Comprehensive Analysis of Energy Minima of the Twenty Natural Amino Acids.

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY A · AUGUST 2014

Impact Factor: 2.69 · DOI: 10.1021/jp503460m · Source: PubMed

4 AUTHORS, INCLUDING:

Paul L A Popelier
The University of Manchester
190 PUBLICATIONS **7,122** CITATIONS
SEE PROFILE

Frank Jensen
Aarhus University
153 PUBLICATIONS **6,344** CITATIONS
SEE PROFILE

Matthew J L Mills
Joint BioEnergy Institute
13 PUBLICATIONS **107** CITATIONS
SEE PROFILE

THE JOURNAL OF PHYSICAL CHEMISTRY A

Article

pubs.acs.org/JPCA

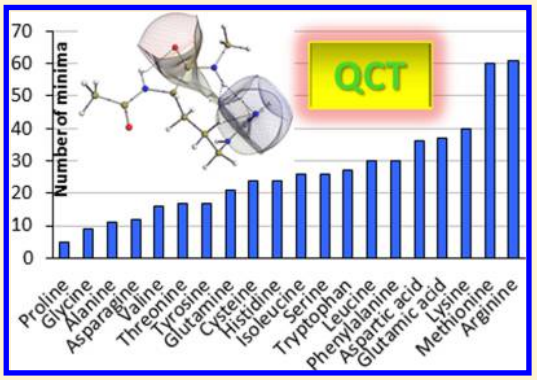
Comprehensive Analysis of Energy Minima of the 20 Natural Amino Acids

Yongna Yuan,[†,‡,§] Matthew J. L. Mills,[†,‡,∥] Paul L. A. Popelier,*[,†,‡] and Frank Jensen[⊥]

[†]Manchester Institute of Biotechnology (MIB), 131 Princess Street, Manchester M1 7DN, Great Britain and
[‡]School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, Great Britain
[⊥]Department of Chemistry, Aarhus University, Aarhus DK-8000, Denmark

Supporting Information

ABSTRACT: Energy minima of the 20 natural amino acids (capped by a peptide bond at both the N and the C termini, $CH_3-C(=O)-N(H)-(H)C_\alpha(R)-C(=O)-N(H)-CH_3$), were obtained by ab initio geometry optimization. Starting with a large number of minima, quickly generated by MarvinView, geometry optimization at the HF/6-31G(d,p) level of theory reduced the number of minima, followed by further optimization at the B3LYP/apc-1 and MP2/cc-pVDZ levels, which caused some minima to disappear and some stable minima to migrate on the Ramachandran map. There is a relation between the number of minima and the size and the flexibility of the side chain. The energy minima of the 20 amino acids are mainly located in the regions of β_L, γ_L, δ_L, and α_L of the Ramachandran map. Multipole moments of atoms occurring in the fragment $[-NH-C_\alpha-C(=O)-]$ common to all 20 amino acids were calculated at the three levels of theory mentioned above. The near parallelism in behavior of these moments between levels of theory is beneficial toward estimating moments with the more expensive B3LYP and MP2 methods from data calculated with the cheaper HF method. Finally, we explored the transferability of properties between different amino acids: the bond length and angles of the common fragment $[-NH-C_\alpha(H_\alpha C_\beta)-C'(=O)-]$ in all amino acids except Gly and Pro. All bond lengths are highly transferable between different amino acids, and the standard deviations are small.



1. INTRODUCTION

As the fundamental building blocks of proteins, individual amino acids (AAs) continue to receive much attention from both experiment and computation. For example, in 2013, a state-of-the-art microwave spectroscopy study[1] established that *N*-acetyl-L-alaninamide (Ac-Ala-NH_2) exists in the gas phase as a mixture of two conformers stabilized by a CO···HN intramolecular hydrogen bond, closing either a seven- or a five-membered ring. In another example published in the same year, researchers combined[2] resonance-enhanced multiphoton ionization, ultraviolet−ultraviolet hole burning, infrared dip spectra, and quantum chemical calculation to find that 12 tyrosine conformers coexist in the supersonic jet of their laser spectroscopy.

The purely computational investigation of the potential energy surfaces of the 20 naturally occurring AAs in terms of the familiar backbone dihedral angles ψ and φ spans more than two decades. It is straightforward to quote at least one study that targeted one (or sometimes more[3]) AA at a time: asparagine,[4] glycine,[5] valine,[6,7] alanine,[8−10] threonine,[11] phenylalaline,[12,13] serine,[14−16] glutamine,[17] isoleucine,[18] methionine,[19] cysteine,[20−22] proline,[23] aspartic acid,[24] histidine,[25] tryptophan,[26−28] tyrosine,[29,30] glutamic acid,[31] leucine,[32] lysine,[33] and arginine,[34] to which one can add selenocysteine[35] for completeness. Each of the computational studies in this nonexhaustive list demonstrates the complexity of the potential energy surface of a single AA, either with neutral NH_2 and COOH terminal groups or capped by peptide bonds at both the N and the C terminus. It is common for an AA to have several dozen local energy minima on its potential energy surface. This complexity surely adds to the flexibility and rich behavior that proteins reveal in their structure and function.

The limited computer resources in the 1990s meant that studies of that era were often carried out at levels of theory nowadays considered as very modest. However, very low levels of theory such as HF/3-21G already identified[23] the global energy minimum of, for example, *N*-formyl-L-prolinamide. This global minimum was designated by γ_L using a nomenclature referring to a square region covering 1/9th of the Ramachandran map (g^+ $0° < \psi < 120°$; g^- $-120° < \varphi < 0°$). Continuing with this typical example of a single-amino-acid study, two other energy minima in the same ψ interval (denoted α_L and ε_L) appeared at the HF/3-21G level but disappeared at two higher levels of theory used (HF/6-31G* and B3LYP/6-31G*). The vanishing and appearing of minima upon a change in the level of theory is characteristic for all AAs. This variation in potential energy surface demonstrates the subtlety found in the interaction

Received: April 8, 2014
Revised: July 18, 2014
Published: August 1, 2014

   dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

between various energy contributions, which differ in both the physical nature and the molecular fragments involved. A second variation that is typically observed (e.g., for phenylalanine[12]) is the migration of a preserved energy minimum upon a change in the level of theory.

An understanding of the conformational preferences of a single AA in terms of the dihedral angles ψ and φ is the very first step in attempting to understand local conformational preferences in proteins. For those AAs with rotatable side chains, these side chains (described by the parameter χ) also need to be considered as they play an important role in the stabilization of the AAs. It is well known that side chains can form intramolecular hydrogen bonds with the protein backbone, and a single AA study can already reveal such important effects. Work that focuses on the conformational landscape of covalently bonded AA dimers such as tyrosine-glycine[36] identified the presence of an OH···O hydrogen bond between Tyr and Gly to be a defining structural characteristic in the 20 most stable TyrGly conformers. Studies on tripeptides such as TyrGlyGly[37] or GlyGlyGly[38] expectedly identified more elaborate intramolecular hydrogen-bond motifs as hallmarks of stability. Work on nine tripeptides[39] obtained by combining Gly, Ala, and Ser around a central glycine residue showed significant variations of the geometry and atomic properties of the central glycine residue when it is attached to a serine residue whose side chain is involved in a hydrogen bond. Because the current article focuses on single AAs, those effects cannot be observed, in contrast to the side-chain stabilization effect.

The geometrical flexibility of the AAs causes a large number of local minima to appear in the potential energy surface. Even the simplest AA (glycine) shows a Ramanchandran map that is not easily understood,[40] and identification of this AA's specific interactions will improve current force fields and help understand structural motifs. Revisiting the standard architecture that underpins all popular protein force fields is important to make sustainable progress toward a more reliable treatment of interatomic interactions. The design of a novel protein force field is perhaps best based on ab initio information because it can comprehensively and consistently cover all AAs. This paper provides ab initio information on all AAs, carried out at the same levels of theory for each AA.

Some time ago we introduced multipolar electrostatics[41−45] for AAs as a route to tackle the electrostatic energy contribution more accurately than atomic point charges do.[46] High-rank atomic multipole moments (up to hexadecapole moment) were used and defined via the spherical tensor formalism.[42,47] The atoms themselves are obtained as naturally occurring subspaces in the gradient vector field of the electron density, as detailed by the "quantum theory of atoms in molecules",[48,49] which is subsumed in Quantum Chemical Topology (QCT).[50,51] Other than defining atomic properties, QCT is also able to characterize bonding patterns by local properties evaluated at so-called bond critical points. Two other types of critical point (ring and cage) will be featured in the current article. Early QCT work on peptides[52] quantified the transferability of the glycyl fragment in GlyGlyGly and later systematically studied[53] the effect of twisting a polypeptide on its geometry, electron distribution, and hydrogen bonding in the 3.6_{13} α-helical geometry of *N*-formyltriglycine. A few years later Matta and Bader published[54] an exhaustive QCT analysis on the effects of conformation and tautomerisation on geometric, atomic, and bond properties of all 20 AAs, as "uncapped" zwitterions and geometry optimized at the HF/6-31+G* level. Their subsequent study[55] demonstrated that the geometric parameters for the bonds and common functional groups of both the (protein) backbone and the side chains exhibited a high degree of transferability. Their third and final study[56] revealed several remarkable correlations between QCT atomic properties (volumes, sum of absolute values of atomic charges) and high-level properties such as the octanol−water partitioning coefficient and single-site mutation-induced changes in protein stability as measured by scanning calorimetry.

In the present work we comprehensively investigate hydrogen bonds between the side chain and the backbone in order to obtain insight into the effect of the size of the side chain on the energy and geometry of each AA. Moreover, in order to demonstrate that the properties of single AAs are transferable to a whole protein, we study the degree of transferability of the common structural elements between different AAs. We compute the minima of all 20 naturally occurring AAs capped by peptide bonds at both termini or $CH_3-C(=O)-N(H)-(H)C_\alpha(R)-C(=O)-N(H)-CH_3$, where R represents the residue that marks the AA. All calculations were carried out at the HF/6-31G(d,p), B3LYP/apc-1,[57] and MP2/cc-pVDZ[58] levels of theory. These levels are admittedly not state of the art, but the overall purpose of this study is to provide consistent training data for a novel protein force field that is under construction in Manchester and currently referred as QCTFF. The apc-1 basis set (which is a polarization-consistent (pc) double-ζ plus polarization basis set with diffuse functions) was used for the DFT calculations, since this family of basis sets has been specifically optimized for DFT. The cc-pVDZ was used for the MP2 calculations, since this family of basis sets has been specifically optimized for correlated wave function methods. MP2 calculations are computationally more expensive than DFT, and given the large number of structures, diffuse functions were not employed at the MP2 level.

There are four more detailed aims to the current work. First, we establish all possible minimum energy molecular geometries of each AA and then optimize these minima at the three levels of theory. The second aim is an analysis of the geometrical parameters, including the dihedral angles, total energies, and side-chain effects. This will be helpful in the investigation of the stabilization of the molecular geometries of all AAs. The third aim is to study the influence of the different levels of theory on atomic multipole moments. The atomic multipole moments of the global minimum of each AA are calculated at the three levels of theory. The fourth and final goal is to assess the degree of transferability of the properties of the common parts between different AAs.

