PROTEINS: Structure, Function, and Genetics 40:310–329 (2000)

An Atoms-In-Molecules Study of the Genetically-Encoded Amino Acids: I. Effects of Conformation and of Tautomerization on Geometric, Atomic, and Bond Properties

Chérif F. Matta and Richard F.W. Bader*
Chemistry Department, McMaster University, Hamilton, Ontario, Canada

ABSTRACT **The theory of Atoms-In-Molecules (AIM) is a partitioning of the real space of a molecule into disjoint atomic constituents as determined by the topology of the electron density, ρ(r). This theory identifies an atom in a molecule with a quantum mechanical open system and, consequently, all of the atom's properties are unambiguously defined. AIM recovers the basic empirical cornerstone of chemistry: that atoms and functional groups possess characteristic and additive properties that in many cases exhibit a remarkable transferability between different molecules. As a result, the theory enables the theoretical synthesis of a large molecule and the prediction of its properties by joining fragments that are predetermined as open systems. The present article is the first of a series (in preparation) that explore this possibility for polypeptides by determining the transferability of the building blocks: the amino acid residues. Transferability of group properties requires transferability of the electron density ρ(r), which in turn requires the transferability of the geometric parameters. This article demonstrates that these parameters are conformation-insensitive for a representative amino acid, leucine, and that the atomic and bond properties exhibit a corresponding transferability. The effects of hydrogen bonding are determined and a set of geometrical conditions for the occurrence of such bonding is identified. The effects of transforming neutral leucine into its zwitterionic form on its atomic and bond properties are shown to be localized primarily to the sites of ionization. Proteins 2000;40:310–329.** © 2000 Wiley-Liss, Inc.

Key words: Atoms-In-Molecules; AIM; amino acids; leucine; zwitter-ion; transferability; ab initio; hydrogen-bonding; atomic properties; bond properties

INTRODUCTION

Functional Groups and the Properties of Macromolecules

The single most important concept in chemistry is that of a functional group, the idea that a linked grouping of atoms can exhibit a set of characteristic geometrical and chemical properties. These properties enable one to detect the presence of a group in a molecule and to predict the effect its presence has on the molecule's chemistry. Nowhere is this concept more useful than in understanding and cataloguing the chemistry of biological macromolecules, which consist of polymeric combinations of a small number of building blocks: amino acids, nucleotides, simple sugars, and phosphates. We know from experiment that the geometrical parameters, bond lengths and bond angles, of the individual members of these sets and combinations of them can be assigned with some certainty when present in a macromolecule, but our ability to predict the chemical consequences of their presence is less precise. It is the purpose of this series of articles to demonstrate how the theory of atoms in molecules[1] can remedy this situation because of its unique provision of a definition of a functional group and its properties.

Theory identifies an atom in a molecule with a region of real space bounded by a surface that arises as a consequence of the dominant topological property of the electron density ρ(**r**), that is, it exhibits a local maximum at the position of a nucleus.[1] An atom in a molecule is therefore a bounded open system, one that is free to exchange electrons and energy with its neighbors. The region associated with an atom in a molecule is termed a proper open system because it and its properties are defined by the quantum action principle.[2,3] Atoms so defined are identified with the atoms of chemistry because i) their properties are characteristic and additive, summing to yield the corresponding values for the total system; ii) they are as transferable from one system to another as are their forms in real space; and iii) the surface that defines an atom in a molecule has the important property of maximizing the transferability of its form and hence its properties between molecules. Point ii) is a consequence of the truism that two identical pieces of matter exhibit identical properties, coupled with the realization that the form of matter is determined by its spatial distribution of charge. Thus to both understand and utilize the concept of a functional group through theory, the group must be defined in real space by its charge distribution, which is what physics does.

*Correspondence to: Richard F.W. Bader, Chemistry Department, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4M1. E-mail: bader@mcmail.cis.mcmaster.ca

Received 14 December 1999; Accepted 14 March 2000

The contributions of a group to the total properties of a molecule are accessible to experimental measurement in those instances where the properties of the group, as well as being additive, are transferable between molecules. It is these experimental group contributions to the volume, energy, polarizability, and magnetic susceptibility that are recovered by theory, providing the ultimate test of the quantum theory of Atoms-In-Molecules (AIM).[3] It is not generally appreciated just how high a degree of transferability the form of a group, and hence its properties, can exhibit. The methyl and methylene groups of saturated hydrocarbons played a historically important role in the development of the notion of a functional group, with early experiments providing evidence of the additivity of their volumes, polarizabilities, heats of formation, and magnetic susceptibilities. A theoretical calculation of the properties of such groups, defined as proper open systems at various computational levels with and without electron correlation, demonstrate that their charge distributions and hence their geometries are essentially superimposable through a homologous series of hydrocarbons, with their electron populations differing by less than 0.001e and their energies by less than 1 kcal/mole, the latter result being particularly remarkable considering that the total energy of a methylene group is of the order of 25,000 kcal/mole.

This series of articles reports the characteristic properties of the atoms and functional groupings of atoms found in the genetically-encoded amino acids and in their residues when bonded as the central member of a tripeptide. The tabulated atomic and group properties, including a set of atomic multipole moments, should facilitate the modeling of the behavior of a polypeptide and its intermolecular interactions. The atomic properties obtained in this study represent the set with the maximal degree of transferability. In the present article, the first in this series, leucine is used to study the dependence of the geometrical, bond, and atomic properties on molecular conformation and ionization-state. This is a necessary prelude to the second article (in preparation), which tabulates and compares the atomic and bond properties for each of the genetically-encoded amino acids in a stable conformation. The third article (in preparation) presents a parallel study for the properties of each amino acid (Aa) bound as the central residue in the tripepetide Gly-Aa-Gly.

Definition of an Atom in a Molecule

Only definitions essential to the present application of the theory are presented here. The region of space (Ω) dominated by the attractive force of the contained nucleus is bounded by an atomic surface through which there is no local flux in the gradient vectors of the density $\nabla\rho(\mathbf{r})$, a "zero-flux" surface, as illustrated in Figure 1 for atoms in the leucine molecule and defined in Eq. (1),

$$\nabla\rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = 0 \quad \text{for all points } \mathbf{r_s} \text{ on the surface } S(\mathbf{r_s}, \Omega), \quad (1)$$

where the vector $\mathbf{n}(\mathbf{r})$ denotes a unit vector normal to the atomic surface at the point $\mathbf{r_s}$. In words Eq. (1) states that the flux of the gradient vector field of the charge density, $\nabla\rho(\mathbf{r})$, is zero at every point on the surface. Thus as a result of the density that exhibits a local maximum at each nucleus, space is partitioned into a set of non-overlapping atomic basins where each basin is bounded by one or more interatomic surfaces. Inter-atomic surfaces are shown for most stable conformation of zwitter-ionic leucine in the plane of the N-Cα-COO nuclei.

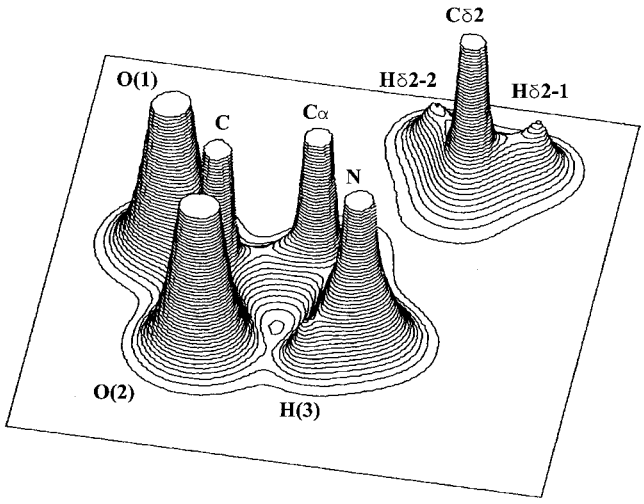
Definition of Bond Path and Properties

A trajectory of $\nabla\rho$ is the path generated by following the vector starting at some initial point, and a zero-flux surface is defined by a particular set of trajectories that satisfies Eq. (1). All of the members of this set terminate at a single point, a critical point in the density where $\nabla\rho(\mathbf{r})= 0$. There is one such critical point (cp) that lies between each pair of atoms that share a common interatomic surface. It is termed a bond critical point (bcp) since, in addition to the set of trajectories that terminate there and define an interatomic surface, it has associated with it a pair of trajectories originating at the cp, with each member of the pair terminating at one of the neighboring nuclei. This latter pair of trajectories defines the bond path, a line through space along which the density is a maximum with respect to any neighboring line and the atoms so linked are bonded to one another.[4] The collection of bond paths defines the molecular graph. Only atoms that share a common interatomic surface are linked by a bond path and are therefore bonded to one another. The ability to uniquely delineate the molecular graph for any density distribution is of particular importance in the case of weak interactions such as hydrogen bonding,[4] since the presence of a bond path unambiguously establish the presence or absence of bonding (see Fig. 1, for example).

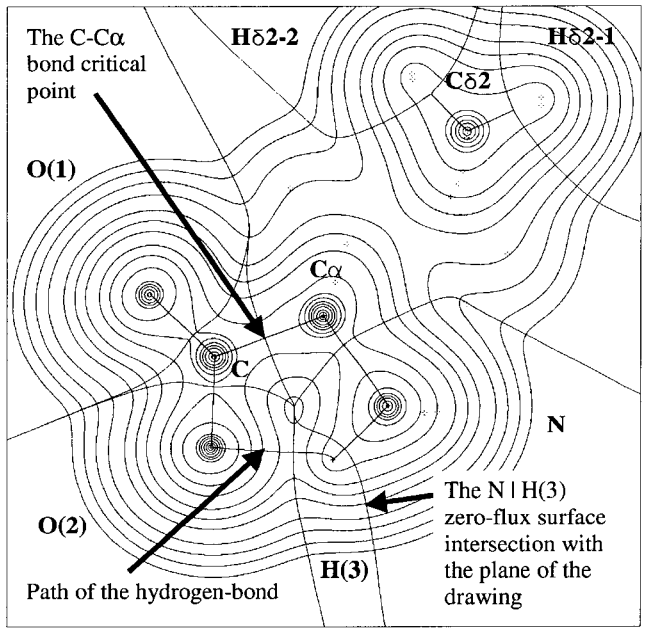
The interaction between a pair of bonded atoms is characterized by the properties exhibited by the density at the bcp:

a) The density at the bcp, ρ_b, provides a measure of bond order for the bonding between a given pair of atoms. In general, $\rho_b > 0.20$ atomic units (a.u.) for shared or polar interactions and < 0.10 a.u. for interactions between closed shell atoms, such as ionic and hydrogen bonded interactions. The distance of a bcp from nucleus A determines the bonded radius of atom A with respect to the interaction defined by the bcp and is denoted by $r_b(A)$. If the bond path is coincident with the internuclear axis, then the sum of the two associated bond radii, termed the bond path length, equals the bond length. If however, the bond path is curved, strained in the chemical sense, then the bond path length will exceed the bond length. Present examples of this latter behavior are found for hydrogen bonded interactions.

