1975). Towards the development of the quantum mechanics of a subspace. *J. Chem. Phys.* **63**, 3945–3961.

Stone, A.J. (1981). Distributed multipole analysis, or how to describe a molecular charge distribution. *Chem. Phys. Lett.* **83**, 233–239.

Stone, A.J. (1985). Distributed multipole analysis methods and applications. *Molec. Phys.* **56**, 1047–1064.

Walker, P. D. and P. G. Mezey (1993). Molecular Electron Density Lego Approach to Molecule Building. *J. Am. Chem. Soc.* **115**, 12423–12430.

Weimer, S.J., P.A. Kollman, D.T. Nguyen and D.A. Case (1986). An all-atom force field for the simulation of proteins and nucleid acids. *J. Comp. Chem.* **7**, 230–252.

Yang, D.S.C., W-H. Hon, S. Bubanko, Y. Xue, J. Seetharaman, C. L. Hew and F. Sicheri (1998). Identification of the ice-binding surface on a type iii antifreeze protein with a "flatness function" algorithm. *Biophys. J.* **74**, 2142–2151.

Zhang, E. and R.A. Laursen (1998). Structure-Function Relationships in a Type I Antifreeze Polypeptide. *J. Biol. Chem.* **273**, –.