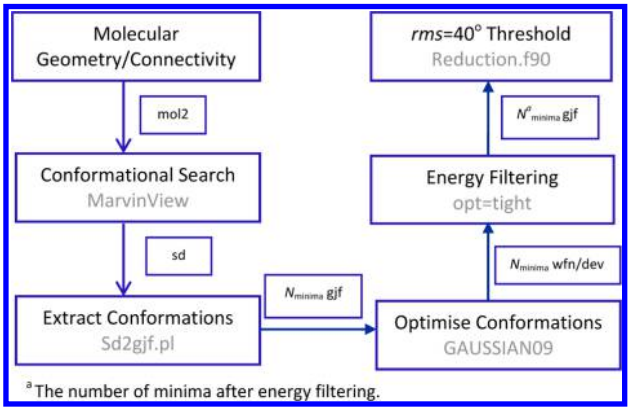
2. BACKGROUND AND COMPUTATIONAL DETAILS

2.1. Location of Energy Minima. Scheme 1 shows how the energy minima were obtained for all of the 20 AAs. The molecular geometries of the AAs are usually described by two Ramachandran angles φ and ψ as shown in Figure 1. The symbol φ characterizes the $C_{18}-N_1-C_2-C_3$ dihedral angle, while ψ denotes the $N_1-C_2-C_3-N_{12}$ dihedral angle. For the AAs with rotatable side chains, the χ_k dihedral angles appearing in the side chains are also considered because they play a key role in the stabilization of both AAs and proteins. The number of χ angles, m (see eq 1), varies from AA to AA, and these angles are labeled by index k (= 0, 1, 2, ..., m).

We will now discuss the various steps in Scheme 1. First, each AA is capped by an *N*-acetyl group at the N terminus and an *N*-methylamino at the C terminus resulting in the schematic

dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

Scheme 1. Various Stages of Generating Energy Minima for Each of the 20 AAs with Details of the in-House Fortran 90 Programs and Perl Scripts Used



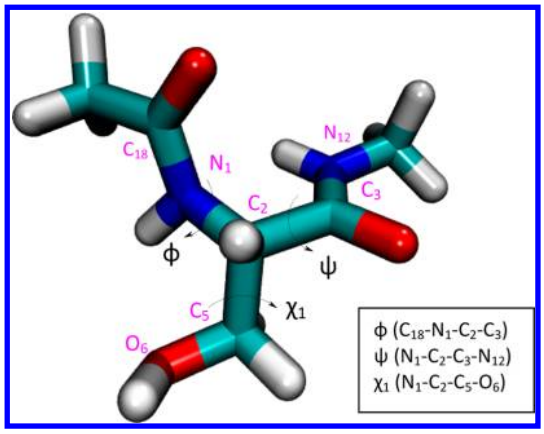


Figure 1. Molecular geometry of serine labeled with the φ, ψ, and χ dihedral angles.

structure $CH_3-C(=O)-N(H)-(H)C_\alpha(R)-C(=O)-N(H)-CH_3$, where R group marks the AA's side-chain. Note that each terminal methyl group represents the C_α atom of the AA adjacent to the central one (marked by R) in a peptide or protein environment. Subsequently, the program MarvinView (ChemAxon Ltd.)[59] uses the Dreiding force field[60] to produce a series of stable molecular geometries (in the vicinity of a minimum) by allowing for all internal single-bond rotations with a specified diversity limit. Then all geometries of the 20 AAs are optimized at the HF/6-31G(d,p) level of theory using the program GAUSSIAN09[61] with the qualifier "tight" in the geometry optimization input instruction. This geometry optimization may cause two minima to collapse to the same minimum. Hence, duplicate molecular geometries (i.e., those with the same energy) are filtered out. The remaining geometries are grouped into subgroups based on energy. Each subgroup contains more than one geometry, which means that all subgroups are possible minima. One geometry from each subgroup is arbitrarily selected as a possible minimum geometry. The harmonic frequencies of these selected geometries are subsequently calculated to make sure all geometries are true minima. For most of the 20 AAs, a large number of minima still exist after this filtering. Taking arginine as an example shows that a total of 172 geometries still exist after application of the energy filter. The need to further reduce the number of minima of each AA remains and is achieved via a root-mean-square (rms) threshold of the torsion angles of φ, ψ, and χ_k between different geometries. The details of this rms measure are discussed later. During the reduction of the number of minima, only non-hydrogen atoms are rotated in the 20 AAs while torsion angles involving hydrogen atoms are not varied because of their small influence on the flexibility of the whole AA. For example, in Ser only χ_1 is considered and χ_2 ($C_2-C_5-O_6-H$) is not (see Figure 1).

In this study, the geometrical difference between two molecular configurations is calculated on the basis of the rms of all torsion angles (φ, ψ, and χ_k) that are present in each AA. This difference is expressed by means of an rms value defined in eq 1

$$\mathrm{rms}_{ij} = \sqrt{\frac{(\varphi_i - \varphi_j)^2 + (\psi_i - \psi_j)^2 + \sum_{k=1}^{m} (\chi_{ki} - \chi_{kj})^2}{n}} \quad (1)$$

where i and j represent two different molecular geometries and n ($n = m + 2$) is the total number of torsional angles used to express their geometrical difference. Calculations were performed by an in-house Fortran 90 code called *Reduction*. Equation 1 allows for a varying number of χ angles depending on the AA. For example, Gly has no χ angles ($m = 0$), while Asp has three χ angles ($m = 3$). Guidance toward a cutoff value of rms = 40° in order to decide if two minima can be considered as clearly separable is found in work by Perczel et al., who extensively studied the distribution of Ser,[14] Phe,[12] and Pro[23] minima in the Ramachandran map. They retrieved the geometries of the three AAs they studied from thousands of proteins in the Protein Data Bank (PDB). The distributions (in the Ramachandran map) of the minima of these three residues obtained using the rms = 40° threshold in our work are similar to the distributions in their work.[12,14,23] The final number of minima was decided for each of the 20 AAs through calculation of the corresponding rms values, which are then optimized at the HF/6-31(d,p) level. The harmonic frequencies of these geometries are subsequently computed at that level, and the geometries with imaginary frequencies are removed to make sure that the optimized structures are true minima. Subsequently, these minima are further optimized at the B3LYP/apc-1[57] and MP2/cc-pVDZ[58] levels.

Two important comments are in place here. First, all five ionizable AAs were treated with their side-chain functional group in the neutral state. In particular, Asp and Glu were both protonated (i.e., COOH), while Arg, Lys, and His did not have an extra proton on their basic nitrogen atom (i.e., NH_2 and unprotonated imidazole). We are aware of the fact that these amino acids appear in a rare protonation state, and incorporating the right protonation state is important in EVB work, for example, particularly for His. The difference between the protonation states of these five amino acids is treated in terms of charge localization in a different forthcoming publication,[62] and also in terms of QCTFF's machine learning training, which is currently ongoing.

The second comment concerns the value of sampling gas-phase geometries for the training by machine learning of QCTFF. Work in the research group has already started with an eye on obtaining an answer to this challenging question while working with the PDB. While the geometries in this database are locally often unrealistic (due to poor crystallographic refinement) they provide a more realistic and informed spectrum of secondary structure than gas-phase structure. We are working on a sampling technique that alleviates this tension and thus hopefully provides the relevant geometries for

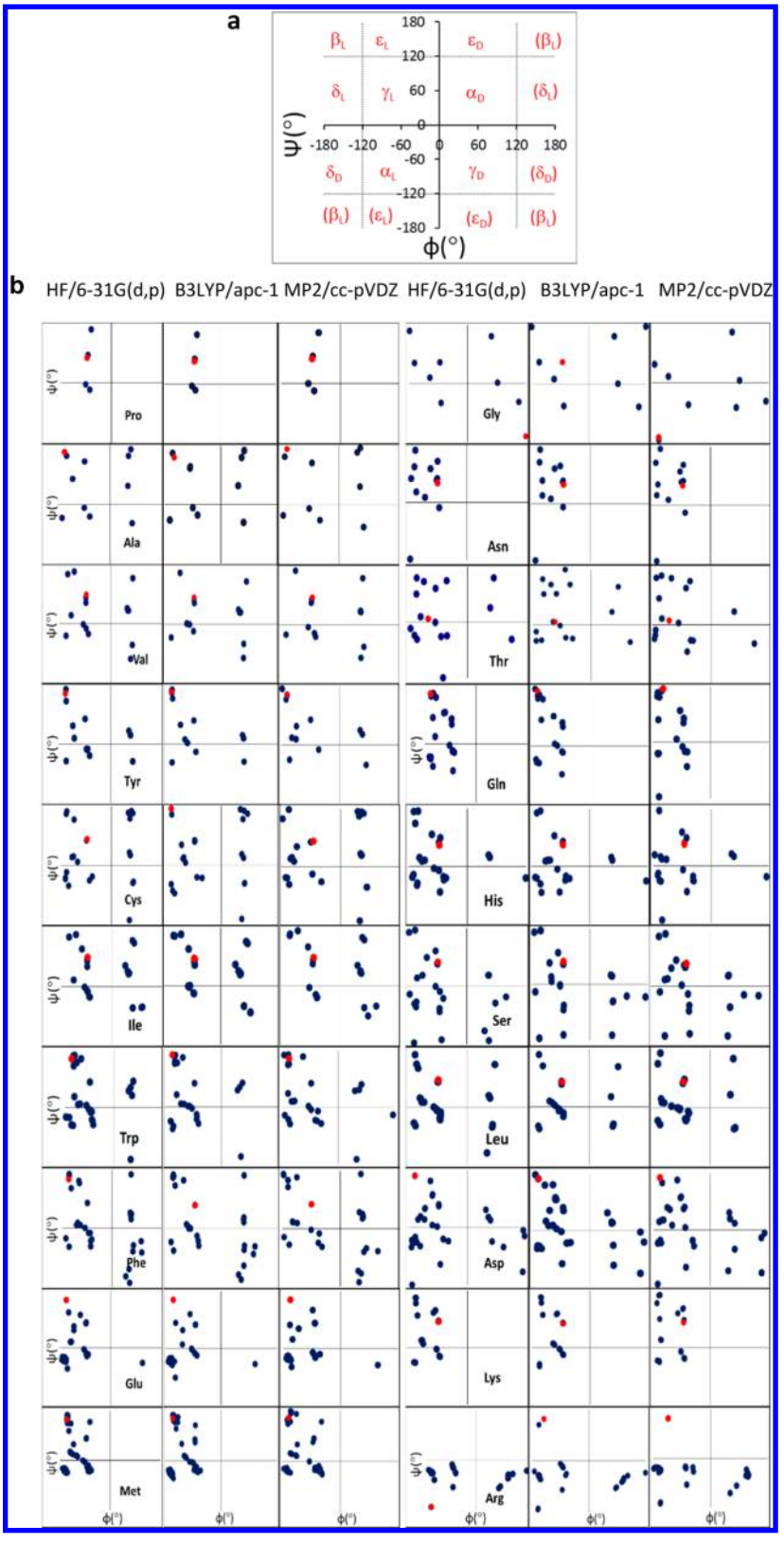


Figure 2. Ramachandran map of the minimum energy geometries of the 20 natural AAs ($E = E[\varphi,\psi]$) at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ levels of theory. Red dot represents the position of the global minimum in each case.

the training of the machine learning method at the heart of QCTFF.