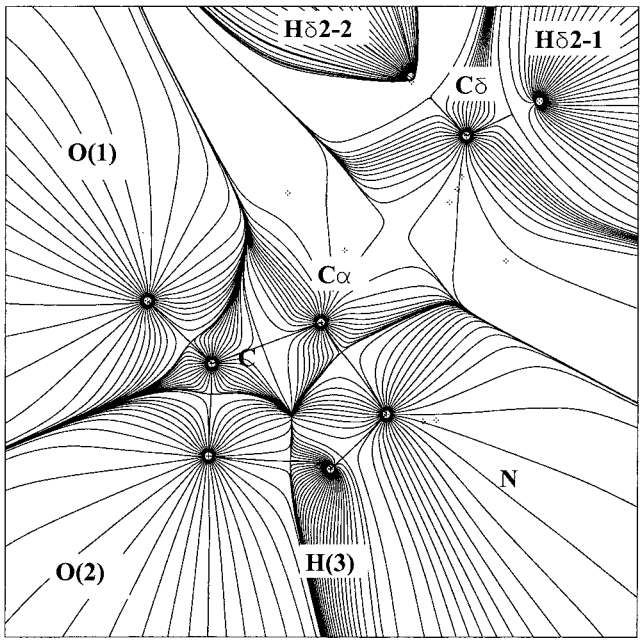
b) The Laplacian of the density at the bcp, $\nabla^2\rho_b$, is the sum of the three curvatures of the density at the critical point, the two perpendicular to the bond path, $\lambda1$ and $\lambda2$ being negative while the third, $\lambda3$, lying along the bond path, is positive. The negative curvatures measure the degree to which density is concentrated along the bond path, while the positive curvature measures the extent to



(a)



(b)



(c)

Fig. 1. Displays of the electron density $\rho(\mathbf{r})$ and its gradient vector field for a given plane of the leucine zwitter-ion **zA**. The plane contains the nuclei of the atoms comprising the hydrogen-bonded 5-membered ring consisting of -C(O)O. . .HNCα- , with H(3) bonded to O(2), one of the two oxygen atoms of the $-COO^-$ group. The nuclei of a carbon and two hydrogens belonging to a terminal methyl group of the side chain happen to be only slightly displaced from the same plane. **a:** A relief map of $\rho(\mathbf{r})$ plotted in intervals of 0.02 a.u. and truncated at 1.0 a.u. This figure illustrates the principal topological feature of the density: that it exhibits a maximum at the position of each nucleus. **b:** A contour display of $\rho(\mathbf{r})$ overlaid with the bond paths and the pairs of surface trajectories showing the intersections of the interatomic surfaces with the plane of the diagram. Note that the bond path for the hydrogen bonded interaction is bent in the region of the proton, a characteristic feature when the acid, hydrogen and base nuclei are not co-linear. The density (in atomic units) increases in value from the outermost 0.001 contour inwards in steps of $2\times10n$, $4\times10n$, and $8\times10n$ with n starting at -3 and increasing in steps of unity. **c:** A gradient vector field map of $\rho(\mathbf{r})$. Each line represents the path traced out by a trajectory of a vector $\nabla\rho(\mathbf{r})$. Since $\rho(\mathbf{r})$is a maximum at each nucleus and $\nabla\rho(\mathbf{r})$ points in the direction of maximum increase in $\rho(\mathbf{r})$, all paths in the vicinity of a given nucleus terminate at the nucleus. The region of space traversed by the trajectories that terminate at a given nucleus define its basin and an atom is the union of the nucleus and its basin. Basins of neighboring atoms are separated by trajectories that terminate at an intervening critical point (cp) where $\nabla\rho(\mathbf{r}) = 0$, called a bond cp. Only one such pair of trajectories is contained in a given symmetry plane. They are members of an infinite set of trajectories, all of which terminate at the bond cp and define an interatomic surface of zero-flux. A unique pair of tajectories originate at each bond cp and terminate one each at the neighboring nuclei. They define the bond path, the line of maximum density linking nuclei that share an interatomic surface and are bonded to one another. The methyl protons are sufficiently out of plane such that the bond paths linking them to the carbon are not displayed, except for the portions of these bond paths that fall in the carbon basin. The cp between Cα and the first atom of the side chain does lie out of this plane and consequently the associated interatomic surface is not indicated in the map.

which it is depleted in the region of the interatomic surface and concentrated in the individual atomic basins. In a shared interaction, density is accumulated between the nuclei and concentrated along the bond path so that ρ_b is large and $\nabla^2\rho_b < 0$, as exemplified by C-H for which $\rho_b = 0.29$ a.u. and $\nabla^2\rho_b = -1.1$ a.u. For a closed-shell interaction density is removed from the region of contact of the two atoms and hence ρ_b is small and $\nabla^2\rho_b > 0$, as

exemplified by the hydrogen bond formed between N-H and C=O for which $\rho_b = 0.01$ a.u. and $\nabla^2\rho_b = +0.03$ a.u. Polar interactions such as C=O, while characterized by significant charge accumulations between the nuclei typical of shared interactions, are dominated by charge transfer and the bcp falls in the region where the valence density borders the core of the electropositive atom. The Laplacian of the density rises steeply in this region undergoes a change in sign and $\nabla^2\rho_b$ can be of either sign for polar interactions.

c) The ellipticity measures the extent to which density is preferentially accumulated in a given plane containing the bond path. It is defined as $\epsilon = (\lambda_1/\lambda_2) - 1$ where λ_1 is the perpendicular curvature of greatest magnitude. The ellipticity provides a measure of double bond character with $\epsilon = 0.0$, 0.23, and 0.45, respectively, at a C-C bcp in ethane, benzene and ethene, respectively.

These properties of the density are now routinely used in the analysis and characterization of electron densities of crystal structures determined experimentally by X-ray diffraction methods, a field recently reviewed by Spackman.[6] Energy densities determined by the one-electron density matrix (as opposed to just the density, which is its diagonal element) are used to summarize the mechanics of the interaction. These quantities are introduced below in the section describing the definition of the electronic energy of an atom.

Atomic Properties

Eq. (1) serves as the boundary condition for the definition of a proper open system.[1,2] It is important to appreciate that since the atomic properties are defined by quantum mechanics, the same theorems that are obtained for the molecule, such as the virial theorem, apply to each of its constituent atoms. Every property is represented by a "dressed" real space density, one that replaces the value of the property for a single electron at some point in space with a corresponding density that describes its average interaction with all of the remaining particles in the molecule, the virial density being an example. In this way the contribution to property A from atom Ω is given by the integration of its corresponding density $\rho_A(\mathbf{r})$ over the atomic basin, as indicated in Eq. (2),

$$A(\Omega) = \int_\Omega \rho_A(\mathbf{r})\, d\mathbf{r}. \tag{2}$$

If for example, the property A is represented by the unit operator, one obtains $N(\Omega)$ the average number of electrons in atom Ω, its atomic population. It follows from Eq. (2) that the atomic properties are additive, that is, the value of any property A for the total system denoted by $\langle A \rangle$ is obtained by summing its atomic contributions as indicated in Eq. (3),

$$\langle A \rangle = \sum_\Omega A(\Omega). \tag{3}$$

As a consequence of the atomic statement of the virial theorem, the electronic energy of an atom, $E_e(\Omega)$, is defined in terms of the atomic average of the electronic kinetic energy, $T(\Omega)$. $T(\Omega)$ can be defined and readily calculated in terms of a kinetic energy density $G(\mathbf{r})$ that is everywhere positive:

$$T(\Omega) = \int_\Omega G(\mathbf{r})\, d\mathbf{r} = -E_e(\Omega). \tag{4}$$

The second identity in Eq. (4) follows from the atomic statement of the virial theorem. For an equilibrium geometry, the atomic sum of the $E_e(\Omega)$ equals E, the total energy of the molecule, which includes the nuclear-nuclear repulsion energy. Theory also defines a potential energy density $V(\mathbf{r})$, the virial field, that is everywhere negative and that integrates to the total potential energy of the molecule, the sum $G(\mathbf{r}) + V(\mathbf{r})$ that yields the energy density $E_e(\mathbf{r})$ that integrates to $E_e(\Omega)$. The magnitude of the virial field exhibits the same topology as does the electron density.[7] This observation has the important implication that every bond path linking a pair of nuclei is mirrored by a virial path, a line connecting the same nuclei along which the potential energy density is maximally attractive.

The local statement of the virial theorem in Eq. (5) ties the kinetic and potential energy densities to the topology of the density through the relation

$$(\hbar^2/4m)\nabla^2\rho(\mathbf{r}) = 2G(\mathbf{r}) + V(\mathbf{r}), \tag{5}$$

where $G(\mathbf{r}) > 0$ and $V(\mathbf{r}) < 0$. Thus interactions for which $\nabla^2\rho_b < 0$ are dominated by a local lowering of the potential energy, while those for which $\nabla^2\rho_b > 0$ are dominated by a local excess in the kinetic energy as measured by the two to one ratio required for the local satisfaction of the virial theorem, which for the total system states $2T = -V$. The energy density $E_e(\mathbf{r}) = G(\mathbf{r}) + V(\mathbf{r})$ evaluated at a bcp compares the kinetic and potential energies on an equal footing and is found to yield negative values for all interactions with significant sharing of the electron density, its magnitude reflecting the "covalent character" of the interaction.[8]

Atomic Multipole Moments

Modeling electrostatic forces plays an important role in molecular modeling of biochemical processes such as molecular recognition. The electrostatic potential can be represented to acceptable accuracy through a multipole moment expansion in terms of the atomic moments, defined by theory, by including terms up to the quadrupole moment.[9–11] The monopole, dipole, and quadrupole moments for all of the atoms of the free amino acids and of their residues in the tripeptide Gly-Aa-Gly will be reported in this series of articles. The monopole, or atomic charge, is defined in Eq. (6) as the difference between the nuclear charge Z_Ω and the atomic population $N(\Omega)$ obtained by setting $\rho_A(\mathbf{r})$ in Eq. (2) equal to the density $\rho(\mathbf{r})$,

$$q(\Omega) = Z_\Omega - N(\Omega). \tag{6}$$

The atomic dipole is a vector with three components defined in Eq. (7),

$$\boldsymbol{\mu}(\Omega) = -e \int_{\Omega} \mathbf{r}_{\Omega} \rho(\mathbf{r})\, d\mathbf{r}, \tag{7}$$

with origin for vector $\mathbf{r}_{\Omega}$ at the nucleus of atom Ω. An axial component of the quadrupole moment tensor is defined as in Eq. (8) for the z-axis,

$$Q_{zz}(\Omega) = -e \int_{\Omega} (3z_{\Omega}^2 - r_{\Omega}^2)\rho(\mathbf{r})\, d\mathbf{r}, \tag{8}$$

where again the origin is at the atomic nucleus. The sum of the diagonal elements is zero, i.e., $\mathbf{Q}$ is a traceless tensor when defined in this manner. A set of axes can always be determined such that off-diagonal components vanish. The magnitude of the quadrupole moment is defined[12] as

$$|\mathbf{Q}| = \sqrt{\frac{2}{3}(Q_{xx}^2 + Q_{yy}^2 + Q_{zz}^2)}, \tag{9}$$

where Q_{xx}, Q_{yy}, and Q_{zz} are the eigenvalues of the diagonalized quadrupole tensor.

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[13]or of Mulliken atomic charges obtained from SCF calculations[14] or of overlap normalized partial charges obtained from CNDO/2 calculations[15] or of charges derived by a fitting of an independently derived potential.[16] An interatomic transfer of electronic charge results in a polarization of the atomic densities in a direction counter to the charge transfer and the resulting multipolar field cannot be adequately represented by a set of monopoles. More recently, there has been a resurgence in methods used to represent the electrostatic field in terms of a set of multipole moments. In one approach, cumulative atomic multipole moments (CAMMs) are derived from the corresponding set of molecular multipole moments,[17] which were used to obtain libraries of amino acid atomic multipole moments.[18] Price et al.[19] use the distributed multipole analysis (DMA) of Stone[20,21] in the determination of the multipole representations of the charge densities of blocked residues of the naturally occurring amino acids. These authors use standard force field bond lengths and bond angles to fix the molecular geometries. The results reported in the present series of articles represent the first set of moments and other atomic properties determined at fully optimized geometries.

With the exception of a few cases,[9–11] 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 Eqs. (6) to (9). Stone's DMA[20] 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[21] argue that the latter procedure should in general produce faster convergence to the true potential since atomic boundaries are generally non-spherical. This conclusion is, however, not borne out in practice as shown in the work of Cooper and Stutchbury[10] in the modeling of van der Waals complexes and more recently in Popelier's[11] direct comparison of the electrostatic potential for a number of molecules using the multipole moments obtained from the disjoint partitioning in AIM and a basis set representation of DMA. Popelier concludes that the atomic multipoles obtained from the theory of Atoms-In-Molecules 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. An even more stringent test of the stability of AIM moments to the choice of basis set is provided by their use in the calculation of distributed polarizabilities both static and frequency dependent, at the coupled perturbed Hartree-Fock level of theory.[22] These moments are necessary in the determination of intermolecular interaction potentials derived from perturbation theory. The authors[22] found that basis set partitioning schemes yielded polarizability parameters so unstable with respect to basis set extension as to be unusable. They demonstrated two important advantages that result from the use of the disjoint atoms defined in the theory of Atoms-In-Molecules for the determination of the induced dipole and charge flow contributions to the atomic polarizabilities: a) the atomic contributions to the moments maintain a remarkable stability with respect to basis set extension, exhibiting a stability equal to that exhibited by the total polarizability components and b), in the case of the n-alkanes, the atomic contributions yield methyl and methylene group polarizability parameters that are transferable between any members of series.

It is readily shown that a quantum mechanical definition of any property for an open system requires that it be bounded in real space and that the failure of the previous methods to meet this theoretical requirement is their principal shortcoming.[23] Any partitioning of a molecule into spatially overlapping atoms eschews the very property of the density that forms the core of the concept of an atom in a molecule, namely, that an atom exhibits characteristic properties in spite of changes in its neighbors, a property that is faithfully recovered by the atoms of theory. However, in the case of contributions defined by overlapping atoms or atomic basis functions, the density assigned to atom A when bonded to B contains contributions from the basin of B, while when bonded to different atom C, it contains contributions from the basin of atom C. Thus the density distribution assigned to atom A is markedly different in each of the two cases and A is assigned different properties even though its total density can be almost identical in the two cases. Models of overlapping atoms or overlapping atomic contributions, as used in the definition of Mulliken charges, do not account

for the essential observation that atoms and functional groups can exhibit characteristic properties in spite of changes in their environments. Mulliken charges can appear transferable (but still basis set dependent) between molecules in those cases where the environments of the atoms remain unchanged, a situation encountered in the repeating units of biological molecules. However, even in these instances the charges are strongly dependent on conformation.

The properties of the Cα atom in an amino acid exhibit a remarkable transferability as the side-chain is changed through the 20 amino acids (data to be published in the second article). The average charge, magnitude of the dipole, and the atomic energy of this atom are 0.581 (0.015), 0.499 (0.023), and −37.497 (0.014) a.u., respectively, values in parentheses being standard deviations (std). Other atomic properties exhibit similar transferability, as evidenced by the small std, and will be discussed in fuller detail in the following article. The atomic charges assigned on the basis of Mulliken population or DMM do not exhibit the same degree of transferability. For example, Price et al. report Mulliken net charges (q) for the CαH group for 14 residues blocked by CH_3CO- and -$NHCH_3$ and calculated using a standard set of geometries. A comparison of these charges with those obtained using AIM at the optimized geometries (values in brackets) show that the latter are more transferable: av. q = 0.079e (0.577e), std = 0.044e (0.023e), max(q) − min(q) = 0.169e (0.087e).

Transferability and the Synthesis of Large Molecules From Fragments

Computer resource requirements of self-consistent-field (SCF) calculations scale as to the fourth power of the contracted basis set, which limits the size of molecules that can be handled. Since large polarized basis sets are necessary if realistic densities are to be obtained, accurate densities can be directly calculated for only relatively small-size molecules at the present. This situation excludes a multitude of biologically important molecules from the reach of high-level SCF calculations. Fortunately, however, complex biological molecules usually consist of a small number of simple building blocks or fragments. It is an empirical fact that these building blocks retain their chemical, physical, and spectroscopic properties in large molecules and, hence, are recognizable regardless of their neighbors. Since large molecules exhibit the characteristic properties of their constituent fragments, it is enticing to mimic nature theoretically, i.e., to construct the density of a large molecule in a given geometry in terms of density distributions (experimental or theoretical) of properly tailored fragments. If this is feasible for a small peptide, then, in principle, the method applies to a protein of a much larger size. The size of the computational problem is reduced from that of the fourth power of a large number to that of the sum of the fourth powers of small numbers. A number of approaches that attempt to take advantage of the building-block principle have been described in the literature. All of these use a fragment partitioning based upon the expansion of the density in terms of atomic centered basis functions. Only two of these are described here, a brief review of these methods being given in Massa et al.[24]

Massa, Huang, and Karle[24] 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.[24]

Another group[25] is in the process of building a data bank of transferable multipolar electron density parameters obtained from a fitting of the experimental X-ray structure factors. This is accomplished using the MOLLY program,[26] which expands the density in terms of a set of overlapping atom-centered basis functions. This group has shown that similar atoms in different environments, such as the atoms of the four different peptide groups of Leu-enkephalin, have experimental multipolar density parameters that are statistically equal. In other words, within this modeling of the electron density, similar atoms in different environment are shown to be transferable. These researchers seek to exploit this transferability to enable the calculation of electrostatic potentials of large proteins once a data bank of these parameters for various types of atoms that occur in proteins becomes available.[25]

Within the theory of Atoms-In-Molecules, the synthesis of a large molecule is accomplished by the linking of fragments by matching their zero-flux surfaces and there is no question of overlapping basis functions from neighboring fragments. In the synthesis of a protein for example, this involves a matching of the amidic surfaces that bound and define each residue. Figure 2 is an example of an amino acid residue, the serinyl residue |HNCαH(CH_2OH)C(=O)|, defined by its two amidic surfaces of zero-flux as denoted by the vertical bars. This residue is obtained by its cleavage from an accurately determined density distribution for the tripeptide "mold" Gly|Ser|Gly. The third article in this series reports the properties of all the genetically-encoded amino acid residues |Aa| determined in the Gly|Aa|Gly mold. The success of this method is determined by the extent to which the amidic surfaces are insensitive to a change in the nature of the side chain defining Aa. It has been previously demonstrated that the properties of di- and tri-peptides can be successfully predicted by linking of

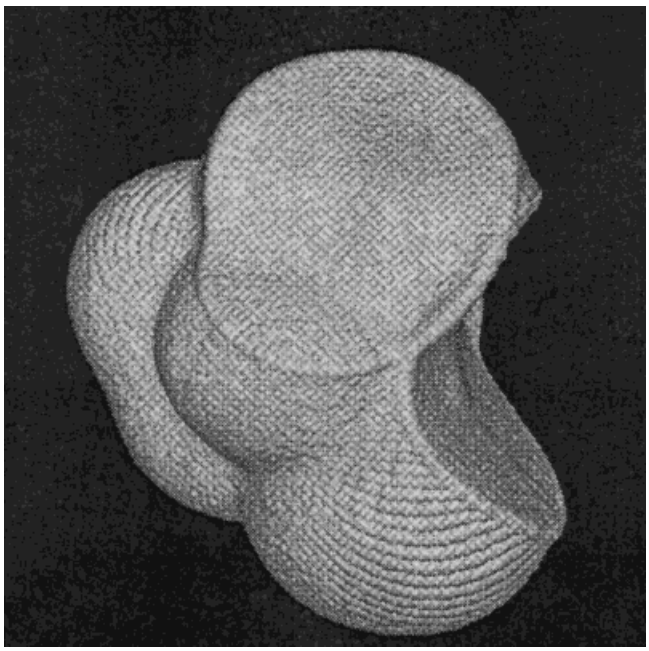


Fig. 2. The density of a serinyl residue cut at the two amidic zero-flux surfaces from a Gly|Ser|Gly mold. The fragment is represented by the intersection of the 0.001 a.u. density envelope, the van der Walls surface, with the -C(=O)| interatomic surface on the bottom right and the |NH- surface in the center at the top.

such fragments.[27] It is a consequence of the underlying theory that the distribution of charge of an open system and hence its properties change only in response to changes in its bounding surfaces. Therefore, a matching of the amidic surfaces on transfer implies an equal degree of transferability of the residue and its properties. For this to be possible, the two amidic surfaces bounding a residue in a polypeptide must be the opposite sides of one and the same surface. This in turn requires that any flux in the electric field through one surface for a residue without a formal charge must be the opposite of that through the other. In other words, the sum of the atomic charges on the residue must be zero (or equal to the net charge borne by the group) if it is to be transferable, a condition that is closely met. This and other surface constraints on transfer have been previously discussed.[28] The third article will demonstrate the feasibility of this method of construction and determine its limitations.

COMPUTATIONAL METHODS

In this article, the effects of conformation of the side chain on the total energy, geometrical parameters, bond, and atomic properties of non-zwitter-ionic leucine are studied. [Henceforth non-zwitter-ionic amino acids will be referred to as "neutral" amino acids (to be distinguished from amino acids with non-ionized side chains).] The changes in these properties incurred on the formation of the zwitter-ion are also investigated. In the second article, the genetically-encoded amino acids are studied in their neutral form. The neutral forms are chosen to eliminate the effects of charge separation on Cα since the environment of this atom in a polypeptide does not involve formal charge separation.

The Conformational Problem

Amino acids are generally floppy molecules that can fall into numerous stable conformations as they explore their respective potential energy hyper-surfaces (PES). Each well on a PES corresponds to the most stable local geometry, i.e., the geometry of one particular rotamer at which the forces on the nuclei vanish. Such a stable rotamer is referred to as a local minimum. The lowest energy local minimum is known as the "global minimum." There is no theoretical method at the present that guarantees locating the global minimum except for the smallest molecules. The global minimum is located usually by following paths of steepest descent (geometry optimization) after generating different starting rotamers either by i) random statistical sampling of the geometrical parameter space or ii) systematically rotating groups around single bonds to get different staggered conformers. These methods of generating initial conformers scale rapidly with the number of rotatable bonds, for example, rotations around four saturated carbon-to-carbon bonds would generate $3^4 = 81$ starting staggered conformers.

The rotamers of the side chains of the amino acids have been repeatedly been shown to be separated by low energy barriers on the PES. In an exhaustive conformational study at the MP2/6-31+G*//HF/6-31G* level, Gronert and O'Hair obtained ten energy minima for alanine, 51 for serine, and 42 for cysteine, all of which are relatively small amino acids.[29] More recently, the same group found 26 conformations of valine at the same level of theory.[30] In both studies, the minima of each molecule are found to span an energy range of about 10 kcal/mol and the results indicate that many of these have energy near the global minimum.[29,30] Even in the case of glycine, the smallest amino acid, seven local minima on the potential energy surface were identified by using a 4-31G basis set spanning a range 10.7 kcal/mol wide.[31] This difference between the global minimum and the least stable conformer in glycine is 0.006% of the total energy of the global minimum reported in their work. The effect of conformation on the total SCF energy is small in the present work as well: the nine neutral rotamers span a range of 5.1 kcal/mol, while their zwitter-ionic counterparts span a range of 3.9 kcal/mol.

The small differences between the energies of rotamers of the side chains account for their floppiness and for their ability to populate several rotameric conformations even in the solid state. For example, the side chains of crystalline [Leu^5]-enkephalin are shown to co-exist in four different conformations while the peptide backbones adopt essentially the same conformation.[32] In protein crystallography, the electron density of some side chains cannot be determined presumably as a result of the numerous conformations that they adopt in the crystal.[33]

It is also important to note that the gas-phase global minimum does not necessarily correspond to the water-

solvated global minimum. For example, a Monte Carlo simulation of four conformations of zwitter-ionic glycine in the gas-phase and in aqueous environment has shown that the addition of the free energies of solvation of each conformer to its gas-phase total energy (to obtain the total energy of the corresponding solvated species) results in two different global minima for the gas and for the solvated states.[34]

It is concluded that the location of the global minimum is an expensive endeavor with no guarantees. In addition, the gas-phase energy ranking may be different from the more biologically relevant solvent-phase ranking. Although interesting for their own sake, thorough conformational studies would be non-essential to study the AIM bond and atomic properties if these can be shown not to be excessively dependent upon conformation, i.e., if they exhibit an acceptable degree of transferability across rotamers. In other words, the insensitivity of the total energy, bond distances and angles, and bond and atomic properties to conformation transforms the problem of locating the global minimum to one of locating a local minimum on the PES. It is the purpose of the current article to establish this insensitivity and to show that one can rely upon geometrical, bond, and atomic properties of a reasonable local minimum. This will be done through the systematic generation of multiple energy minima for leucine followed by a comparison of the geometrical, bond, and atomic properties of these rotamers.

Choice of Basis Set

Unconstrained geometry optimizations were performed at the single-determinant Hartree-Fock level using a 6-31+G* basis set whose suitability for such optimizations was demonstrated in a benchmark article by Head-Gordon et al.[35] The wavefunctions were then recalculated at these geometries by using a 6-311++G** basis set for the determination of the density and the bond and atomic properties, a similar strategy having been used in a study of hydrogen bonding in related molecules.[36] The diffuse functions present in the large basis set are necessary to describe the density of anionic forms of aspartate and glutamate and were applied to all amino acids in this study for consistency.[37] In addition, the inclusion of diffuse function along with polarization functions has been shown to provide a balanced basis set that allows a better description of the potential energy hypersurface of glycine.[31]

Generation of Leucine Rotamers and Wavefunctions

For a fixed conformation of the -CαH(NH_2)COOH group, the leucine side chain can exist in nine (3×3) staggered rotamers around the Cα-Cβ and Cβ-Cγ bonds. A starting geometry for leucine was obtained from the amino acid structures library of Hyperchem05.[38] This geometry was fully optimized at the Hartree-Fock level implemented in Gaussian94[39] using the 3-21G basis set first, followed by optimization using a 6-31+G* basis set. This optimized rotamer was used to generate the starting geometries of the remaining eight rotamers by varying the necessary dihedral angles. The new rotamers were optimized directly at the 6-31+G* level. A single point calculation was performed at each of these nine geometries to obtain 6-311++G**//6-31+G* wavefunctions and densities. The optimized structures of the nine neutral rotamers were then modified to the corresponding zwitter-ionic forms and optimized using a 6-31+G* basis set followed by a single-point calculation using a 6-311++G** basis set. Figure 3a depicts the geometries of the lowest energy rotamers of both neutral and zwitter-ionic leucine along with the numbering scheme, and Figure 3b shows the remaining eight optimized minimum energy structures of neutral leucine (their optimized zwitter-ionic counterparts are not shown as they adopt the same side-chain conformation).

Atoms-In-Molecules (AIM) Calculations

The AIMPAC[40] suite of programs was used to calculate the bond and atomic properties. The bond properties were determined from the calculated electron densities. Bond paths were followed to their nuclear termini to ensure correct assignment of the two contributing atoms in the case of hydrogen bonded interactions. The atomic properties where obtained by integrations over each atomic basin.