2.2. Quantum Chemical Topology (QCT). Atomic multipole moments fully represent the electron density of a given atom inside a molecule. In this paper we delineate atoms following quantum chemical topology, which is a generalization of the quantum theory of atoms in molecules.[48,49] The only mathematical object that is necessary to define an atom in

dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

a molecule is the gradient of $\rho(\vec{r})$. When operating on the electron density the gradient naturally carves out subspaces that are called topological atoms. This minimal partitioning decision creates a wealth of atomic shapes with a precise imprint of the chemical environment the atoms find themselves in. An important feature of topological atoms is that they do not overlap with each other and that they do not leave gaps between each other either; in other words, they exhaust real space. A molecule thus falls apart into disjoint regions in real 3D space, separated by sharp boundaries called interatomic surfaces. Atomic multipole moments are obtained by integrating the electron density over the volume of a topological atom. The atomic charge is essentially the zeroth-order multipole moment (or monopole moment) corrected for the nuclear charge that the atom contains. Atomic multipole moments are formulated within the context of regular spherical harmonics (see Appendix of ref 42). This formalism is more compact than the perhaps more familiar Cartesian formalism[63] and shares the symbols of the familiar angular solutions of the Schrödinger equation of the hydrogen atom (i.e., s, p, d, f orbitals). Indeed, the general atomic multipole moment Q_{lm} has an index l that refers to its rank and an index m referring to any of its $2l + 1$ independent components. For example, the rank of the quadrupole moment is 2, which has therefore 5 components. The program MORPHY[64,65] was used (default settings) to obtain all atomic properties.

3. RESULTS AND DISCUSSION

3.1. Number of Minima for the 20 Amino Acids and Their Side-Chain Effects. An energy minimum is considered different from another minimum if the torsion-angle-based rms measure (see eq 1) is greater than 40°. The Ramachandran maps $E[\varphi, \psi]$ with $-180° < \varphi < +180°$ and $-180° < \psi < +180°$ at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ levels are schematically depicted for each AA in Figure 2. The $E[\varphi, \psi]$ map is divided into nine regions: α_L, α_D, β_L, δ_L, δ_D, γ_L, γ_D, ε_L, and ε_D (Figure 2a). This division helps in distinguishing the geometries of the energy minima and labeling the preferred regions of population in the Ramachandran map. Figure 2b shows the distribution of energy minima in this map, where the red dot represents the global minimum of each of the 20 AAs. The β_L, γ_L, δ_L, and α_L regions are the most populated for most of the 20 AAs. For most AAs, the φ dihedral angles of energy minima predominantly occur in the left region of the Ramachandran map in the range from around −50° to −175°. For all AAs, most of their minima are located in the left region in the Ramachandran map. The density of minima in the right region is low, and the minima are mainly in the φ range from 50° to 120°. When seen from the point of view of ψ, most minima occur in the region from −55° to 175°. Moreover, most energy minima move only a little in the Ramachandran map when going from one level of theory to another. Finally, for Arg, many minima overlap, which is why this panel appears to show fewer than the 61 minima mentioned in Table 1, which we discuss now.

The number of minima found for each of the 20 AAs is listed in Table 1 and also shown in Figure 3. The presence of the side chains modifies the number of minima and their distribution in the Ramachandran map. Table 1 ranks all 20 AAs by increasing number of minima, according to the energies obtained at the HF/6-31G(d,p) level. The number of minima for seven AAs (i.e., Val, Ile, Trp, Leu, Glu, Lys, and Met) is reduced by one, two, or three when the HF/6-31G(d,p) geometry is reoptimized at the B3LYP/apc-1 and MP2/cc-pVDZ levels.

Table 1. Number of Minima for Each of the 20 AAs Obtained at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ Levels

		no. of minima[b]		
AAs	no. of minima[a]	HF/6-31G(d,p)	B3LYP/apc-1	MP2/cc-pVDZ
Pro	17	5	5	5
Gly	19	9	9	9
Ala	31	11	11	11
Asn	28	12	12	12
Val	62	16	15	15
Thr	85	17	17	17
Tyr	67	17	17	17
Gln	51	21	21	21
Cys	142	24	24	24
His	81	24	24	24
Ile	99	26	25	25
Ser	142	26	26	26
Trp	90	27	26	26
Leu	121	30	28	28
Phe	61	30	30	30
Asp	100	36	36	36
Glu	98	37	36	36
Lys	187	40	39	39
Met	167	60	57	57
Arg	172	61	61	61

[a]Number of minima obtained after application of the energy filter. [b]Number of minima obtained based on the threshold of rms = 40°.

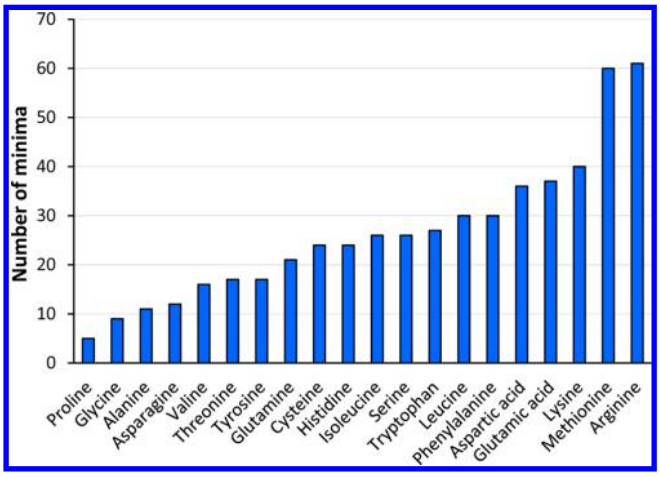


Figure 3. Number of minima of all 20 AAs optimized at the HF/6-31G(d,p) level. The 20 AAs are ranked by an increasing number of minima.

When the geometry of such a disappearing minimum is re-optimized at a correlated (non-HF) level of theory then an imaginary frequency arises, which turns this HF energy minimum into a transition state. Note that the number of minima is the same in going from B3LYP/apc-1 to MP2/cc-pVDZ for all AAs. Depending on the different types of side chain, the 20 AAs are divided into seven subgroups. These subgroups will be discussed next, one by one, and always referring to the energy minima obtained at the HF/6-31G(d,p) level.

The first subgroup only contains one AA, namely, Pro. Proline is unique among the 20 natural AAs as its side chain is bonded to the amide nitrogen and forms a five-membered ring, which hampers free rotation of Pro. Therefore, proline's flexibility is limited by the constrained ring. Because of this

geometrical restriction, Pro has far fewer minima than any of the other AAs, which can be seen in Table 1 and Figure 3.

The second subgroup includes glycine (R = −H) (Gly) and alanine (R = $-CH_3$) (Ala), which are the simplest residues yielding 9 and 11 minima, respectively. Many computational studies[66] have been published on glycine and alanine. Structures of Gly and Ala behave in a similar way[66] as the only difference between them is that the H atom in Gly is replaced by the methyl group in Ala. The methyl group is small and nonpolar and, similar to H in Ala, cannot form hydrogen bonds with the backbone. Therefore, CH_3 has a weak effect on the geometry of Ala.[67]

Valine (Val), leucine (Leu), isoleucine (Ile), and methionine (Met), which make up the third subgroup, are collated since their side chains are quite similar to each other. The influence of the side chain in Val,[68,69] Leu,[70] and Ile[71] has been reported before. It was concluded that the size and geometries of these side chains significantly affect their minimum energy molecular geometries. The isopropyl side chain in Val is much more bulky and flexible than the methyl group in Ala, and thus, Val has 16 minima. Hence, the size of the side chain plays a major role in the flexibility of AAs. This assertion is corroborated by comparing the number of minima between Val and Leu. Leucine has one more CH_2 group in its side chain than Val and 30 minima, which is many more than the number of minima of Val, which is only 16. Moreover, the bonding pattern within the side chain also has a great influence on the rotation of AAs. The number of atoms in the side chains in Leu and Ile is the same, but their arrangement is different. Isoleucine has 4 fewer minima than Leu. Another possible factor may be the steric repulsion between the *sec*-butyl side chain and the backbone in Ile, which is larger than that between the isobutyl side chain and the backbone in Leu. Methionine (Met) has a large number of minima due to its large and flexible side chain: 167 after initial energy filtering, which becomes 60 based on the rms = 40° threshold. Methionine has the second largest number of minima of all 20 AAs.

AAs with a polar side chain, including serine (Ser), cysteine (Cys), and threonine (Thr), are grouped together into the fourth subgroup. Serine and cysteine have the simplest polar side chains, specifically $-CH_2OH$ and $-CH_2SH$, respectively. Together with Thr, these three AAs contain hydrogen bonds between OH/SH and the backbone, that is, OH/SH···NH or OH/SH···O═C. Table 1 shows that Ser and Cys both have the same number of minima, 142, after energy filtering. The number of minima of Ser and Cys is then reduced based on the rms = 40° threshold. The number of minima for Ser and Cys, which end up as 26 and 24, respectively, are very similar. As the side chain ($-C_\beta H(CH_3)OH$) of Thr is somewhat larger than that in Ser and Cys, the steric repulsion between the side chain and the backbone is unfavorable for its dihedral rotations. This is probably why Thr has a smaller number of minima (17) than either Ser or Cys.

The fifth group consists of aspartic acid (or aspartate, Asp), asparagine (Asn), glutamic acid (or glutamate, Glu), and glutamine (Gln). Aspartate and glutamate are classified as acidic AAs. These AAs can form hydrogen bonds between the side chain and the backbone. Aspartate and asparagine have quite a different number of minima, although their side chains are only slightly different (−OH in Asp and $-NH_2$ in Asn). After energy filtering, the number of minima of Asp and Asn is 100 and 28, respectively, and the number of minima then decreases to 36 and 12 respectively, based on the rms = 40° threshold. The difference of the number of minima between Asp and Asn illustrates that −OH is much more flexible than $-NH_2$. Glutamate and glutamine have an extra methylene group compared to Asp and Asn, respectively, and a higher number of minima as well. This shows again that, for the same atomic connectivity within the side chain, the larger the side chain the larger the number of minima.

The sixth subgroup includes tyrosine (Tyr), histidine (His), tryptophan (Trp), and phenylalanine (Phe). Of the 20 common AAs, only these four possess aromatic rings. There are many intramolecular interactions in these AAs, such as the interaction between the backbone and the ring plane. Tyrosine (which contains a phenolic side-chain group) and Phe (which contains a phenyl group) have nearly the same number of minima after energy filtering. The number of minima of Tyr is 67, which is slightly larger than the number for Phe, which is 61. This similarity in number of minima is due to the fact that both the size and the bonding pattern of their side chains are similar. Second, Tyr has 17 minima based on the rms = 40° threshold, while Phe has 30 minima. This large difference is due to the fact that the hydrogen bonds between the phenolic −OH and the peptidic N/O atoms in the backbone in Tyr make the phenol group less flexible (i.e., rotatable) compared with the phenyl group in Phe. As a result, Tyr has a smaller number of minima than Phe. Histidine (which contains an imidazole group) and tryptophan (with an indole group) both have more minima than Tyr but fewer minima than Phe. The number of minima of His is 2, and for Trp it is 27. This is because, on one hand, His and Trp possess hydrogen bonds between the side chain and the backbone. On the other hand, the size of the side chain is larger (particularly in Trp) and thus has a larger steric effect than the side chain in Phe.