Quality of Atomic Integrations

Theory provides a stringent test of the numerical accuracy of the atomic integrations. The kinetic energy can be expressed locally in terms of a function that is everywhere positive G($\mathbf{r}$) or in terms of another that has both positive and negative contributions K($\mathbf{r}$). Both functions yield the same average kinetic energy when integrated over all space, but will yield different values when integrated over an arbitrary subspace, Ω. The difference in their average values over Ω reduces to an integral of the flux in $\nabla\rho(\mathbf{r})$ through the surface of Ω, that is, to the surface integral of the term appearing in Eq. (1), $\nabla\rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r})$. Since this term vanishes at every point on the surface of an atom, free or bound, both K($\mathbf{r}$) and G($\mathbf{r}$) should yield identical values for the atomic kinetic energies when integrated over an atomic basin, just as they do when integrated over the complete molecule. The difference between the integrated values of K($\mathbf{r}$) and G($\mathbf{r}$) for an atom, L(Ω), should vanish for a properly defined atomic surface that satisfies Eq. (1). Deviations in L(Ω) from zero provide a measure of the numerical accuracy of the atomic integrations.

According to Eq. (3), the atomic sum of each property should equal the molecular value. Deviations from this equality beyond an acceptable numerical accuracy can be traced to the atom or atoms that exhibit elevated $|L(\Omega)|$ values. All integration reported in this work have $|L(\Omega)| < 5.0 \times 10^{-3}$. When necessary, the robust PROMEGA integration algorithm was used to remain within this accuracy upper-bound. The average $|L(\Omega)|$ for all atoms integrated in the present article (264 atoms) is approx. 7.0×10^{-4}. At this accuracy, the sums of the atomic populations N(Ω) for the leucine rotamers (nine neutral and three zwitter-ionic) differ on average by 0.005e from the total

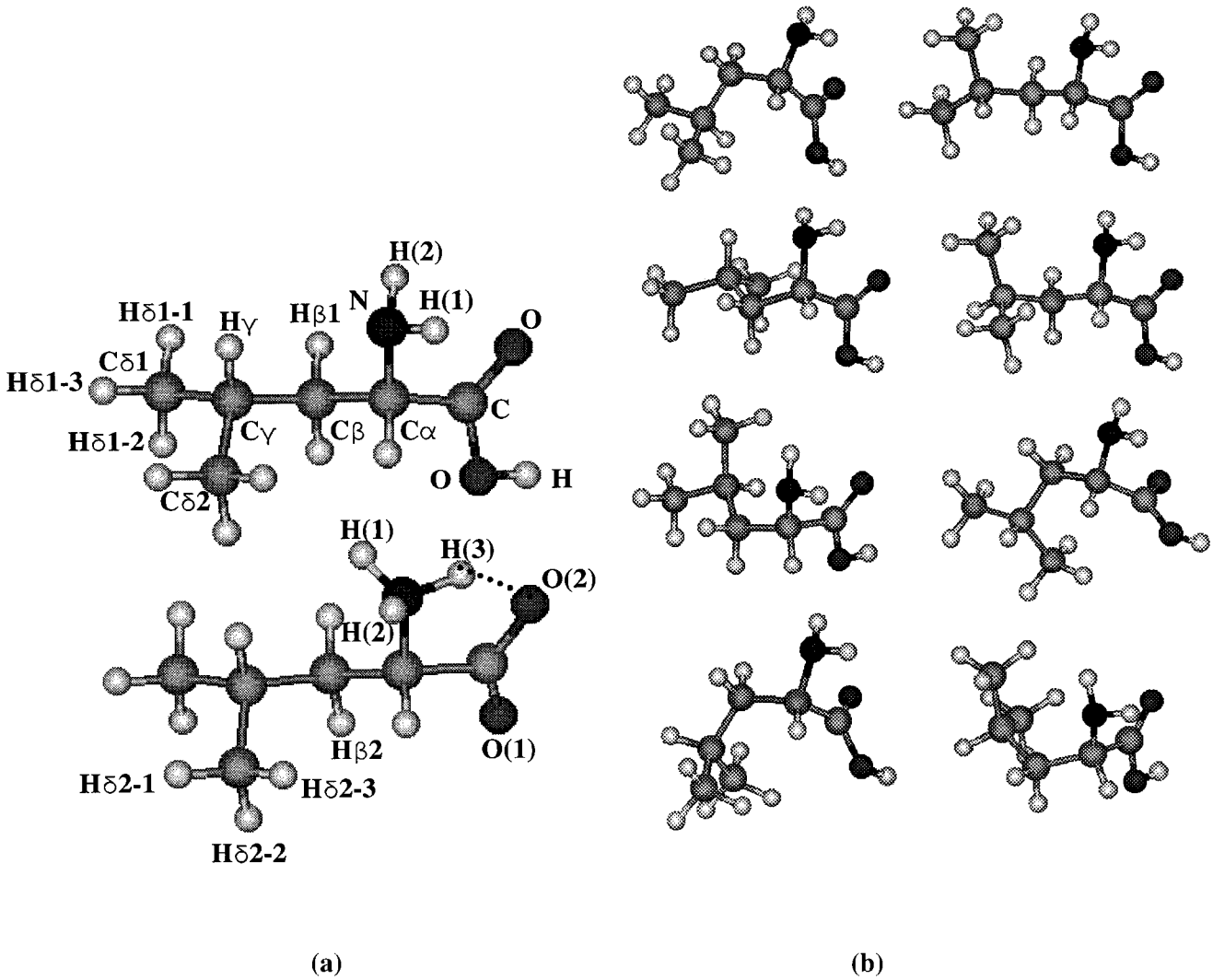


Fig. 3. **a:** Optimized geometries and numbering system for rotamers **nA** and **zA** (the global minima of the neutral and zwitter-ionic set). The numbering of the other rotamers follow the same conventions (summarized in the Appendix). Note the close geometrical similarity of the side chains of the two tautomers. **b:** Rotamers of the neutral (**n**)-set: **nB-nI** (ordered from left to right and then from top to bottom). Rotamers of the zwitter-ionic (**z**)-set are not shown, since the geometry of their side-chains are almost identical to the corresponding rotamers of the **n**-set.

number of electrons N (std ~ 0.003e). The sum of the atomic energies $E_e(\Omega)$ for each conformer differs from the SCF total energy on average by 0.7 kcal/mol (std ~ 0.6 kcal/mol).

RESULTS AND DISCUSSION

The effect of Conformation and Tautomerization on Energy and Geometrical Parameters

The geometry optimizations described in the present work provide nine local minima (rotamers) for the neutral set, designated as (**n**), and nine corresponding minima for the zwitter-ionic set, designated as (**z**). The rotamers are labeled **A** to **I** in order of increasing energy in the n-set. The relative energies and molecular dipole magnitudes of the rotamers are given in Table I and their respective torsion angles in Table II. The optimized **z**-tautomers exhibit side-chain geometries that are essentially the same as their **n** counterparts, justifying the use of the same label. The relative energy ranking is identical for each corresponding members in both sets, except for the interchange of the rotamers **H** and **I** by 0.1 kcal/mol in the **z**-set.

The nine minima of **n**-leucine span an energy range of 5 kcal/mol (Table I). The relative energies using both basis sets are almost identical. The enlargement of the basis set at the 6-31+G* optimized geometry lowers the energy by 75.4 kcal/mol. The **z**-rotamers span an energy range of 4 kcal/mol. The energy lowering due to basis set enlargement is about 73.3 kcal/mol in the **z**-set. The change in the total energy upon the migration of the proton from the neutral carboxylic group to the neutral amino group to form the zwitter-ion will be referred to as energy of tautomerization and is given by

$$E_\tau = E_{\mathbf{z}} - E_{\mathbf{n}}, \qquad (10)$$

TABLE I. Molecular Dipole Moments and Total Single Determinant SCF Energies of the Neutral (n) and Zwitter-Ionic (z) Leucine Rotamers at (a) 6-311++G//6-31+G*, (b) 6-31+G*//6-31+G*, and (c) 3-21G//3-21G Levels**

Leucine rotamer	Molecular $\|\mu\|$ (D) (a)–(**n**)	Rel. E (kcal/mol) (a)–(**n**)*	Rel. E (kcal/mol) (b)–(**n**)*	Rel. E (kcal/mol) (c)–(**n**)*	Molecular $\|\mu\|$ (D) (a)–(**z**)	Rel. E (kcal/mol) (a)–(**z**)*	Rel. E (kcal/mol) (b)–(**z**)*	Rel. E (kcal/mol) (c)–(**z**)*
A	1.279	0	0	0	10.521	0	0	0
B	1.581	0.7	1.1	1.3	9.923	0.7	0.7	0.3
C	1.410	1.9	1.9	1.3	10.442	0.8	0.6	1.4
D	1.390	2.0	2.1	1.7	10.114	1.3	1.2	0.7
E	1.412	2.5	2.5	2.0	10.470	1.5	1.5	2.2
F	1.483	2.7	2.8	2.7	10.007	1.6	1.6	2.0
G	1.369	3.1	3.1	3.5	10.014	3.1	2.8	0.6
H	1.416	3.9	3.9	4.0	10.057	3.8	3.8	2.7
I	1.520	5.1	5.1	4.9	10.018	3.9	3.6	4.1
***E (a.u.)**		−439.1015	−438.9815	−436.5280		−439.0599	−438.9429	−436.5271

TABLE II. Torsion Angles (°) of the Leucine Rotamers and the Corresponding Crystallographic Values†

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

†Crystal **B1** and **B2** refer to two closely related but crystallographically independent low-temperature X-ray structures[43] that correspond to the **B** rotamer in the present study. Enk. **A1–A3** refer to three crystallographically independent X-ray structures,[32] small letters refer to labeling in the original paper.

where $E_{\mathbf{n}}$ and $E_{\mathbf{z}}$ are the total energies of the **n** and the z forms of the amino acid with the same conformation. The average E_τ over the nine rotamers is +25.5(std=0.6) kcal/mol at the RHF/6-311++G**//6-31+G* level, showing that the **z**-form is less stable than the **n** counterpart in the gas-phase. This destabilization is more than compensated for by the energy of aqueous solvation estimated to be about 150 kcal/mol[41] for the solvated carboxyl and quaternized amino groups. The small value of the std indicate that E_τ is rather independent of the conformation of the side chain. For comparison, E_τ was calculated for glycine at the same level of theory and is found to be +28.3 kcal/mol, a value similar to that reported previously[42] (+29 kcal/mol) at the RHF/4-31G level. These results show that the replacement of one of the hydrogen atoms bonded to the Cα of glycine by the side-chain of leucine reduces E_τ by less than 3 kcal/mol, that is by less than 10%.

The side-chain conformations of rotamers **A** and **B** are in good agreement with x-ray structure of leucine[43] (see Table II). Leucine crystallizes in two slightly different symmetry independent conformations that we label **B1** and **B2** since both are close to the optimized rotamer B. On the other hand, three slightly different variants of conformation **A** are observed in the crystal of leu-enkephalin, a natural brain opiate pentapeptide.[32] The crystallographic side-chain dihedral angles differ on the average by 6° from their optimized counterparts (Table II), which is a relatively small difference that could be accounted for, at least in part, by crystal packing effects.

The calculated bond angles are also in good agreement with the X-ray values and the trends in the calculated bond lengths also agree with the X-ray results, but the X-ray bond lengths are generally longer than the calculated values that refer to the vibrational-less molecule (Tables III–VI). The average value for each of these geometrical parameters over the set of rotamers is given in these tables along with the std. The optimized bond lengths and angles show a marked constancy across rotamers of both **n** and **z** sets, std ranging from 0.000 to 0.005Å for bond lengths and from 0.0 to 2.5° for bond angles for both **n** and **z** sets. Comparison of the optimized geometrical parameters of the side chains of each **n** rotamer with its **z** counterpart shows that ionization has little effect and that the geometrical parameters are almost constant within each set and from one set to the other. This transferability of bond lengths and angles leads one to anticipate an underlying transferability of

TABLE III. Bond Lengths (Å) of the —CαH(NH_2)COOH Group in the Neutral (n) and Zwitter-Ionic (z) Leucine Rotamers and Corresponding Crystallographic Values (See Footnote of Table II)

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

[a]Bond length N—H(1,2) is an average of the averages of N—H(1) and N—H(2).

TABLE IV. Bond-Lengths (Å) of the Side-Chain of the Neutral (n) and Zwitter-Ionic (z) Leucine Rotamers and Corresponding Crystallographic Values (See Footnote of Table II)

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

[a]For Cβ—H and Cδ—H bonds, the values given are averages of averages.

TABLE V. Bond Angles (°) Involving the Terminal Cα(NH_2)COOH Groups of the Neutral (n) and Zwitter-Ionic (z) Leucine Rotamers and Corresponding Crystallographic Values (See Footnote of Table II)

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

[a]Averages of O(1)—C—Cα and O(2)—C—Cα.

TABLE VI. Bond Angles (°) of the Side Chains of the Neutral (n) and Zwitter-Ionic (z) Leucine Rotamers and Corresponding Crystallographic Values (See Footnote of Table II)

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

the electron density and hence of the atomic and bond properties, as illustrated below for the rotamers of leucine. The following articles (in preparation) illustrate that a similar transferability of geometries, bond, and atomic properties holds not only for the same side chain in different conformations but for corresponding bonds and atoms belonging to different genetically-encoded amino acids.