Lysine (Lys) and arginine (Arg) are classified as the seventh subgroup. Of all the 20 natural AAs, geometrical study of Arg is the most challenging because of the large flexibility (i.e., rotational freedom) of its long and straight side chain. This side chain generates many energy minima. In addition, the more dihedral angles (n in eq 1) are active, the more minima are obtained. Of all 20 AAs, arginine has the largest number of dihedral angles according to the rms = 40° threshold and ends up with 61 minima. Intramolecular interactions are important factors in keeping the minima relatively stable. The pattern of the intramolecular interactions (i.e., the hydrogen bonds) in Arg is shown in Figure 4. Three types of hydrogen bonds are

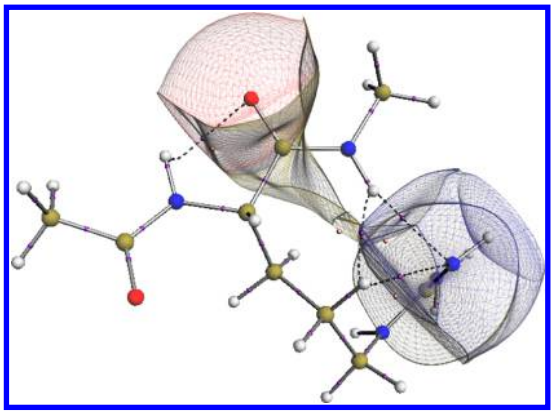


Figure 4. Pattern of atomic interaction lines in Arg with some topological atoms. Bond critical points are shown in purple, and ring critical points are in pink. Dashed lines represent nonbonded interactions.

dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

Table 2. Global Minimum Geometries for the 20 AAs[a]

AA	φ	ψ	χ_1	χ_2	χ_3	χ_4	χ_5	E_{total} (au)[b]	ΔE (kJ mol^{-1})[c]	dipole (au)[d]
Gly	179	−180						−453.84477	47	1.122
	−82	60						−456.50522	43	1.271
	−159	−172						−455.25204	46	1.248
Ala	−157	159	−178					−492.88467	59	1.005
	−154	157	−178					−495.81540	54	1.213
	−155	169	−179					−492.89940	59	1.165
Val	−86	88	−178					−570.95890	101	0.876
	−84	82	−177					−574.43362	86	0.882
	−81	85	−177					−570.97249	102	0.805
Thr	−116	12	53					−606.77930	94	1.754
	−110	11	50					−610.33815	81	1.682
	−128	13	53					−606.79860	95	1.532
Cys	−87	79	−172					−890.39436	73	1.513
	−159	172	−162					−894.02576	69	0.784
	−81	79	−173					−890.42851	72	1.358
Ser	−85	75	54					−567.74714	87	0.774
	−82	74	57					−571.03362	80	0.781
	−80	78	54					−567.76585	84	0.676
Asn	−86	62	56	118				−660.68333	70	0.119
	−81	59	51	102				−664.51696	61	0.354
	−85	61	52	107				−660.70598	68	0.297
Tyr	−157	155	−165	73				−797.29859	55	0.740
	−155	159	−167	70				−802.05756	51	1.004
	−162	152	−176	74				−797.32490	56	0.992
His	−86	66	50	70				−716.56229	80	1.032
	−82	69	49	68				−720.82673	77	1.117
	−84	70	47	69				−716.58558	79	1.062
Ile	−86	90	−56	172				−609.99406	66	0.845
	−84	84	−55	172				−613.74070	60	0.875
	−81	86	−54	175				−610.00727	67	0.792
Trp	−156	150	−168	−112				−853.21110	67	1.511
	−155	159	−165	−113				−858.39760	58	1.781
	−158	156	−170	−112				−853.23877	69	1.691
Leu	−86	84	−175	63				−609.99681	66	0.955
	−83	78	−170	67				−613.74351	63	0.925
	−82	81	−55	−179				−610.00952	68	0.934
Phe	−157	151	−169	72				−733.43904	64	0.961
	−84	72	−57	111				−726.83995	56	1.155
	−90	76	−52	111				−722.46043	60	1.198
Asp	−161	164	−160	178				−680.51003	87	0.854
	−161	166	−161	174				−684.37808	75	1.052
	−161	167	−162	176				−680.53698	86	0.983
Pro	−86	75	32	−38	29			−569.78721	24	1.322
	−83	70	31	−37	29			−573.21943	25	1.290
	−86	75	33	−41	33			−569.80317	24	1.106
Gln	−159	154	−175	54	−86			−699.71795	68	1.139
	−160	172	−106	−68	−9			−703.82452	68	1.246
	−148	173	−99	−68	−8			−699.74297	74	0.977
Glu	−156	151	−176	58	93			−719.54734	73	0.728
	−154	154	−177	56	95			−723.68704	65	1.000
	−159	154	−177	58	91			−719.57337	73	0.822
Met	−157	150	−175	52	56			−968.46852	65	0.326
	−154	152	−175	49	54			−972.64199	60	0.526
	−160	152	−176	50	52			−968.49876	67	0.439
Lys	−86	82	−171	179	−180	179		−665.02280	55	0.610
	−83	77	−167	180	−180	−180		−669.08752	51	0.836
	−81	80	−172	179	179	177		−665.03836	66	0.590
Arg	−161	−172	−141	−169	51	62	−5	−773.92832	90	2.284
	−142	−147	−92	−103	51	63	−139	−778.59222	83	1.457
	−133	−148	−91	−98	52	63	−139	−773.95095	101	1.299

[a]For each AA, the first, second, and third lines are the values optimized at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ levels, respectively. [b]The single-point energy of the global minimum of each of the 20 AAs calculated at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ levels. [c]The relative energy difference between the global minimum and the local minimum with the highest energy at each of the three levels. [d]The dipole moment of the global minimum of each of the 20 AAs at the three levels.

dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

clearly visible, including N···HC, N···HN, and O···HN. These types of hydrogen bonds have also been reported by Gutowski et al.,[34] be it in Arg without peptide bond capping (i.e., as a neutral AA, with COOH and NH_2 termini). There is also one hydrogen−hydrogen bond between the Cγ−H group and a N−H group in the peptide backbone. This important (and for a long time overlooked) type of bond is extensively discussed by Matta et al.[72] Lysine, which has a structure similar to that of Arg and also has a large side chain, has 40 minima, the third largest number of minima of all 20 AAs.

3.2. Dihedral Angles and the Energy of the Global Minimum of Each of the 20 AAs. Table 2 gives the values for φ, ψ, and χ_k (k = 0, 1, 2, ...) of the global minimum of each of the 20 AAs, the absolute energies, and the relative energy difference between the highest energy local minimum and the global minimum. These values are calculated at three levels of theory: HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ. The intramolecular interaction between a large or polar side chain and the backbone makes an AA with such a side chain more stable than an AA with only small substituents (including just a hydrogen). Figure 2 marks the position in the Ramachandran map of the global minima by red dots and the position of all other local minima by blue dots. Figure 1 and the corresponding geometric values in Table 2 show that the global minimum geometries of most AAs optimized at the HF/6-31G(d,p) and B3LYP/apc-1 levels are close to the geometries obtained at the MP2/cc-pVDZ level with few exceptions.

In Figure 2b, the geometry of the global minimum of Gly at the B3LYP/apc-1 level is very different from that at the HF/6-31G(d,p) and MP2/cc-pVDZ levels. However, this large difference is caused by a relatively small difference in energy between two minima. In terms of absolute energy difference, the two lowest energy minima of Gly differ by 1.1, 2.8, and 2.6 kJ mol^{-1} for the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ levels, respectively. At the B3LYP/apc-1 level the two lowest energy minima have swapped compared to the two other levels, causing the global minimum to be geometrically very different. Note that the energy range (difference between the lowest and the highest energy) for Gly is about 45 kJ mol^{-1} for any of the three levels, which is about 20 times larger than the energy difference between the two lowest energies.

The φ, ψ, and χ_k dihedral angles of the global minima of Arg and Phe at the HF/6-31(d,p) level are very different from the global minima calculated at the B3LYP/apc-1 and MP2/cc-pVDZ levels (see Figure 2b and Table 2). However, the dihedral angles of the B3LYP/apc-1 and MP2/cc-pVDZ global minima of these two AAs are close. In contrast, the global minimum of Cys optimized at the B3LYP/apc-1 level differs significantly from the global minima at the HF/6-31G(d,p) and MP2/cc-pVDZ levels. Moreover, the φ and ψ dihedral angles of the global minima of Leu and Gln do not change much at these three levels of theory, while their χ_k dihedral angles are very different. Of course, this cannot be seen in Figure 2, but it is clear from Table 2.

The influence of the orientation of the side chain on the geometry of the global minimum for each of the 20 AAs is investigated. It is interesting to compare the (single-point) energy difference between the global minimum and the other minima of each of the 20 AAs. The energy range (ΔE), which is the energy difference between the global minimum and the (local) minimum with the highest energy, is calculated at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ levels. The values are listed in Table 2 and shown in Figure 5.

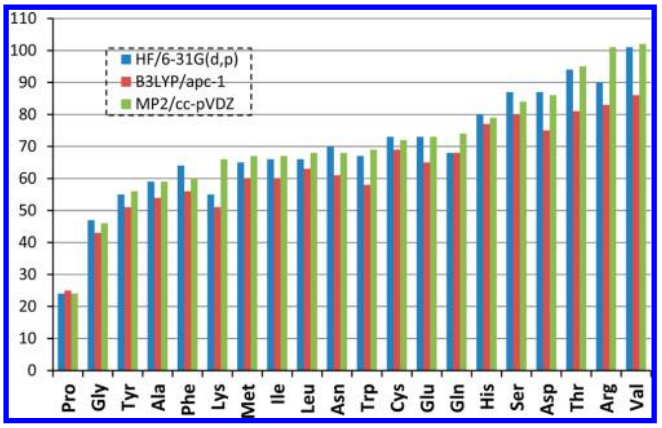


Figure 5. Energy range (in kJ mol^{-1}) of each of the 20 AAs between the global minimum and the local minimum with the highest energy at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ levels. Energies are ranked in increasing order according to the MP2 values.

In Figure 5 the order of the AAs is determined by the increasing energy range calculated at the MP2/cc-pVDZ level. Valine and arginine possess the largest energy range, while Pro has the smallest energy range, which is most likely related to its rigidity, given that it is the only AA whose side chain is covalently bonded to the backbone. Moreover, the energy range of each of the 20 AAs does not change much from one level of theory to another. However, the energy ranges are different enough to locally alter the ordering of the ranges. For example, Lys would precede Tyr based on the HF energy ranges.

The molecular dipole moment of each of the 20 global minima at the three levels is also listed in Table 2 and shown in Figure 6. In Figure 6 the order of the 20 AAs is determined

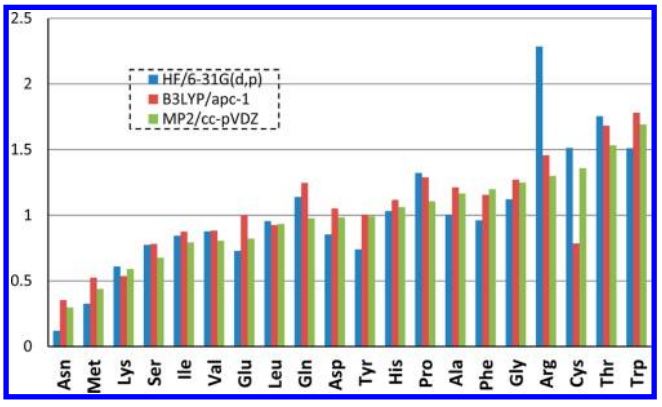


Figure 6. Dipole moment (au) of each of the 20 AAs at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ levels. The 20 AAs are ranked by the increasing dipole moments calculated at the MP2/cc-pVDZ level.

by increasing dipole moment, again calculated at the MP2/cc-pVDZ level. Tryptophan and threonine have the highest dipole moments, while asparagine and methionine have the lowest. Figure 6 shows that, in general, dipole moments broadly follow the same trend across the three levels of theory. However, there are two exceptions: Arg and Cys. Arginine's dipole moment calculated at the HF/6-31G(d,p) level is much larger than that obtained at the two other levels. This is because the HF/6-31G(d,p) global minimum geometry is very different from the B3LYP/apc-1 and MP2/cc-pVDZ global minimum geometries.

dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

Table 3. Charge (au) of the C_α Atom Calculated at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ Levels