The Effect of Conformation on Bond Properties

The presence of a bond path joining the nuclei of a pair of atoms provides a rigorous criterion that the two atoms are bonded to one another, a criterion that is applicable to all types of bonding.[4] The properties of the electron and energy densities at the bcp, which serves as the origin of the trajectories of the gradient of the density that define

the bond path, characterize the bonding. With the exception of hydrogen bonds, the amino acids exhibit shared interactions with $0.21 < \rho_b < 0.44$ a.u., $\nabla^2\rho_b < 0$, and $E_e(\mathbf{r}_b) < 0$. The value of $\nabla^2\rho_b$ for the polar C=O bond in the ionized leucine rotamers is slightly positive, $E_e(\mathbf{r}_b)$ however remains negative. Bond properties are summarized in Table VII. Considerable accumulation of electronic charge between the nuclei distinguishes a shared interaction from one between two closed-shells, as found for hydrogen-bonded interactions. In such interactions the hydrogen atom is linked by bond paths to two other atoms and shares two interatomic surfaces (see Fig. 1b). The primary interaction, the acidic one, exhibits the characteristics of a shared or polar interaction while the weaker one between the hydrogen and the base atoms, exhibits the characteristics of a closed-shell interaction: $0.00 < \rho_b < 0.06$ a.u., positive $\nabla^2\rho_b$, and positive or near zero $E_e(\mathbf{r}_b)$ (Table VIII).

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

The bond properties are remarkably transferable from one rotamer to the other. In general, the fluctuations in these properties are too small across each rotamer set to exhibit physically meaningful trends. For this reason, only the average value of each property and its standard deviation are given in Table VII. The transferability of the bond length across rotamers is echoed by a parallel transferability in the bond properties. For example, the high transferability of O-H bond lengths in the **n** set (std=0.000) is associated with a high transferability in ρ_b and the remaining bond properties (see small stds in Table VII). In general, the transferability of one property entails the transferability of the others as well, since the various properties are inter-dependent.

The density in the neighborhood of the bcp for nearly all the bonds, with the exception of the hydrogen bonds (discussed separately below), is close to cylindrical symmetry, i.e., the bond ellipticities (ϵ) are close to zero. Two exceptions are the Cα-N(**z**) and C-O(**n**) interactions that have $\epsilon = 0.14$ and 0.16, respectively. With the exception of the hydrogen bonds, the bond paths lie along the internuclear axes and consequently the bond path lengths equal the corresponding bond lengths and there is no bond strain. Hydrogen bonding in these rotamers results in the formation of ring structures in that the bond paths exhibit deviations from linearity, as discussed below.

Hydrogen Bonding

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

All the ionic rotamers show at least one hydrogen bond between H(3) of the quaternized amino group and oxygen O(2). This hydrogen bond, shown in Figure 1, closes a 5-membered ring (5-MR) involving Cα, N, H(3), O(2), and C. It is the strongest such interaction found in this study in terms of its ρ_b value. Other weaker hydrogen bonds are also present, the properties of all are summarized in Table VIII. From this table one can identify distinct geometrical conditions for each type of hydrogen bond with length given as H to base (B) separation and the angle A-H-B where A denotes the acid to yield the following classification:

(1) Nine N-H. . .O=C in 5-MR with lengths 1.6–1.8 Å, angles 123–130° and $\rho_b \sim 0.046-0.059$ a.u.; this type is the strongest hydrogen bond found in this study, it exist in all zwitter-ionic rotamers, and is expected to occur in all zwitter-ionic amino acids.

(2) Six C-H. . .O=C in 6-and 7-MR with lengths 2.4–2.8 Å, angles 117–137° and $\rho_b \sim 0.006-0.011$ a.u.

(3) Two C-H. . .O(H) in 7-MR with lengths 2.7– 3.0 Å, angles 134–119° and $\rho_b \sim 0.007$ and 0.004 a.u., respectively.

(4) Four C-H. . .N in 6-MR with lengths 2.6– 2.8 Å, angles 104– 119° and $\rho_b \sim 0.007-0.011$ a.u.

(5) Three C-H. . .C in 6-MR with lengths 2.6– 2.7 Å, angles 108– 111° and $\rho_b \sim 0.007$ a.u.

(6) Four unusual hydrogen-to-hydrogen C-H. . .H-N bonds (i.e., hydrogen-bonds in which the acceptor atom is a negatively charged hydrogen atom): three in 7-MR with lengths 2.1–2.3 Å, angles 104–132° and $\rho_b \sim 0.009$ a.u.; and one in a 6-MR with length 2.1Å, angle 129° and $\rho_b = 0.010$ a.u. (Dihydrogen bonds involving a boron atom (B-H. . .H-N) were recently described in the literature).[5]

Carrol and Bader[44] reported a wide range of hydrogen bonds between various bases and HF, the strongest being H_3N. . .HF ($\rho_b \approx 0.034$), an intermediate being HF. . .HF ($\rho_b \approx 0.021$) and the weakest being HCl. . .HF ($\rho_b \approx 0.007$). Hydrogen bonds occurring between a methyl group and a negatively charged oxygen in creatine and carbamoyl sarcosine have ρ_b values of 0.035 and 0.033, respectively.[36] In open and cyclic formamide dimers, the values of ρ_b are 0.027 and 0.021 respectively.[45] Destro et al. report X-ray crystallographic ρ_b values for N-H. . .O bonds between different L-alanine molecules calculated from densities determined from a multipole refinement at 23K.[46] These workers found two inter-molecular O. . .H. . .N hydrogen bonds of length 1.83 Å and ρ_b 0.03 a.u. each, these two hydrogen-bonds bifurcate at oxygen O(2) (our numbering). They also found a shorter and stronger inter-molecular O. . .H. . .N hydrogen-bond of 1.72 Å and ρ_b of 0.04 a.u.[46] In comparison to these reported theoretical and experimental values of ρ_b, the hydrogen bonds found in this study fall in the range of

TABLE VII. Selected Bond Properties Averaged Over the Nine Neutral (n) and Nine Zwitter-Ionic (z) Leucine Rotamers, Respectively (Values Are in a.u. Except When Stated Otherwise)

		r_a(Å)	r_b(Å)	R(Å)	ρ_b	$\nabla^2\rho_b$	ϵ	G(**r**)	G(**r**)/ρ_b	E_e(**r**)
Cα—C										
n	av.	0.685	0.839	1.524	0.268	−0.790	0.086	0.059	0.218	−0.256
	std	0.001	0.001	0.001	0.001	0.005	0.027	0.000	0.002	0.001
z	av.	0.848	0.721	1.569	0.248	−0.670	0.067	0.049	0.200	−0.217
	std	0.001	0.002	0.003	0.001	0.008	0.004	0.000	0.000	0.002
Cα—Cβ										
n	av.	0.803	0.738	1.540	0.250	−0.660	0.027	0.051	0.206	−0.216
	std	0.002	0.003	0.004	0.002	0.012	0.005	0.001	0.001	0.004
z	av.	0.809	0.720	1.529	0.254	−0.683	0.026	0.055	0.215	−0.225
	std	0.002	0.003	0.004	0.002	0.009	0.009	0.001	0.002	0.003
Cα—H										
n	av.	0.686	0.398	1.084	0.294	−1.095	0.030	0.040	0.134	−0.313
	std	0.001	0.001	0.002	0.002	0.014	0.004	0.001	0.003	0.003
z	av.	0.689	0.392	1.080	0.296	−1.106	0.055	0.038	0.130	−0.315
	std	0.000	0.001	0.001	0.001	0.008	0.002	0.001	0.003	0.001
Cα—N										
n	av.	0.540	0.908	1.448	0.281	−0.900	0.012	0.139	0.494	−0.364
	std	0.004	0.003	0.005	0.004	0.024	0.003	0.003	0.008	0.009
z	av.	0.487	1.028	1.516	0.216	−0.208	0.142	0.233	1.078	−0.285
	std	0.001	0.002	0.001	0.001	0.015	0.006	0.003	0.017	0.002
Cβ—Cγ										
n	av.	0.784	0.757	1.540	0.250	−0.656	0.010	0.050	0.201	−0.214
	std	0.004	0.004	0.003	0.002	0.008	0.004	0.001	0.001	0.003
z	av.	0.793	0.749	1.541	0.249	−0.652	0.011	0.051	0.204	−0.214
	std	0.014	0.012	0.004	0.002	0.010	0.005	0.001	0.004	0.003
Cγ—Cδ1										
n	av.	0.764	0.769	1.533	0.252	−0.670	0.009	0.051	0.203	−0.219
	std	0.003	0.003	0.001	0.000	0.002	0.002	0.000	0.000	0.001
z	av.	0.766	0.767	1.533	0.252	−0.670	0.009	0.051	0.203	−0.219
	std	0.010	0.011	0.001	0.001	0.003	0.004	0.000	0.000	0.001
Cγ—Cδ2										
n	av.	0.763	0.769	1.533	0.252	−0.670	0.009	0.051	0.203	−0.219
	std	0.003	0.003	0.001	0.000	0.002	0.003	0.000	0.001	0.001
z	av.	0.767	0.766	1.533	0.252	−0.669	0.009	0.051	0.203	−0.219
	std	0.010	0.010	0.001	0.001	0.003	0.003	0.000	0.001	0.001
C—O										
n	av.	0.430	0.900	1.331	0.307	−0.095	0.162	0.438	1.428	−0.462
	std	0.000	0.000	0.000	0.000	0.012	0.001	0.002	0.008	0.001
C═O										
n	av.	0.397	0.794	1.190	0.434	0.131	0.059	0.792	1.824	−0.759
	std	0.000	0.000	0.000	0.000	0.012	0.001	0.002	0.005	0.002
C═O(1)										
z	av.	0.404	0.807	1.211	0.417	−0.080	0.054	0.701	1.680	−0.721
	std	0.000	0.001	0.001	0.001	0.004	0.002	0.003	0.004	0.002
C═O(2)										
z	av.	0.413	0.831	1.244	0.386	−0.194	0.004	0.598	1.548	−0.647
	std	0.000	0.001	0.001	0.001	0.003	0.003	0.002	0.002	0.002
Cβ—H(1)										
n	av.	0.680	0.406	1.086	0.286	−1.037	0.010	0.044	0.153	−0.303
	std	0.002	0.004	0.002	0.002	0.013	0.001	0.001	0.004	0.003
z	av.	0.675	0.416	1.090	0.282	−1.010	0.017	0.046	0.165	−0.299
	std	0.002	0.002	0.001	0.001	0.007	0.004	0.001	0.003	0.002
Cβ—H(2)										
n	av.	0.679	0.407	1.086	0.287	−1.041	0.010	0.044	0.154	−0.304
	std	0.001	0.001	0.001	0.001	0.006	0.001	0.000	0.001	0.002
z	av.	0.688	0.397	1.085	0.289	−1.055	0.009	0.041	0.142	−0.305
	std	0.006	0.008	0.003	0.003	0.019	0.002	0.002	0.008	0.003
Cγ—H										
n	av.	0.675	0.413	1.088	0.287	−1.044	0.003	0.045	0.158	−0.306
	std	0.002	0.003	0.001	0.002	0.009	0.001	0.001	0.004	0.002

TABLE VII. (Continued)

		r_a(Å)	r_b(Å)	R(Å)	ρ_b	$\nabla^2\rho_b$	ϵ	G(**r**)	G(**r**)/ρ_b	E_e(**r**)
z	av.	0.677	0.411	1.088	0.287	−1.047	0.005	0.045	0.156	−0.306
	std	0.007	0.011	0.003	0.004	0.027	0.002	0.003	0.013	0.004
Cδ1—H(2)										
n	av.	0.675	0.412	1.087	0.282	−1.012	0.008	0.046	0.162	−0.298
	std	0.002	0.002	0.001	0.001	0.006	0.001	0.001	0.002	0.002
z	av.	0.678	0.408	1.087	0.283	−1.018	0.009	0.045	0.158	−0.299
	std	0.005	0.008	0.003	0.003	0.021	0.001	0.002	0.009	0.003
Cδ1—H(3)										
n	av.	0.676	0.410	1.086	0.283	−1.018	0.008	0.045	0.160	−0.300
	std	0.001	0.001	0.001	0.001	0.004	0.001	0.000	0.001	0.001
z	av.	0.677	0.409	1.086	0.283	−1.020	0.008	0.045	0.158	−0.300
	std	0.002	0.003	0.001	0.001	0.006	0.002	0.001	0.004	0.001
Cδ2—H(1)										
n	av.	0.676	0.410	1.086	0.283	−1.017	0.007	0.045	0.160	−0.299
	std	0.001	0.001	0.000	0.000	0.002	0.001	0.000	0.001	0.000
z	av.	0.677	0.408	1.086	0.283	−1.021	0.008	0.045	0.157	−0.300
	std	0.002	0.002	0.001	0.001	0.004	0.002	0.001	0.003	0.000
Cδ2—H(2)										
n	av.	0.676	0.411	1.086	0.282	−1.012	0.008	0.045	0.159	−0.299
	std	0.003	0.003	0.002	0.001	0.008	0.001	0.001	0.003	0.003
z	av.	0.677	0.410	1.087	0.282	−1.014	0.009	0.045	0.160	−0.299
	std	0.005	0.007	0.002	0.003	0.019	0.001	0.002	0.009	0.003
Cδ2—H(3)										
n	av.	0.675	0.411	1.086	0.283	−1.019	0.008	0.045	0.160	−0.300
	std	0.002	0.002	0.001	0.001	0.005	0.001	0.001	0.002	0.001
z	av.	0.677	0.409	1.086	0.283	−1.023	0.009	0.045	0.158	−0.300
	std	0.005	0.007	0.002	0.003	0.017	0.001	0.002	0.008	0.003
N—H(1)										
n	av.	0.735	0.266	1.002	0.354	−1.769	0.052	0.059	0.168	−0.502
	std	0.000	0.001	0.000	0.000	0.007	0.001	0.000	0.001	0.002
z	av.	0.756	0.250	1.005	0.354	−1.992	0.010	0.045	0.127	−0.543
	std	0.001	0.001	0.001	0.001	0.004	0.000	0.000	0.001	0.001
N—H(2)										
n	av.	0.738	0.264	1.002	0.354	−1.800	0.050	0.058	0.164	−0.507
	std	0.001	0.001	0.001	0.001	0.016	0.001	0.000	0.001	0.003
z	av.	0.757	0.249	1.005	0.354	−2.004	0.010	0.045	0.126	−0.546
	std	0.001	0.001	0.001	0.001	0.016	0.001	0.000	0.001	0.004
N—H(3)										
z	av.	0.827	0.217	1.044	0.311	−2.010	0.005	0.039	0.125	−0.541
	std	0.005	0.000	0.005	0.005	0.039	0.000	0.001	0.005	0.009
O—H										
n	av.	0.782	0.171	0.953	0.375	−2.825	0.017	0.062	0.164	−0.768
	std	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000

weak interaction ($0.004 \leq \rho_b \leq 0.011$) with the exception of N-H. . .O=C bonds ($0.046 \leq \rho_b \leq 0.059$).

Popelier and Bader found similar hydrogen bonds between the same atoms (N-H. . .O=C), but in a non-ionic state, in the blocked triglycyl peptide HC(=O)|Gly|Gly|Gly|NH_2 using the same level of theory as in the current work.[47] These bonds are found to exhibit values of ρ_b ranging from 0.013 to 0.022 a.u. and all have positive Laplacians ($\nabla^2\rho_b$),[47] as is the case for the Laplacians reported in this work, as expected for such closed shell interactions. In this article, the authors also emphasize that the depletion of electron density on the hydrogen bonded hydrogen is an important contributor in the strength of a hydrogen bond.[47] The charge on the non-hydrogen-bonded H atoms on a quaternized ammonium group in **z**-leucine rotamers is around +0.4, but is +0.6 for the hydrogen bonded one (see Table IX). Arranging the hydrogen bonded hydrogens in order of their charge reproduces the ordering of the strength of the hydrogen-bonds. The charges on the H for the hydrogen-bonds found by Popelier and Bader[47] are 0.48 (weakest) and 0.50 (strongest), while it is 0.59 on the H involved in the strongest hydrogen-bond in the present work. Comparison of Figure 1c to Popelier and Bader's Figure 2 reveals a marked similarity of the shapes of the interatomic surfaces, the bond paths, and the gradient vector field of the density. In particular, the flatness of the hydrogen-bond interatomic surface is consistently found in such interactions (see, for example, Fig. 1 of Carroll and Bader[44] for an F-H. . .F-H bond and Fig. 1 of Cheeseman et al.[45] for N-H. . .O=C bonds).

In addition to stabilizing and holding certain conformations that might have otherwise been unfavorable, hydro-

TABLE VIII. Classification of the Hydrogen-Bonds Based on Selected Geometric and Bond Properties (Values are in a.u. Except When Stated Otherwise)

Hydrogen bond type		R (Å)	Angle°	Y—H—H—X°	ρ_b	$\nabla^2\rho_b$	ϵ	G(**r**)	E_e(**r**)
N—H · · · O(═C)									
zA	N—H(3) · · · O(2)	1.65	129		0.057	0.169	0.052	0.051	−0.009
zB	N—H(3) · · · O(2)	1.69	126		0.052	0.166	0.070	0.047	−0.006
zC	N—H(3) · · · O(2)	1.64	130		0.059	0.171	0.048	0.052	−0.010
zD	N—H(3) · · · O(2)	1.67	128		0.054	0.166	0.057	0.048	−0.007
zE	N—H(3) · · · O(2)	1.64	130		0.058	0.170	0.049	0.052	−0.009
zF	N—H(3) · · · O(2)	1.68	128		0.053	0.165	0.058	0.047	−0.006
zG	N—H(3) · · · O(2)	1.73	124		0.048	0.162	0.091	0.044	−0.003
zH	N—H(3) · · · O(2)	1.75	123		0.046	0.161	0.105	0.042	−0.002
zI	N—H(3) · · · O(2)	1.69	127		0.052	0.165	0.064	0.047	−0.005
Methyne C—H · · · O(═C)									
zB	Cγ—H · · · O(2)	2.54	117		0.010	0.034	0.211	0.008	0.001
Methyne C—H · · · H—N									
zD	Cγ—H · · · H(2)	2.07	129	−17	0.010	0.042	1.798	0.009	0.002
Methyl C—H · · · H—N									
nF	Cδ2—H(3) · · · H(2)	2.11	132	39	0.009	0.033	0.184	0.007	0.001
nI	Cδ1—H(2) · · · H(2)	2.15	130	23	0.009	0.033	0.387	0.007	0.001
zF	Cδ2—H(3) · · · H(2)	2.29	104	120	0.009	0.036	0.565	0.007	0.002
Methyl C—H · · · N									
nC	Cδ1—H(1) · · · N	2.55	119		0.011	0.035	0.150	0.008	0.001
nE	Cδ2—H(2) · · · N	2.61	113		0.010	0.033	0.198	0.007	0.001
zC	Cδ1—H(1) · · · N	2.72	108		0.007	0.030	0.245	0.006	0.001
zE	Cδ2—H(2) · · · N	2.78	104		0.007	0.029	0.405	0.006	0.001
Methyl C—H · · · O(H)									
nB	Cδ1—H(1) · · · O(H)	2.95	119		0.004	0.016	0.250	0.003	0.001
nH	Cδ1—H(2) · · · O(H)	2.72	134		0.007	0.024	2.201	0.005	0.001
Methyl C—H · · · O(═C)									
nD	Cδ1—H(3) · · · O(═C)	2.76	137		0.007	0.028	1.214	0.006	0.001
nI	Cδ2—H(2) · · · O(═C)	2.81	133		0.007	0.026	0.975	0.005	0.001
zG	Cδ2—H(3) · · · O(2)	2.57	126		0.009	0.030	0.040	0.006	0.001
zH	Cδ1—H(2) · · · O(2)	2.42	136		0.011	0.037	0.063	0.008	0.001
zH	Cδ2—H(2) · · · O(2)	2.79	128		0.006	0.020	0.086	0.004	0.001
N—H · · · C (methyl)									
zI	N—H(2) · · · Cδ1	2.56	118		0.009	0.038	1.276	0.008	0.002
C—H · · · C									
nG	Cδ2—H(3) · · · C	2.68	108		0.007	0.029	1.257	0.006	0.002
zD	Cδ1—H(1) · · · C	2.64	111		0.007	0.030	0.934	0.006	0.002
zI	Cδ2—H(2) · · · C	2.64	111		0.007	0.031	1.185	0.006	0.002

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

The presence of intra-molecular hydrogen bonds in leucine rotamers results in the formation of ring(s) and the emergence of a ring critical point(s). In one case (**nH**) the surface of three rings are so disposed as to form a cage, accompanied by the appearance of a cage critical point. For every rotamer, the Poincaré-Hopf relation[1] is satisfied (number of nuclei − number of bonds + number of rings − number of cages = 1).The size of the ring structures formed in the 18 rotamers varies from five to seven membered. As can be expected, ring ρ_{cp} falls sharply with the size of the ring. All studied zwitter-ionic rotamers exhibit at least the hydrogen bond (N-H(3). . .O(2)=C) closing a 5-membered ring, with a ring ρ_{cp} clustering around 0.034.

Atomic Properties

The atomic properties listed in Table IX show a marked transferability for all atoms from rotamer to rotamer within each set and for atoms of the side chains across the

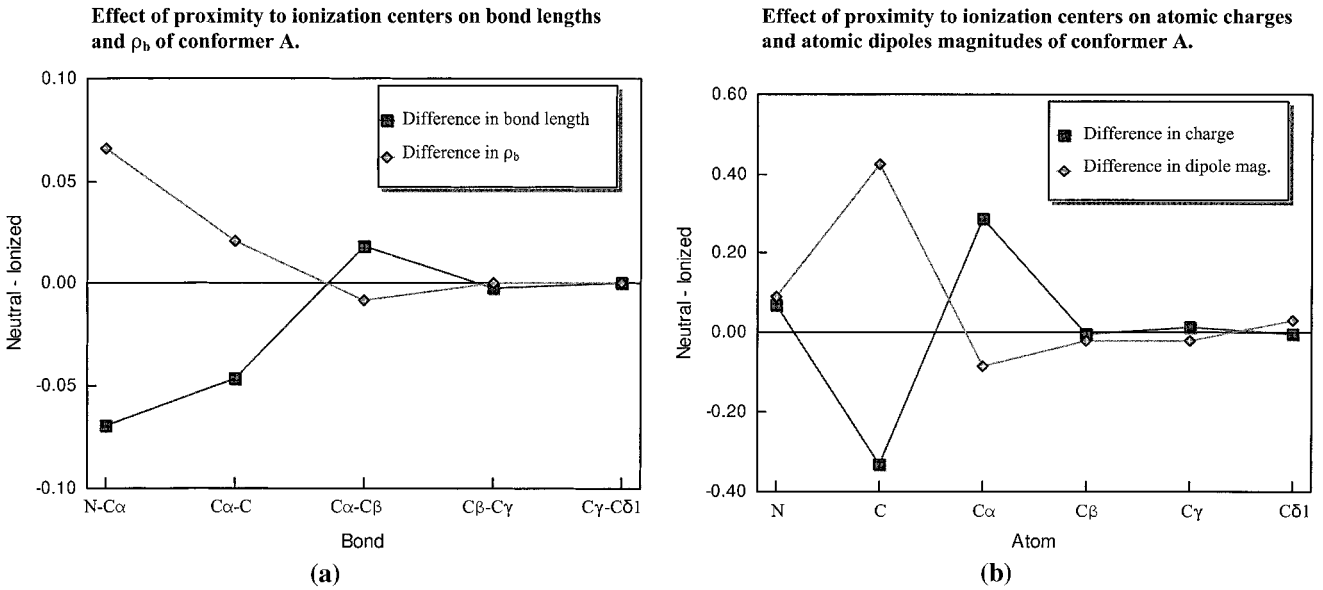


Fig. 4. The effect of tautomerization on (**a**) bond properties, and (**b**) atomic properties. These are plots of the difference between the bond or atomic property of the neutral **A** rotamer (**nA**) and the corresponding value in its zwitter-ionic counterpart (**zA**). Both plots indicate that the effect of the ionization state on bond and atomic properties is significant only on the groups directly involved in the tautometrization.

two sets. Only those atoms that were involved in the tautomerism and the atoms to which they are bonded are non-transferable across the **n** and the **z** sets. Conformation has minor effects on the studied atomic properties, as can be seen by the small magnitudes of the stds. For example, atomic energies show a near-constancy. Typically, the energy of the same atom in different rotamers differ by a few kcal/mol, a very small number when compared to the magnitude of the atomic energy (e.g. $\sim 5\times10^4$ kcal/mol for an oxygen atom). Relatively, properties with very small magnitudes are the ones that show the larger stds. Otherwise, a slightly elevated std always originate from the inclusion of the hydrogen-bonded atoms in the averaging, since the properties of these atoms are generally the most outlying ones. In order to show how hydrogen bonding affects the atomic properties, Table IX lists the properties of those atoms that are involved in hydrogen bonding explicitly in addition to the averages over all rotamers.