	HF/6-31G(d,p)	B3LYP/apc-1				MP2/cc-pVDZ			
AAs	Q_{00}	Q_{00}	$k_{HF/B3LYP}$ [a]	pred.[b]	abs. err.[c]	Q_{00}	$k_{HF/MP2}$ [d]	pred[e]	abs. err.[f]
Gly	0.628	0.394	1.59	0.416	0.021	0.441	1.42	0.448	0.006
Pro	0.502	0.337	1.49	0.333	0.005	0.356	1.41	0.358	0.002
Ala	0.592	0.391	1.52	0.393	0.002	0.431	1.37	0.423	0.008
Val	0.529	0.359	1.47	0.351	0.008	0.379	1.39	0.377	0.002
Ile	0.528	0.359	1.47	0.350	0.009	0.378	1.40	0.377	0.001
Leu	0.534	0.362	1.47	0.354	0.008	0.384	1.39	0.381	0.002
Ser	0.548	0.364	1.50	0.363	0.002	0.386	1.42	0.391	0.005
Cys	0.559	0.372	1.50	0.371	0.001	0.397	1.41	0.399	0.003
Thr	0.544	0.356	1.53	0.360	0.004	0.377	1.44	0.388	0.011
Met	0.580	0.381	1.52	0.385	0.004	0.413	1.40	0.414	0.001
Asn	0.555	0.371	1.50	0.368	0.002	0.401	1.38	0.396	0.004
Asp	0.594	0.388	1.53	0.394	0.005	0.427	1.39	0.424	0.003
Gln	0.562	0.371	1.51	0.373	0.001	0.402	1.40	0.401	0.000
Glu	0.581	0.381	1.53	0.385	0.004	0.416	1.40	0.415	0.001
His	0.557	0.372	1.50	0.369	0.002	0.400	1.39	0.398	0.002
Phe	0.543	0.367	1.48	0.360	0.008	0.391	1.39	0.387	0.004
Try	0.586	0.386	1.52	0.389	0.003	0.419	1.40	0.418	0.001
Tyr	0.591	0.389	1.52	0.392	0.003	0.424	1.39	0.422	0.002
Lys	0.539	0.365	1.48	0.35	0.008	0.390	1.38	0.385	0.005
Arg	0.556	0.362	1.54	0.369	0.007	0.389	1.43	0.397	0.008
Avg[g]			1.51		0.005		1.40		0.004

[a]Ratio of the charges calculated at the HF/6-31G(d,p) level and the B3LYP/apc-1 level, $k_{HF/B3LYP} = Q_{00(HF/6\text{-}31G(d,p))}/Q_{00(B3LYP/apc\text{-}1)}$. [b]Charges predicted based on $Q_{00(HF/6\text{-}31G(d,p))}/k_{HF/B3LYP(Avg)}$. [c]Absolute errors between the predicted values and the true values ($Q_{00(Pred.)} - Q_{00}$) obtained at the B3LYP/apc-1 level. [d]Ratio of the charges calculated at the HF/6-31G(d,p) level and the MP2/cc-pVDZ level, $k_{HF/MP2} = Q_{00(HF/6\text{-}31G(d,p))}/Q_{00(MP2/cc\text{-}pVDZ)}$. [e]Charges predicted based on $Q_{00(HF/6\text{-}31G(d,p))}/k_{HF/MP2(Avg)}$. [f]Absolute errors between the predicted values and the true values ($Q_{00(Pred.)} - Q_{00}$) obtained at the MP2/cc-pVDZ level. [g]Average values of the ratios.

Similarly, the dipole moment obtained at the B3LYP/apc-1 level stands out because its global minimum geometry at this level is very different to that obtained at the HF/6-31G(d,p) and MP2/cc-pVDZ levels.

It is known that HF dipole moments are larger[73] than those obtained at a correlated level of theory when calculated at exactly the same molecular geometry. To investigate this question, extra calculations were performed but are not shown in Figure 6. The dipole moments of the minimum energy geometries obtained at the MP2/cc-pVDZ level are now calculated at the HF/6-31G(d,p) and B3LYP/apc-1 levels for the following four AAs: Gly, Val, Ala, and Ser. Hartree−Fock dipole moments are indeed the largest of the three levels of theory for Ala and Ser. However, this is not true for Gly and Val. Also, if the dipole moment of Ala is evaluated at the HF/6-31G(d,p) level but at the HF/6-31G(d,p) geometry then the HF dipole moment is the smallest (see Figure 6). Hence, the actual geometry is another factor to keep in mind when assessing the relative magnitude of dipole moments.

3.3. Atomic Multipole Moments at Different Levels of Theory. In order to investigate the difference between the atomic multipole moments at different levels of theory, the geometry of the global minimum of all 20 AAs optimized at the MP2/cc-pVDZ level is taken as the reference. Subsequently, the energies of the 20 global minimum geometries are calculated at the B3LYP/apc-1 and HF/6-31G(d,p) levels. In all our calculations, the "nosymm" keyword in GAUSSIAN09 is added to prevent the program from translating and rotating the molecule into the center-of-mass frame. The multipole moments of all five atoms in the [−NH−C_α−C(=O)−] common fragment in each global minimum are calculated according to the wave functions obtained at these three levels of theory. C_α is taken as an example to show the difference between the multipole moments at these three levels. The multipole moments of C_α are listed in Tables 3 (charge), 4 (dipole moment), and 5 (quadrupole moment) and shown in Figure 7. The dipole and quadrupole moment are calculated using $Q_{dipole} = (Q_{10}^2 + Q_{11c}^2 + Q_{11s}^2)^{1/2}$ and $Q_{quadrupole} = (Q_{20}^2 = Q_{21c}^2 + Q_{21s}^2 + Q_{22c}^2 + Q_{22s}^2)^{1/2}$, where real components of the multipole moments were used instead of the original Q_{lm} values (which can be complex).

Figure 7 represents the charge, dipole, and quadrupole moments of C_α in all of the 20 AAs. The three plots in Figure 7 show that the values at the B3LYP/apc-1 level are always the lowest, followed by MP2/cc-pVDZ in the middle and topped by HF/6-31G(d,p). This is due to the fact that HF does not include electron correlation and thereby always exaggerates the polarity of bonds. In other words, negatively charged atoms are more negative when calculated at the HF level than at a level of theory that includes electron correlation (i.e., B3LYP and MP2). The DFT/B3LYP method takes into account electron correlation, which diminishes the absolute value of an atomic charge.[74] However, for the polar molecules that we consider here, the DFT/B3LYP method overestimates the electron correlation, i.e., the absolute values of the atomic charges are lower than those obtained with the MP2 method. Moreover, the three curves in each panel (a, b, and c) of Figure 7 show very similar behavior, although they do not exactly parallel each other. The ratios between the moments calculated at the HF/6-31G(d,p) and B3LYP/apc-1 levels as well as between the HF/6-31G(d,p) and MP2/cc-pVDZ levels are tabulated in Tables 3−5. These ratios remain remarkably constant, which is consistent with the curves in Figure 7 being almost parallel.

dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

Table 4. Dipole Moments (au) of the C_α Atom Calculated at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ Levels

	HF/6-31G(d,p)	B3LYP/apc-1				MP2/cc-pVDZ			
AAs	Q_{dipole}	Q_{dipole}	$k_{HF/B3LYP}$[a]	pred.[b]	abs. err.[c]	Q_{dipole}	$k_{HF/MP2}$[d]	pred.[e]	abs. err.[f]
Gly	0.564	0.453	1.24	0.434	0.019	0.483	1.17	0.467	0.020
Pro	0.574	0.443	1.29	0.442	0.001	0.473	1.21	0.475	0.003
Ala	0.597	0.460	1.30	0.460	0.000	0.493	1.21	0.495	0.002
Val	0.604	0.459	1.31	0.466	0.006	0.494	1.22	0.501	0.007
Ile	0.605	0.460	1.31	0.466	0.006	0.495	1.22	0.501	0.007
Leu	0.598	0.459	1.30	0.461	0.002	0.491	1.22	0.496	0.005
Ser	0.598	0.458	1.30	0.461	0.003	0.497	1.20	0.496	0.001
Cys	0.569	0.441	1.29	0.439	0.003	0.475	1.20	0.472	0.003
Thr	0.567	0.431	1.31	0.437	0.006	0.468	1.21	0.470	0.002
Met	0.602	0.465	1.29	0.464	0.001	0.501	1.20	0.499	0.002
Asn	0.554	0.425	1.30	0.427	0.002	0.462	1.20	0.459	0.002
Asp	0.582	0.454	1.28	0.448	0.005	0.492	1.18	0.482	0.009
Gln	0.588	0.452	1.30	0.453	0.002	0.487	1.21	0.488	0.001
Glu	0.604	0.470	1.28	0.465	0.005	0.506	1.19	0.500	0.006
His	0.587	0.448	1.31	0.453	0.005	0.482	1.22	0.487	0.005
Phe	0.587	0.451	1.30	0.452	0.001	0.484	1.21	0.487	0.002
Try	0.606	0.465	1.30	0.467	0.002	0.501	1.21	0.502	0.002
Tyr	0.606	0.465	1.30	0.467	0.002	0.501	1.21	0.503	0.002
Lys	0.598	0.457	1.31	0.461	0.004	0.489	1.22	0.496	0.007
Arg	0.608	0.476	1.28	0.468	0.008	0.510	1.19	0.504	0.006
Avg.[g]			1.30		0.004		1.21		0.004

[a]Ratio of the dipole moments calculated at the HF/6-31G(d,p) level and the B3LYP/apc-1 level, $k_{HF/B3LYP} = Q_{dipole(HF/6\text{-}31G(d,p))}/Q_{dipole(B3LYP/apc\text{-}1)}$. [b]Dipole moments predicted based on $Q_{dipole(HF/6\text{-}31G(d,p))}/k_{HF/B3LYP(Avg)}$. [c]Absolute errors between the predicted values and the true values ($Q_{dipole(Pred.)} - Q_{dipole}$) obtained at the B3LYP/apc-1 level. [d]Ratio of the dipole moments calculated at the HF/6-31G(d,p) level and the MP2/cc-pVDZ level, $k_{HF/MP2} = Q_{dipole(HF/6\text{-}31G(d,p))}/Q_{dipole(MP2/cc\text{-}pVDZ)}$. [e]Dipole moments predicted based on $Q_{dipole(HF/6\text{-}31G(d,p))}/k_{HF/MP2(Avg)}$. [f]Absolute errors between the predicted values and the true values ($Q_{dipole(Pred.)} - Q_{dipole}$) obtained at the MP2/cc-pVDZ level. [g]Average values of the ratios.