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

Effect of Proton-Transfer (Tautomerization) on Bond and Atomic Properties

The effect of ionization that leads to the formation of the zwitter-ionic tautomer is significant only for the bond and atomic properties involving the ionized atoms or those directly bonded to them. Generally, bond and atomic properties beyond Cβ change little in response to ionization. As an example, we consider the effect of ionization on rotamer **A**. A comparison of the bond properties and atomic properties of **nA** with the corresponding ones in **zA** shows how little the side chain is perturbed by tautomerization. A plot of the of the values of bond length (Å) and ρ_b (a.u.) for **nA** form minus the corresponding value for **zA** results in the difference plot shown in Figure 4a. The plot immediately reveals that the perturbations in these two properties are localized mainly at the N-Cα and Cα-C bonds. The Cα-Cβ bond is still slightly perturbed by the ionization but no bond beyond Cβ-Cγ shows a significant change in properties. Both the C-Cα and the Cα-N bonds are shorter and stronger (higher ρ_b) in the neutral form.

The same localization of the perturbation is also exhibited by the atomic properties. For example, Figure 4b shows the rapid damping of the perturbing effect of ionization on the atomic charge and the magnitude of the dipole moment at Cβ and beyond. The atoms primarily responsible for the increase in energy incurred in the formation of the zwitter-ion of rotamer A (E_τ = +26 kcal/mol) are readily accounted for by comparing the atomic energies in the two tautomers. The formation of the zwitter-ion results in the destabilization of atoms C, O(2), H(1), and H(2) by 146, 56, 26, and 24 kcal/mol, respectively, and in the stabilization of atoms Cα and N by 101 and 130 kcal/mol, respectively. The combined effect of ionization on these atoms alone is a destabilization energy of about 20 kcal/mol accounting for 77% of the destabilizing E_τ. In other words, the destabilization upon ionization is mostly localized at the ionized atoms and their immediate neighborhood. The only atoms whose populations are affected significantly on going from the **n**-tautomer to the **z**-tautomer (i.e., absolute difference is greater than 0.08e) are: C that loses 0.334e, Cα that gains 0.286e, and O(H) that gains 0.086e upon the loss of its hydrogen to the nitrogen. The nitrogen experiences a slight increase

TABLE IX. Selected Atomic Properties of Nine Neutral (n) and Three (A, E, and I) Zwitter-Ionic (z) Leucine Rotamers[†]

		Rotam.	H-bonded to	N(Ω)	E(Ω)	E(Ω) rel.kcal/mol	Vol.(Ω)[a]	q(Ω)	|**μ**(Ω)|	|**Q**(Ω)|	|L(Ω)|
C	**n**	**G**	Hδ2—3	4.231	−36.718	0.6	28.91	1.769	1.009	1.485	0.0034
	n	av.		4.233	−36.716		30.38	1.767	1.006	1.514	0.0022
		std		0.003	0.003		0.79	0.003	0.003	0.014	0.0012
	z	av.		3.900	−36.485		26.59	2.100	0.580	1.547	0.0013
		std		0.002	0.002		0.90	0.002	0.007	0.047	0.0010
Cα	**n**	av.		5.436	−37.505		40.99	0.564	0.506	1.058	0.0020
		std		0.008	0.007		0.40	0.008	0.006	0.101	0.0008
	z	av.		5.717	−37.663		47.65	0.283	0.586	1.686	0.0010
		std		0.002	0.003		0.13	0.002	0.002	0.023	0.0003
Cβ	**n**	av.		5.869	−37.746		52.34	0.131	0.099	0.408	0.0018
		std		0.004	0.007		0.36	0.004	0.004	0.040	0.0011
	z	av.		5.870	−37.747		53.22	0.130	0.128	0.565	0.0028
		std		0.002	0.009		0.31	0.002	0.007	0.013	0.0018
Cδ1	**n**	av.		5.862	−37.723		62.31	0.138	0.038	0.081	0.0009
		std		0.005	0.003		0.41	0.005	0.006	0.010	0.0006
	z	av.		5.865	−37.725		62.09	0.135	0.037	0.146	0.0009
		std		0.010	0.004		0.64	0.010	0.026	0.081	0.0006
Cδ2	**n**	av.		5.861	−37.722		62.25	0.139	0.037	0.094	0.0009
		std		0.003	0.002		0.37	0.003	0.004	0.023	0.0006
	z	av.		5.863	−37.724		62.01	0.137	0.049	0.183	0.0011
		std		0.006	0.001		0.27	0.006	0.021	0.025	0.0005
Cγ	**n**	av.		5.849	−37.705		46.94	0.151	0.043	0.189	0.0023
		std		0.006	0.004		0.53	0.006	0.017	0.052	0.0016
	z	av.		5.857	−37.708		47.43	0.143	0.071	0.328	0.0038
		std		0.003	0.002		0.56	0.003	0.018	0.049	0.0015
H(—O)	**n**	av.		0.352	−0.327		18.61	0.648	0.142	0.066	0.0000
		std		0.001	0.000		0.09	0.001	0.000	0.000	0.0000
H(1)	**n**	av.		0.641	−0.486		31.54	0.359	0.186	0.111	0.0001
		std		0.003	0.001		0.22	0.003	0.001	0.006	0.0000
	z	av.		0.561	−0.444		27.53	0.439	0.158	0.091	0.0001
		std		0.001	0.001		0.20	0.001	0.001	0.002	0.0000
H(2)	**n**	**F**	Hδ2—3	0.627	−0.482	1.6	28.37	0.373	0.175	0.136	0.0001
	n	**I**	Hδ1—2	0.627	−0.482	1.5	28.74	0.373	0.176	0.131	0.0000
	n	av.		0.629	−0.480		30.38	0.371	0.181	0.103	0.0001
		std		0.003	0.002		1.02	0.003	0.003	0.017	0.0000
	z	**I**	Cδ1	0.555	−0.439	2.2	25.83	0.445	0.151	0.165	0.0034
	z	av.		0.558	−0.441		26.84	0.442	0.156	0.113	0.0012
		std		0.002	0.002		0.73	0.002	0.004	0.037	0.0015
H(3)	**z**	av.[b]	O═C	0.406	−0.346		14.85	0.594	0.114	0.168	0.0003
		std		0.004	0.003		0.37	0.004	0.002	0.004	0.0001
Hα	**n**	av.		1.012	−0.646		45.77	−0.012	0.121	0.381	0.0001
		std		0.001	0.002		0.87	0.001	0.003	0.007	0.0001
	z	av.		0.995	−0.637		45.537	0.005	0.12	0.409	0.0001
		std		0.000	0.001		0.594	0.000	0.00	0.010	0.0000
Hβ1	**n**	av.		1.039	−0.646		48.92	−0.039	0.131	0.408	0.0001
		std		0.013	0.004		1.13	0.013	0.001	0.008	0.0000
	z	av.		1.064	−0.655		50.32	−0.064	0.131	0.421	0.0001
		std		0.009	0.004		0.45	0.009	0.002	0.001	0.0000
Hβ2	**n**	av.		1.040	−0.648		47.79	−0.040	0.125	0.420	0.0001
		std		0.004	0.003		0.75	0.004	0.003	0.010	0.0000
	z	av.		0.973	−0.619		44.63	0.027	0.123	0.399	0.0001
		std		0.012	0.005		1.15	0.012	0.002	0.004	0.0000
Hδ1—1	**n**	**B**	O(H)	1.045	−0.645	1.3	48.98	−0.045	0.121	0.489	0.0000
	n	**C**	N	1.004	−0.630	10.7	45.28	−0.004	0.120	0.508	0.0000
	n	av.		1.044	−0.642		49.75	−0.044	0.126	0.479	0.0001
		std		0.016	0.005		1.95	0.016	0.006	0.021	0.0000
	z	av.		1.056	−0.645		51.02	−0.056	0.129	0.467	0.0001
		std		0.025	0.010		1.00	0.025	0.002	0.003	0.0000
Hδ1—2	**n**	**H**	O(H)	1.057	−0.651	1.6	48.36	−0.057	0.116	0.515	0.0000
	n	**I**	H(2)	1.064	−0.653	0.0	49.70	−0.064	0.123	0.462	0.0001
	n	av.		1.057	−0.646		50.92	−0.057	0.130	0.507	0.0001
		std		0.006	0.004		1.28	0.006	0.006	0.123	0.0000
	z	av.		1.056	−0.646		50.28	−0.056	0.133	0.427	0.0016
		std		0.013	0.006		0.69	0.013	0.005	0.020	0.0014

TABLE IX. (Continued)

		Rotam.	H-bonded to	N(Ω)	E(Ω)	E(Ω) rel.kcal/mol	Vol.(Ω)[a]	q(Ω)	|μ(Ω)|	|**Q**(Ω)|	|L(Ω)|
Hδ1—3	n	**D**	O═C	1.049	−0.643	3.1	50.75	−0.049	0.131	0.466	0.0001
	n	av.		1.052	−0.644		51.07	−0.052	0.131	0.468	0.0001
		std		0.004	0.001		0.25	0.004	0.000	0.001	0.0000
	z	av.		1.039	−0.639		50.40	−0.039	0.130	0.463	0.0001
		std		0.009	0.004		0.46	0.009	0.000	0.005	0.0000
Hδ2—1	n	av.		1.052	−0.644		51.02	−0.052	0.131	0.468	0.0001
		std		0.004	0.001		0.28	0.004	0.000	0.002	0.0000
	z	av.		1.044	−0.641		50.63	−0.044	0.130	0.464	0.0001
		std		0.013	0.005		0.55	0.013	0.001	0.004	0.0000
Hδ2—2	n	**E**	N	1.012	−0.632	9.4	46.08	−0.012	0.125	0.480	0.0001
	n	**I**	O═C	1.048	−0.647	0.1	48.20	−0.048	0.116	0.502	0.0000
	n	av.		1.047	−0.642		50.45	−0.047	0.129	0.464	0.0001
		std		0.014	0.005		1.91	0.014	0.006	0.016	0.0000
	z	**A**[c]		1.034	−0.635	15.7	50.44	−0.034	0.130	0.452	0.0001
	z	**E**	N	1.087	−0.660	0.0	51.31	−0.087	0.127	0.474	0.0000
	z	**I**	C	1.009	−0.631	18.4	45.65	−0.009	0.115	0.482	0.0000
	z	av.		1.043	−0.642		49.13	−0.043	0.124	0.470	0.0000
		std		0.033	0.013		2.49	0.033	0.007	0.013	0.0000
Hδ2—3	n	**F**	H(2)	1.055	−0.651	0.8	48.30	−0.055	0.116	0.492	0.0001
	n	**G**	C	1.066	−0.652	0.0	49.47	−0.066	0.112	0.533	0.0031
	n	av.		1.056	−0.647		50.508	−0.056	0.126	0.479	0.0004
		std		0.006	0.003		1.064	0.006	0.007	0.021	0.0009
	z	av.		1.052	−0.645		50.51	−0.052	0.128	0.468	0.0001
	z	std		0.022	0.009		1.04	0.022	0.002	0.016	0.0000
Hγ	n	av.		1.068	−0.666		48.97	−0.068	0.128	0.404	0.0004
		std		0.011	0.004		0.98	0.011	0.003	0.008	0.0007
	z	av.		1.078	−0.670		49.71	−0.078	0.129	0.405	0.0001
		std		0.026	0.009		1.13	0.026	0.001	0.012	0.0000
N		**C**	Hδ1—1	8.147	−54.926	0.8	107.76	−1.147	0.194	2.066	0.0002
		E	Hδ2—2	8.149	−54.927	0.0	107.96	−1.149	0.192	2.072	0.0001
	n	av.		8.146	−54.914		111.43	−1.146	0.207	2.311	0.0006
		std		0.006	0.011		2.18	0.006	0.015	0.148	0.0010
	z	**E**	Hδ2—2	8.221	−55.131	0.0	90.61	−1.221	0.119	1.102	0.0008
	z	av.		8.220	−55.130		91.29	−1.220	0.113	1.091	0.0005
		std		0.001	0.002		0.63	0.001	0.008	0.008	0.0003
O(H)	n	**B**	Hδ1—1	9.302	−75.634	0.0	119.18	−1.302	0.310	0.939	0.0007
	n	**H**	Hδ1—2	9.299	−75.624	6.3	120.18	−1.299	0.304	0.930	0.0005
	n	av.		9.301	−75.625		119.78	−1.301	0.306	0.962	0.0005
		std		0.001	0.004		0.49	0.001	0.003	0.021	0.0007
O(═C)	n	**D**	Hδ1—3	9.358	−75.679	4.2	134.12	−1.358	0.583	0.518	0.0000
	n	**I**	Hδ2—2	9.357	−75.678	5.0	134.76	−1.357	0.583	0.518	0.0001
	n	av.		9.356	−75.677		135.37	−1.356	0.582	0.554	0.0001
		std		0.002	0.004		0.62	0.002	0.003	0.044	0.0000
O(1)	z	av.		9.387	−75.620		139.63	−1.387	0.486	0.721	0.0001
		std		0.001	0.001		0.21	0.001	0.002	0.040	0.0001
O(2)	z	av.		9.401	−75.585		133.55	−1.401	0.377	0.972	0.0001
		std		0.002	0.001		0.25	0.002	0.001	0.054	0.0000

[†]Only averages over the rotamer set are reported for each atom unless the atom is involved in hydrogen bonding, in which case its properties are explicitly given in addition to the set averages. Since this study is primarily concerned with neutral amino acid, only the highest (**A**), lowest (**I**) and middle (**E**) energy rotamers of the (**z**) set was subjected to atomic integrations in view of the computational costs. (Values are in a.u. unless stated otherwise).