Table 5. Quadrupole Moments (au) of the C_α Atom Calculated at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ Levels

	HF/6-31G(d,p)	B3LYP/apc-1				MP2/cc-pVDZ			
AAs	$Q_{quadrupole}$	$Q_{quadrupole}$	$k_{HF/B3LYP}$[a]	pred.[b]	abs. err.[c]	$Q_{quadrupole}$	$k_{HF/MP2}$[d]	pred.[e]	abs. err.[f]
Gly	0.521	0.459	1.13	0.433	0.025	0.500	1.04	0.472	0.028
Pro	0.401	0.312	1.28	0.334	0.021	0.352	1.14	0.364	0.012
Ala	0.510	0.418	1.22	0.424	0.006	0.458	1.11	0.462	0.004
Val	0.477	0.386	1.23	0.397	0.010	0.415	1.15	0.433	0.018
Ile	0.481	0.389	1.23	0.400	0.011	0.417	1.15	0.436	0.019
Leu	0.461	0.377	1.22	0.383	0.006	0.410	1.12	0.418	0.008
Ser	0.439	0.357	1.23	0.365	0.009	0.397	1.11	0.398	0.001
Cys	0.466	0.400	1.16	0.387	0.013	0.433	1.08	0.422	0.010
Thr	0.589	0.516	1.14	0.490	0.026	0.562	1.05	0.534	0.028
Met	0.512	0.424	1.20	0.425	0.001	0.459	1.11	0.464	0.005
Asn	0.476	0.410	1.16	0.396	0.014	0.451	1.05	0.431	0.020
Asp	0.505	0.431	1.17	0.420	0.011	0.470	1.07	0.458	0.012
Gln	0.552	0.464	1.19	0.459	0.005	0.508	1.09	0.501	0.007
Glu	0.487	0.403	1.21	0.405	0.002	0.438	1.11	0.442	0.004
His	0.467	0.390	1.19	0.388	0.002	0.428	1.09	0.423	0.005
Phe	0.454	0.374	1.21	0.377	0.004	0.410	1.10	0.411	0.001
Try	0.548	0.454	1.21	0.456	0.001	0.494	1.11	0.497	0.003
Tyr	0.529	0.438	1.21	0.440	0.002	0.474	1.12	0.480	0.006
Lys	0.473	0.389	1.22	0.394	0.004	0.422	1.12	0.429	0.007
Arg	0.506	0.419	1.21	0.421	0.001	0.456	1.11	0.459	0.003
Avg.[g]			1.20		0.009		1.10		0.010

[a]Ratio of the quadrupole moments calculated at the HF/6-31G(d,p) level and the B3LYP/apc-1 level, $k_{HF/B3LYP} = Q_{quadrupole(HF/6\text{-}31G(d,p))}/Q_{quadrupole(B3LYP/apc\text{-}1)}$. [b]Quadrupole moments predicted based on $Q_{quadrupole(HF/6\text{-}31G(d,p))}/k_{HF/B3LYP(Avg)}$. [c]Absolute errors between the predicted values and the true values ($Q_{quadrupole(Pred.)} - Q_{quadrupole}$) obtained at the B3LYP/apc-1 level. [d]Ratio of the quadrupole moments calculated at the HF/6-31G(d,p) level and the MP2/cc-pVDZ level, $k_{HF/MP2} = Q_{quadrupole(HF/6\text{-}31G(d,p))}/Q_{quadrupole(MP2/cc\text{-}pVDZ)}$. [e]Quadrupole moments predicted based on $Q_{quadrupole(HF/6\text{-}31G(d,p))}/k_{HF/MP2(Avg)}$. [f]Absolute errors between the predicted values and the true values ($Q_{quadrupole(Pred.)} - Q_{quadrupole}$) obtained at the MP2/cc-pVDZ level. [g]Average values of the ratios.

dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

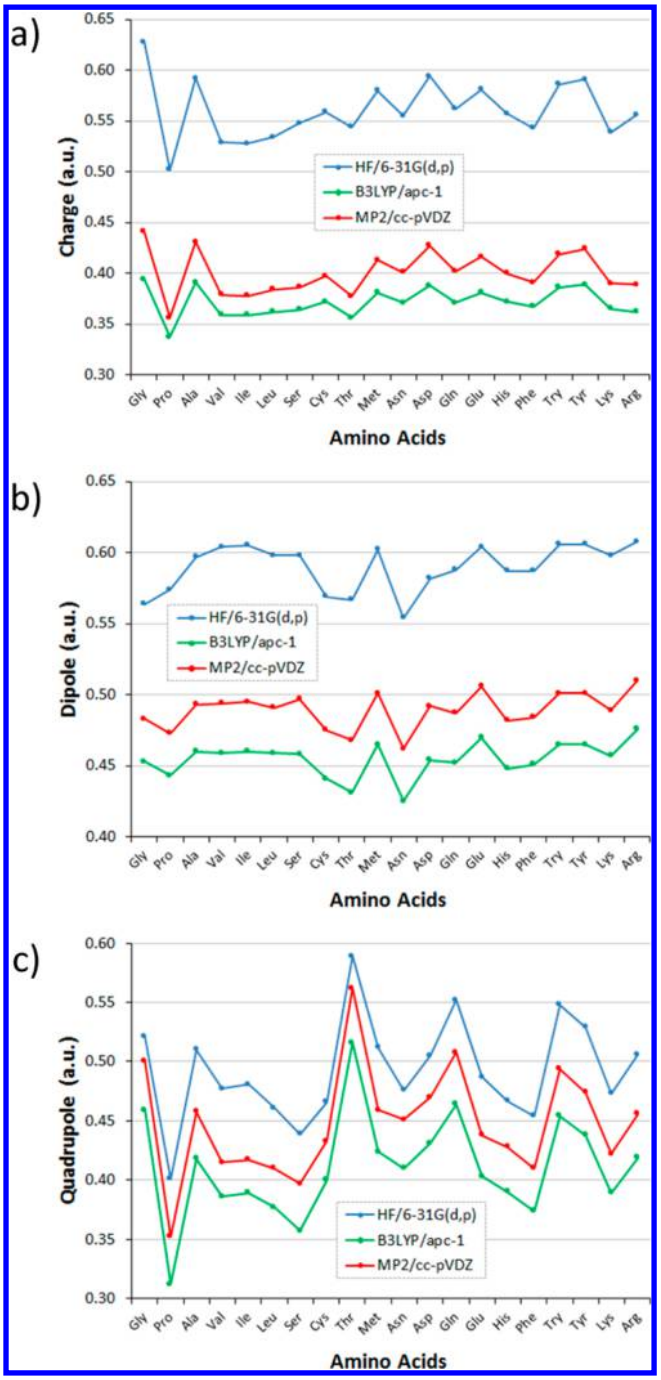


Figure 7. Magnitudes of the multipole moments (au) of C_α in all 20 global energy minima at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ levels: (a) charge, (b) dipole moment, and (c) quadrupole moment.

This near parallelism is beneficial for estimation and even prediction of multipole moments of the more expensive B3LYP and MP2 methods, from the computationally cheaper HF method. For that purpose we use the ratios $Q_{HF/6\text{-}31G(d,p)}/k_{HF/B3LYP(Avg.)}$ and $Q_{HF/6\text{-}31G(d,p)}/k_{HF/MP2(Avg.)}$, where $Q_{HF/6\text{-}31G(d,p)}$ represents the moments calculated at HF/6-31G(d,p) while $k_{HF/B3LYP(Avg.)}$ and $k_{HF/MP2(Avg.)}$ are the respective ratios, each averaged over the 20 AAs. From Tables 3−5 we can see that the predicted charges are excellent. Indeed, the absolute average error between the predicted and the original charges is only 0.005 au at the B3LYP/apc-1 level and 0.004 au at the MP2/cc-pVDZ level. For the dipole moment, the absolute average errors are both 0.004 au at these two levels of theory. The absolute average errors at the two levels of theory are 0.009 au and 0.010 au, respectively. The absolute average errors of quadrupole moments are higher than the corresponding values for the charge and dipole moments, which means that the quadrupole moments are more easily influenced by the geometries than the charge and dipole moments.

The average values of the ratios and the absolute average errors between the predicted moments and the true values for the remaining four atoms in the common fragment, other than C_α, are listed in Table 6. This table shows that the average errors for these atoms are again very small. Once more, the absolute average errors of quadrupole moments are higher than the values of charge and dipole moments for each of the four atoms. In summary, in order to reduce the computational cost, the multipole moments at a higher level of theory can be predicted quite accurately, based on the values calculated at the lower level of HF/6-31G(d,p).

Table 6. Average Values of Ratios and Errors (au) for the Remaining Four Atoms in the Common Fragment Calculated at the HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ Levels

atoms	multipole moments	$k_{HF/B3LYP}$ [a]	abs. err [b]	$k_{HF/MP2}$ [c]	abs. err. [d]
C	charge	1.27	0.004	1.21	0.003
	dipole	1.17	0.003	1.17	0.003
	quadrupole	0.78	0.004	0.72	0.006
H	charge	1.08	0.005	1.00	0.008
	dipole	0.89	0.002	0.85	0.001
	quadrupole	0.98	0.009	1.44	0.011
O	charge	1.15	0.004	1.18	0.003
	dipole	1.60	0.004	1.35	0.002
	quadrupole	0.60	0.014	0.87	0.021
N	charge	1.36	0.010	1.22	0.011
	dipole	1.40	0.020	1.27	0.013
	quadrupole	0.82	0.037	0.96	0.022

[a] Average values of ratios at the HF/6-31G(d,p) and B3LYP/apc-1 levels, $k_{HF/B3LYP} = Q_{moments(HF/6\text{-}31G(d,p))}/Q_{moments(B3LYP/apc\text{-}1)}$. [b] Absolute average errors between the predicted values and the true values at the B3LYP/apc-1 level ($Q_{moments(Pred.)} - Q_{moments}$). [c] Average values of ratios at the HF/6-31G(d,p) and MP2/cc-pVDZ levels $k_{HF/MP2} = Q_{moments(HF/6\text{-}31G(d,p))}/Q_{moments(MP2/cc\text{-}pVDZ)}$. [d] Absolute average errors between the predicted values and the true values at the MP2/cc-pVDZ level ($Q_{moments(Pred.)} - Q_{moments}$).

3.4. Transferability of Properties between Amino Acids. It has been observed before[54] that the bond lengths and valence angles of leucine are not significantly altered as various dihedral angles are sampled. This conservation of geometry in turn ensures conservation of the electron density of corresponding atoms in different rotamers. Hence, the bond, atomic, and group properties defined by theory are equally insensitive to conformational changes. Eventually, the molecular properties and energy of a protein can in principle be predicted from small fragments. Moreover, the molecular properties depend on the electron density, $\rho(\mathbf{r})$, which is subsequently dependent on the geometries of the molecules. Consequently, a high degree of similarity of the molecular geometries is the most important factor to obtain a high degree of transferability of the molecular properties between different systems.

It is clear that 18 out of the 20 AAs (that is, excluding Gly and Pro) possess the common fragment [−NH−C_α($H_\alpha C_\beta$)−C′(=O)−],

dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

Table 7. Fifteen Geometrical Parameters of the Common Fragment $[-NH-C_\alpha(H_\alpha C_\beta)-C'(=O)-]$ in the Global Minima of the 18 AAs, Geometry Optimized at the MP2/cc-pVDZ Level[a]

AAs	$R_{C\alpha N}$	$R_{C\alpha C'}$	$\theta_{NC\alpha C'}$	$R_{C\alpha H\alpha}$	$\theta_{C\alpha H\alpha}$	$\varphi_{C\alpha H\alpha}$	$R_{C\alpha H}$	$\theta_{C\alpha H}$	$\varphi_{C\alpha H}$	$R_{C\alpha O}$	$\theta_{C\alpha O}$	$\varphi_{C\alpha O}$	$R_{C\alpha C\beta}$	$\theta_{C\alpha C\beta}$	$\varphi_{C\alpha C\beta}$
Ala	1.445	1.529	107	1.106	146	−125	2.096	86	26	2.418	84	82	1.535	38	−127
Asp	1.445	1.540	106	1.106	145	−126	2.090	87	26	2.418	84	81	1.540	37	−124
Gln	1.450	1.540	107	1.106	148	−127	2.074	87	26	2.424	87	81	1.550	40	−127
Glu	1.445	1.535	107	1.101	146	−124	2.101	82	25	2.408	78	83	1.545	37	−125
Arg	1.450	1.524	108	1.101	147	−123	2.096	78	23	2.413	75	87	1.550	37	−127
Met	1.445	1.529	107	1.101	146	−124	2.101	81	24	2.408	77	84	1.540	37	−126
Tyr	1.445	1.529	107	1.106	145	−124	2.101	82	25	2.408	78	84	1.550	37	−128
Trp	1.445	1.524	108	1.101	146	−124	2.101	83	25	2.408	80	84	1.550	37	−127
Val	1.461	1.540	108	1.101	148	−121	2.133	65	5	2.429	65	110	1.529	39	−125
Ile	1.461	1.540	108	1.101	149	−121	2.127	65	5	2.429	65	109	1.529	39	−125
Leu	1.461	1.540	109	1.101	149	−122	2.133	65	3	2.434	65	113	1.524	38	−124
Ser	1.455	1.540	110	1.101	146	−123	2.106	64	0	2.429	65	116	1.529	35	−123
Cys	1.461	1.540	110	1.101	147	−122	2.133	65	4	2.429	65	115	1.524	37	−123
Asn	1.461	1.545	113	1.106	148	−124	2.117	65	−9	2.440	68	127	1.535	38	−122
Lys	1.461	1.540	109	1.101	148	−122	2.133	65	2	2.429	65	113	1.524	38	−124
His	1.455	1.545	111	1.106	148	−125	2.122	65	−5	2.450	67	121	1.540	39	−121
Phe	1.445	1.535	107	1.106	148	−123	2.127	67	−11	2.429	65	117	1.529	37	−124
Thr	1.455	1.540	113	1.101	148	−123	2.085	64	3	2.408	85	139	1.529	38	−124
Avg.[b]								**83**[c]	**25**[c]	2.424	**80**[c]	**83.**[c]			
	1.455	1.535	109	1.101	147	−124	2.111	65[d]	3[d]		65[d]	116[d]	1.535	38	−125
sd[e]								**3**[f]	**1**[f]	0.011	**4**[f]	**2**[f]			
	0.005	0.005	2	0.005	1	2	0.016	1[g]	6[g]		1[g]	5[g]	0.011	1	2

[a]Bond lengths are in Angstroms; polar angles are in degrees. [b]Average values of each of the 15 geometrical parameters. [c]Average values of $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $\theta_{C\alpha O}$, and $\varphi_{C\alpha O}$ in the first group. [d]Average values of $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $\theta_{C\alpha O}$, and $\varphi_{C\alpha O}$ in the second group. [e]Standard deviations of each of the 15 geometrical parameters. [f]Standard deviations of $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $\theta_{C\alpha O}$, and $\varphi_{C\alpha O}$ in the first group. [g]Standard deviations of $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $\theta_{C\alpha O}$, and $\varphi_{C\alpha O}$ in the second group.

which consists of seven atoms. The global minima of these 18 AAs calculated at the MP2/cc-pVDZ level are utilized in the following calculation. The geometries of the 18 global minima are first converted from the global frame to the so-called atomic local frame (ALF).[74,75] The C_α atom is taken as the origin of the ALF, the $C_\alpha N$ bond determines the x axis, while the $C_\alpha C'$ bond determines the xy plane together with N. It is then straightforward to construct the y axis, which is orthogonal to the x axis. The z axis is then made to be orthogonal to the xy plane, forming a right-handed axis system. Atoms H, O, H_α, and C_β are non-ALF atoms. There are in total $3 \times 7 - 6 = 15$ geometrical parameters describing each atomic position in this common fragment, which are $R_{C\alpha N}$, $R_{C\alpha C'}$, $\theta_{NC\alpha C'}$, $R_{C\alpha H\alpha}$, $\theta_{C\alpha H\alpha}$, $\varphi_{C\alpha H\alpha}$, $R_{C\alpha H}$, $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $R_{C\alpha O}$, $\theta_{C\alpha O}$, $\varphi_{C\alpha O}$, $R_{C\alpha C\beta}$, $\theta_{C\alpha C\beta}$, and $\varphi_{C\alpha C\beta}$, where R are distances and θ and φ angles. The values of the 15 geometrical parameters for each AA are shown in Table 7 and Figure 8.

Figure 8a shows that all the bond lengths ($R_{C\alpha N}$, $R_{C\alpha C'}$, $R_{C\alpha H\alpha}$, $R_{C\alpha H}$, $R_{C\alpha O}$, and $R_{C\alpha C\beta}$) in the 18 AAs lie on an almost straight curve, which means that all bond lengths in the global minima of the 18 AAs have a high degree of transferability. For example, in Table 7 the average bond lengths of $R_{C\alpha N}$ and $R_{C\alpha C'}$ are 1.45 and 1.54 Å, respectively, and their corresponding standard deviations are both less than 0.01 Å. The average bond angle of $\theta_{NC\alpha C'}$ in Table 7 is 108.9°, and the standard deviation is 2.29°. The other geometrical parameters (with the exception of $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $\theta_{C\alpha O}$, and $\varphi_{C\alpha O}$) exhibit high transferability as well, as evidenced by their small standard deviations.

On the basis of the values of the $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $\theta_{C\alpha O}$, and $\varphi_{C\alpha O}$ polar angles, these 18 AAs are classified into three groups. The first group includes eight AAs, which are Ala, Asp, Gln, Glu, Arg, Met, Tyr, and Trp. The angles $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, and $\theta_{C\alpha O}$ in the

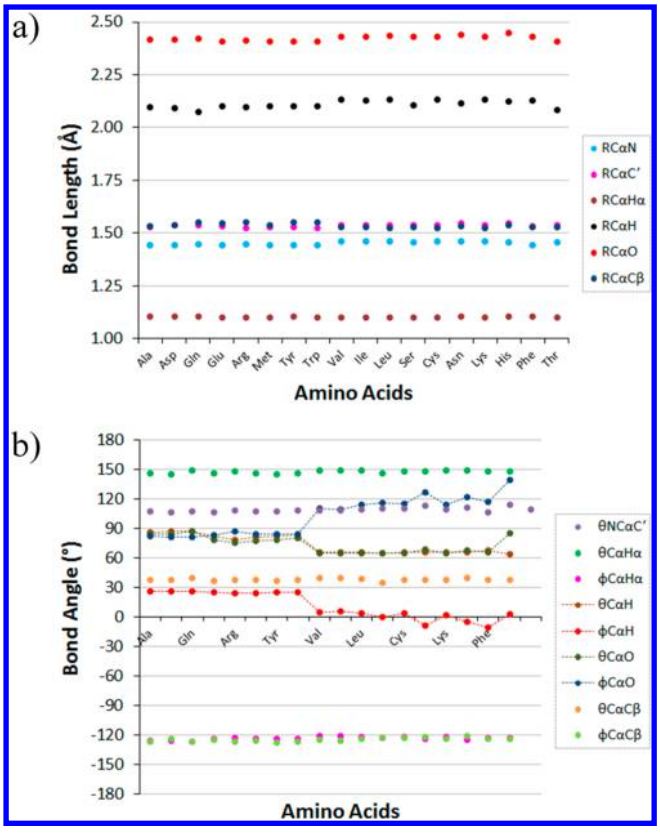


Figure 8. Variation of 15 geometrical parameters of the common fragment $[-NH-C_\alpha(H_\alpha C_\beta)-C'(=O)-]$ across the global energy minima of the 18 amino acids: (a) bond lengths between the C_α atom and the other six atoms in the fragment; (b) three angles spanned by C_α.

dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

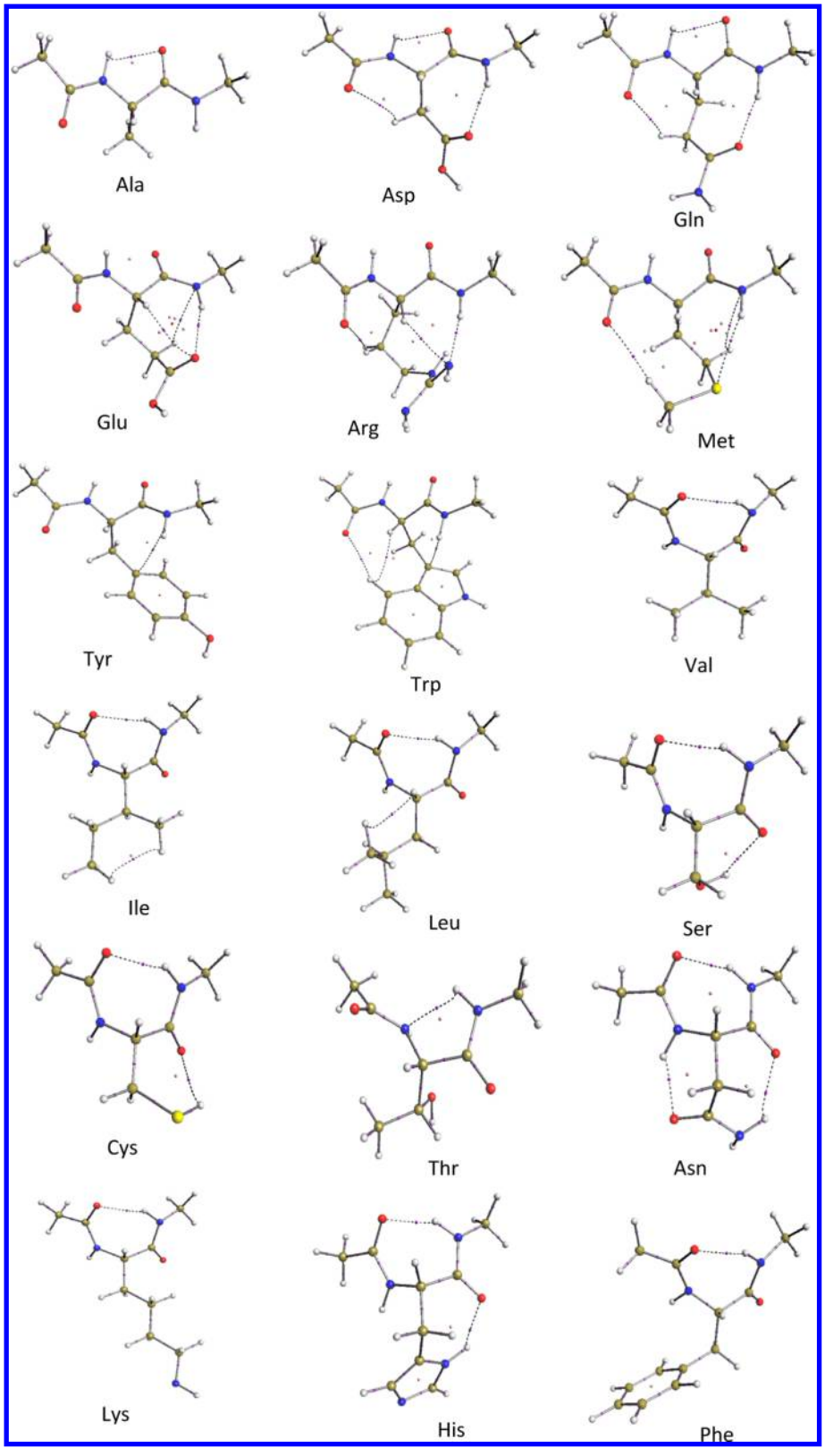


Figure 9. Intramolecular interaction patterns in the global minimum of each of the 18 AAs (all natural amino acids excluding Pro and Gly). The level of theory is MP2/cc-pVDZ level.

first group are higher than the values in the second group, while the values of $\varphi_{C\alpha O}$ are lower than the values in the second group. The different values of the four angles are due to the different orientation of the atoms H and O in the common fragment. The geometries of the 18 global minima and their intramolecular interactions are shown in Figure 9, with the AAs of the first group at the top of the figure. Atoms H and O in the first group are in a region at the opposite side of the side chains, while in the second group they lie in a region at the same side of the side chains.

dx.doi.org/10.1021/jp503460m | J. Phys. Chem. A 2014, 118, 7876−7891

The angles $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $\theta_{C\alpha O}$, and $\varphi_{C\alpha O}$ in the first group are close but differ from the values in the second group. Taking $\varphi_{C\alpha H}$ as an example, the average value of $\varphi_{C\alpha H}$ in the first group (of eight AAs), as shown in bold in Table 7, is 24.6°. The average value of $\varphi_{C\alpha H}$ in the second group is 5.7°. The standard deviations of $\varphi_{C\alpha H}$ in each group are both 1.1°.