[a]Volume of the atomic basin integrated up to the 0.001 a.u. isodensity envelope.

[b]All three averaged (**z**) rotamers exhibit this hydrogen-bond representing no exception to be listed separately. (In fact all the (**z**) rotamers exhibit this particular hydrogen-bond).

[c]This is the only one of the three averaged atoms that is *not* hydrogen bonded. It is thus the exception and deserves to be listed individually.

(0.069e) in population and hence is slightly more negative upon protonation. It achieves this increase in population by extracting electrons from the neighboring Cα forcing the less electronegative carbon to bear the increase in positive charge on the quaternized amino group. The functional group properties are the sums the atomic properties of the individual atoms making up the functional group. The functional group charges are -NH_2 (-0.420e), -COOH (-0.238e), -NH_3^+ (+0.257e), and -COO^- (-0.683e). In other words, -NH_3^+ is more positive relative to -NH_2 by 0.658e, and -COO^- is more negative by 0.445e relative to -COOH. The CαH fragment loses 0.270e upon

nA→zA tautomerization. The corresponding values for glycine are very similar: the amino group gains 0.672 positive charge, the carboxylic group gains 0.442 negative charge and a CαH fragment loses 0.266e. A comparison of the changes in the functional group populations upon tautomerization in leucine and in glycine shows how little the side chain perturbs the functional groups' response to this dramatic change in structure.

CONCLUSIONS

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

The present results show that the numerous possible conformations that are energetically accessible to the side chain does not present a serious problem in the use of amino acid residues in the modeling of proteins. Based on our results and those of earlier workers, the conformers accessible from an energy minimum that is characterized by normal bond lengths and bond angles, differ little in energy, and the geometric parameters are not significantly altered as various dihedral angles are sampled. This conservation of geometry in turn ensures the conservation of the form of the density distribution of corresponding atoms in different rotamers. As a consequence, the bond, atomic, and group properties defined by theory are found to be equally insensitive to conformational changes. The results of this article set the stage for the following one, wherein the bond and atomic properties of each of the free genetically-encoded amino acids are determined for a single minimum energy conformer.

The presence of strong hydrogen bonds, those with ρ_b values in excess of 0.02 a.u., do change the properties of the atoms directly involved in the interaction. The results presented here provide geometric parameters characterizing such interactions for each set of ρ_b values that enable one to predict the presence of such interactions and the resulting changes in the atomic properties. This information will be provided for the free amino acids and corrections will be included where necessary for the amino acid residues |Aa| in the following articles (in preparation).

APPENDIX: NUMBERING CONVENTIONS

The IUPAC-IUB conventions[48] were adopted unless otherwise stated. The hydrogens of the leucine rotamers were numbered as follows: i) nitrogen's H(1) has the smallest H-N-Cα-Hα angle; ii) Cβ's H(1) has the smallest Hβ-Cβ-Cα-Hα angle; and iii) Cδ's H(1) has the smallest Hδ-Cδ-Cγ-Hγ angle; these angles being those necessary to rotate the H clockwise when the viewer is from its side. For the hydrogens attached to a nitrogen of the zwitterion, the one involved in the H-bonding with the carboxyl oxygen [O(2)] is H(3); the two others are labeled as described above.

REFERENCES

1. Bader RFW. Atoms in molecules: a quantum theory. Oxford: Clarendon Press; 1990.
2. Bader RFW. Principle of stationary action and the definition of a proper open system. Phys Rev B 1994;49:13348–13356.
3. Bader RFW, Popelier PLA, Keith TA. Theoretical definition of a functional group and the molecular orbital paradigm. Ang Chem Int Ed Eng 1994;33:620–631.
4. Bader RFW. A bond path: a universal indicator of bonded interactions. J Phys Chem 1998;102:7314–7323.
5. Popelier PLA. Characterization of a dihydrogen bond on the basis of the electron density. J Phys Chem A 1998;102:1873–1878.
6. Spackman MA. Charge densities from x-ray diffraction data. R Soc Chem Ann Rep Section C 1998;94:177–207.
7. Keith TA, Bader RFW, Aray Y. Structural homeomorphism between the electron density and the virial field. Int J Quantum Chem 1996;57:183–198.
8. Cremer D, Kraka E. Chemical bonds without bonding electron density: does the difference electron-density analysis suffice for a description of the chemical bond? Angew Chem 1984;23:627–628.
9. Messer RR. The virial partitioning theory and its applications to molecular systems. Ph.D. thesis, McMaster University; 1977.
10. Cooper DL, Stutchbury NCJ. Distributed multipole analysis from charge partitioning by zero-flux surfaces: the structure of HF complexes. Chem Phys Lett 1985;120:167–172.
11. Popelier PLA. Integration of atoms in molecules: a critical examination. Molec Phys 1996;87:1169–1187.
12. Gray CG, Gubbins KE. Theory of molecular fluids, vol 1. Oxford: Clarendon Press; 1984. p 67–68.
13. Hagler AT, Huler E, Lifson S. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. J Am Chem Soc 1977;96:5319–5327.
14. Clementi E, Cavallone F, Scordamaglia R. Analytical potential from "ab initio" computations for the interaction between biomolecules. 1. Water with amino acids. J Am Chem Soc 1977;99:5531–5545. Sordo JA, Probst M, Corongiu G, Chin S, Clementi E. Ab initio pair potentials for the interactions between aliphatic amino acids. J Am Chem Soc 1987;109:1702–1708.
15. Momany FA, McGuire RF, Burgess AW, Scheraga HA. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, non-bonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J Phys Chem 1975;79:2361–2381.
16. Weiner SJ, Kollman PA, Nguyen DT, Case DA. An all atom force-field for simulation of proteins and nucleic acids. J Comp Chem 1986;7:230–252.
17. Sokalski WA, Poirier RA. Cumulative atomic multipole representation of the molecular charge distribution and its basis set dependence. Chem Phys Lett 1983;98:86–92.
18. Sokalski WA, Maruszewski K, Hariharan PC, Kaufman JJ. Library of cumulative atomic multipole moments. II. Neutral and charged amino acids. Int J Quant Chem Quant Biol Symp 1989;16:119–164.
19. Price SL, Faerman CH, Murray CW. Toward accurate transferable electrostatic models for polypeptides: a distributed multipole study of blocked amino acid residue charge distributions. J Comp Chem 1991;12:1187–1197.
20. Stone AJ. Distributed multipole analysis, or how to describe a molecular charge distribution. Chem Phys Lett 1981;83:233–239.
21. Stone AJ, Alderton M. Distributed multipole analysis methods and applications. Molec Phys 1985;56:1047–1064.
22. Ángyán JG, Jansen G, Loos M, Hättig C, Hess BA. Distributed polarizabilities using the topological theory of atoms in molecules.

Chem Phys Lett 1994;219:267–273. Stone AJ, Hättig C, Jansen G, Ángyán JG. Transferability of topologically partitioned polarizabilities: the case of n-alkanes. Mol Phys 1996;89:595–605. Hättig C, Jansen G, Hess BA, Ángyán JG. Intermolecular interaction energies by topologically partitioned electric properties. II. Dispersion energies in one-centre and multicentre multipole expansions. Molec Phys 1997;91:145–160.
23. Bader RFW. Why are there atoms in chemistry. Can J Chem 1998;76:973–988.
24. Massa L, Huang L, Karle J. Quantum crystallography and the use of kernel projector matrices. Int J Quantum Chem Quantum Chem Symp 1995;29:371–384. Huang L, Massa L, Karle J. Kernel projector matrices for Leu1-zervamicin. Int J Quantum Chem Quantum Chem Symp 1996;30:479–488.
25. Pichon-Pesme V, Lecomte C, Lachekar H. On building a data bank of transferable experimental density parameters: application to polypeptides. J Phys Chem 1995;99:6242–6250.
26. Hansen, NK, Coppens P. Testing aspherical atom refinement on small molecules data sets. Acta Cryst 1978;A34:909–921.
27. Chang C, Bader RFW. Theoretical construction of a polypeptide. J Phys Chem 1992;96:1654–1662.
28. Bader RFW, Martín F. Interdeterminacy of basin and surface properties of an open system. Can J Chem 1998;76:284–291.
29. Gronert S, O'Hair RAJ. Ab initio studies of amino acid conformations. 1. The conformers of alanine, serine, and cysteine. J Am Chem Soc 1995;117:2071–2081.
30. Shirazian S, Gronert S. The gas-phase conformations of valine: an ab initio study. J Molec Struct (Theochem) 1997;397:107–112.
31. Ramek M, Cheng VKW. On the role of polarization functions in SCF calculations of glycine and related systems with intramolecular hydrogen bonding. Int J Quantum Chem Quantum Biol Symp 1992;19, 15–26.
32. See for example: Karle IL, Karle J. [Leu^5]enkephalin: four cocrystallizing conformers with extended backbones that form an antiparallel β-sheet. Acta Cryst B 1983;39:625–637.
33. Creighton TE. Proteins: structures and molecular principles. New York: W.H. Freeman and Co.; 1983. p 249.
34. Watanabe T, Hashimoto K, Takase H and Kikuchi O. Monte Carlo simulation study on the conformation and interaction of the glycine zwitterion in aqueous solution. J Mol Struct (Theochem) 1997;397:113–119.
35. Head-Gordon T, Head-Gordon M, Frisch MJ, Brooks C III, Pople JA. Theoretical study of blocked glycine and alanine peptide analogues. J Am Chem Soc 1991;113, 5989–5997.
36. Popelier PLA, Bader RFW. The existence of an intramolecular C-H-O hydrogen bond in creatine and carbamoyl sarcosine. Chem Phys Lett 1992;189, 542–548 .
37. Chipot C., Maigret B, Rivail J-L, Scheraga H. Modelling amino acid side chains. 1. Determination of net atomic charges from ab initio self-consistent-field molecular electrostatic properties. J Phys Chem 1992;96:10276–10284 (and references therein).
38. Hyperchem, Version 5.01. Waterloo, Canada: Hypercube, Inc.; 1996.
39. Gaussian 94, Revision B.3. Frisch MJ, Trucks GW, Schlegel HB, Gill PMW, Johnson BG, Robb MA, Cheeseman JR, Keith T, Petersson GA, Montgomery JA, Raghavachari K, Al-Laham MA, Zakrzewski VG, Ortiz JV, Foresman JB, Peng CY, Ayala PY, Chen W, Wong MW, Andres JL, Replogle ES, Gomperts R, Martin RL, Fox DJ, Binkley JS, Defrees DJ, Baker J, Stewart JP, Head-Gordon M, Gonzalez C, Pople JA. Pittsburgh PA: Gaussian Inc.; 1995.
40. Biegler-König FW, Nguyen-Dang TT, Tal Y, Bader RFW, Duke AJ. Calculation of the average properties of atoms in molecules. J Phys B 1981;14:2739–2751. Biegler-König FW, Bader RFW, Tang T-H. Calculation of the average properties of atoms in molecules. II. J Comp Chem 1982;13:317–328. AIMPAC program suite can be downloaded free of charge from http://www.chemistry.mcmaster.ca/aimpac/
41. Kebarle P. Solvent effects on acidity and basicity from gas phase ion equilibria measurements. Jerusalem Symp Quantum Chem Biochem (1975) 1976;8:81–94.
42. Tse Y-C, Newton MD, Vishveshwara S and Pople JA. Ab initio studies of the relative energetics of glycine and its zwitterion. J Am Chem Soc 1978;100:4329–4331.
43. Görbitz CH, Dalhus B. Redetermination of L-leucine at 120 K. Acta Cryst C 1996;52:1754–1756.
44. Carroll MT, Bader RFW. An analysis of the hydrogen bond in BASE-HF complexes using the theory of atoms in molecules. Molec Phys 1988;65:695–722.
45. Cheeseman JR, Carrol MT, Bader RFW. The mechanics of hydrogen bond formation in conjugated systems. Chem Phys Lett 1988;143:450–458.
46. Destro R, Bianchi R, Gatti C, Merati F. Total electronic charge density of L-alanine from x-ray diffraction at 23 K. Chem Phys Lett 1991;186:47–52.
47. Popelier PLA, Bader RFW. Effect of twisting a polypeptide on its geometry and electron distribution. J Phys Chem 1994;98:4473–4481.
48. CRC Handbook of biochemistry and molecular biology, vol. II: proteins (third edition). Fasman GD, editor. Cleveland, Ohio: CRC Press, Inc.; 1976; p 63–78. CRC handbook of biochemistry. Sober HA, editor; Harte RA, chairman. Cleveland, Ohio: The Chemical Rubber Co.; 1968. p A–69.