The hydrogen bonds in the first group (bold in Table 7) can be formed between H and O in the backbone (five-membered ring) or between the side chain and the H or O atom in the capping groups of the backbone (e.g., Asp in Figure 9). In this first group, all AAs possess very strong hydrogen bonds between the side chain and the backbone, in particular, Gln, Glu, Arg, and Met, as can be seen in Figure 9. Both Glu and Met have a cage critical point. Met has the most complicated ring pattern because there is one cage critical point and six ring critical points involving three hydrogen bonds. Finally, Thr is classified into the third group.

In the second group (not bold in Table 7), a hydrogen bond is always (except for Thr) formed between the O and the H atoms in the capping group of the backbone (e.g., Ser in Figure 9). Besides this hydrogen bond, another one or two hydrogen bonds may be formed between the side chain and the H or O atom in the common fragment. All AAs in this group have a seven-membered ring in the backbone. Only Ser, Cys, Asn, and His have hydrogen bonds between the side chain and the backbone in this second group, and these hydrogen bonds are weak. Moreover, Ile and Leu have weak hydrogen−hydrogen bonds.[72]

Threonine is classified into a third group as the $\theta_{C\alpha O}$ and $\varphi_{C\alpha O}$ values are different from the values in the second group, though the remaining geometrical parameters are all close to those in the second group. This is due to the fact that Thr has a five-membered ring in the backbone, which is different from the seven-membered ring in the AAs in the second group.

Generally speaking, the bond lengths and angles (excluding $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $\theta_{C\alpha O}$, and $\varphi_{C\alpha O}$) in the common fragment have a high degree of transferability among the 18 AAs. The values of $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $\theta_{C\alpha O}$, and $\varphi_{C\alpha O}$ are also close to each other in each of the two groups. Because a protein backbone consists of the common fragment, the transferability degree of this fragment is definitely important in the design of QCTFF.

4. CONCLUSIONS

The present study analyzes the energy minima of each of the 20 AAs. All minima are optimized at three levels of theory: HF/6-31G(d,p), B3LYP/apc-1, and MP2/cc-pVDZ. The number of minima of each AA is influenced by the size and flexibility of the side chain. In general, the larger and more flexible the side chain, the more minima an AA has. Moreover, most positions (in the Ramachandran map) of all local minima change little going from one level of theory to another.

The multipole moments of atoms in the common fragment [−NH−C_α−C(=O)−] in each global minimum of the 20 AAs are calculated at the same three levels of theory listed earlier. The multipole moments keep the same trend at these three levels. The multipole moments at higher level of theory (specifically B3LYP/apc-1 and MP2/cc-pVDZ) can be predicted from the values calculated at a lower level, HF/6-31G(d,p). The average absolute errors calculated between the predicted and the true values are all very small.

The transferability of the properties of the common fragment [−NH−C_α($H_\alpha$$C_\beta$)−C′(=O)−] between different AAs (except Gly and Pro) has also been investigated. These 18 AAs are classified into two groups on the basis of the orientation of the H and O atoms. There are in total 15 geometrical parameters in the common fragment: $R_{C\alpha N}$, $R_{C\alpha C'}$, $\theta_{NC\alpha C'}$, $R_{C\alpha H\alpha}$, $\theta_{C\alpha H\alpha}$, $\varphi_{C\alpha H\alpha}$, $R_{C\alpha H}$, $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $R_{C\alpha O}$, $\theta_{C\alpha O}$, $\varphi_{C\alpha O}$, $R_{C\alpha C\beta}$, $\theta_{C\alpha C\beta}$, and $\varphi_{C\alpha C\beta}$. All bond lengths are highly transferable between different AAs, and the standard deviation values are very small. The angles are also close to each other in each of the two groups, but the $\theta_{C\alpha H}$, $\varphi_{C\alpha H}$, $\theta_{C\alpha O}$, and $\varphi_{C\alpha O}$ values are different between these two groups.

■ ASSOCIATED CONTENT

Supporting Information

List of Cartesian coordinates and energies of all structures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: pla@manchester.ac.uk.

Present Addresses

[§]School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China.

[∥]Department of Chemistry, University of Southern California, Los Angeles, CA 90089, United States.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the China Scholarship Council (CSC) for financial support. The authors thank the High Throughput Computing (CONDOR) facility of the Faculty of the Engineering and Physical Science (EPS) at the University of Manchester and are thankful for support from the Danish Center for Scientific Computation and the Danish Natural Science Research Council.

■ REFERENCES

(1) Cabezas, C.; Varela, M.; Cortijo, V.; Jimenez, A. I.; Pena, I.; Daly, A. M.; Lopez, J. C.; Cativiela, C.; Alonso, J. C. The alanine model dipeptide Ac-Ala-NH2 exists as a mixture of Ceq 7 and C5 conformers†. *Phys. Chem. Chem. Phys.* **2013**, *15*, 2580−2585.

(2) Shimozono, Y.; Yamada, K.; Ischuichi, S.; Tsukiyama, K.; Fujii, M. Revised conformational assigments and conformational evolution of tyrosine by laser desorpton supersonic jet laser spectroscopy. *Phys. Chem. Chem. Phys.* **2013**, *15*, 5163−5175.

(3) Gronert, S.; O'Hair, R. A. J. Ab Initio Studies of Amino Acid Conformations. 1. The Conformers of Alanine, Serine, and Cysteine. *J. Am. Chem. Soc.* **1995**, *117*, 2071−2081.

(4) Berg, M. A.; Salpietro, S. J.; Perczel, A.; Farkas, O.; Csizmadia, I. G. Side-chain conformational analysis of N-formyl-l-asparaginamide and N-acetyl-l-asparagine N-methylamide in their gl backbone conformation. *J. Mol. Struct.: THEOCHEM* **2000**, *504*, 127−139.

(5) Baldoni, H. A.; Zamarbide, G. N.; Enriz, R. D.; Jauregui, E. A.; Farkas, Ö.; Perczel, A.; Salpietro, S. J.; Csizmadia, I. G. Peptide models XXIX. cis−trans Isomerism of peptide bonds: ab initio study on small peptide model compound; the 3D-Ramachandran map of formylglycinamide. *J. Mol. Struct.: THEOCHEM* **2000**, *500*, 97−111.

(6) Viviani, W.; Rivail, J.-L.; Perczel, A.; Csizmadia, I. G. Peptide Models. 3 Conformational potential energy hypersurface of formyl-L-valinamide. *J. Am. Chem. Soc.* **1993**, *115*, 8321−8329.

(7) Hernandez, B.; Pflueger, F.; Nsangou, M.; Ghomi, M. Vibrational Analysis of Amino Acids and Short Peptides in Hydrated Media. IV. Amino Acids with Hydrophobic Side Chains: L-Alanine, L-Valine, and L-Isoleucine. *J.Phys.Chem.B* **2009**, *113*, 3169−3178.

(8) Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J.-L.; Marcoccia, J.-F.; Csizmadi, I. G. Peptide Models. 1. Topology of Selected Peptide Conformational Potential Energy Surfaces (Glycine and Alanine Derivatives). *J. Am. Chem. Soc.* **1991**, *113*, 6256−6265.
(9) Stepanian, S. G.; Reva, I. D.; Radchenko, E. D.; Adamowicz, L. Conformational Behavior of α-Alanine. Matrix-Isolation Infrared and Theoretical DFT and ab Initio Study. *J. Phys. Chem. A* **1998**, *102*, 4623−4629.
(10) Perczel, A.; Farkas, O.; Jákli, L.; Topol, L. A.; Császár, I. G. Peptide Models. XXXIII. Extrapolation of Low-Level Hartree-Fock Data of Peptide Conformation to Large Basis Set SCF, MP2, DFT, and CCSD(T) Results. The Ramachandran Surface of Alanine Dipeptide Computed at Various Levels of Theory. *J. Comput. Chem.* **2003**, *24*, 1026−1042.
(11) Sahai, M. A.; Fejer, S. N.; Viskolcz, B.; Pai, E. F.; Csizmadia, I. G. First-Principle Computational Study on the Full Conformational Space of l-Threonine Diamide, the Energetic Stability of Cis and Trans Isomers. *J. Phys. Chem. A* **2006**, *110*, 11527−11536.
(12) Jákli, I.; Perczel, A.; Farkas, Ö.; Hollósi, M.; Csizmadia, I. G. Peptide models XXII. A conformational model for aromatic amino acid residues in proteins. A comprehensive analysis of all the RHF/6-31+G* conformers of For-L-Phe−NH2. *J. Mol. Struct.: THEOCHEM* **1998**, *455*, 303−314.
(13) von Helden, G.; Compagnon, I.; Blom, M. N.; Frankowski, M.; Erlekam, U.; Oomens, J.; Brauer, B.; Gerber, R. B.; Meijer, G. Mid-IR spectra of different conformers of phenylalanine in the gas phase. *Phys. Chem. Chem. Phys.* **2008**, *10*, 1248−1256.
(14) Jakli, I.; Perczel, A.; Farkas, O.; Csaszar, A. G.; Sosa, C.; Csizmadia, I. G. Peptide Models XXIII. Conformational Model for Polar Side-Chain Containing Amino Acid Residues: A Comprehensive Analysis of RHF, DFT, and MP2 Properties of HCO-L-SER-NH2. *J. Comput. Chem.* **2000**, *21*, 626−655.
(15) Farkas, O.; Perczel, A.; Marcoccia, J. F.; Hollosi, M.; Csizmadia, I. G. Peptide models XIII. Side-chain conformational energy surface E = E(chi1,chi2) of N-formyl-L-serinamide (For-L-Ser-NH2) in its gamma-L or C7eq backbone conformation. *J. Mol. Struct.: THEOCHEM* **1995**, *331*, 27−36.
(16) Perczel, A.; Farkas, Ö.; Csizmadia, I. G. Peptide Models. 18. Hydroxymethyl Side-Chain Induced Backbone Conformational Shifts of l-Serine Amide. All ab Initio Conformers of For-l-Ser-NH2. *J. Am. Chem. Soc.* **1996**, *118*, 7809−7817.
(17) Tarditi, A. M.; Klipfel, M. W.; Rodriguez, A. M.; Suvire, F. D.; Chasse, G. A.; Farkas, O.; Perczel, A.; Enriz, R. D. An ab initio exploratory study of side chain conformations for selected backbone conformations of N-acetyl-l-glutamine-N-methylamide. *J. Mol. Struct.: THEOCHEM* **2001**, *545*, 29−47.
(18) Barroso, M. N.; Cerutti, E. S.; Rodríguez, A. M.; Jáuregui, E. A.; Farkas, O.; Perczel, A.; Enriz, R. D. Side-chain conformations for selected backbone conformations of N-acetyl-l-isoleucine-N-methylamide and N-acetyl-l-nor-isoleucine-N-methylamide. An exploratory ab initio study. *J. Mol. Struct.: THEOCHEM* **2001**, *548*, 21−37.
(19) Lang, A.; Csizmadi, I. G.; Perczel, A. Peptide Models XLV: Conformational peptides of N-formyl-L-methioninamide and its relevance to methionine in proteins. *Proteins: Struct., Funct., Bioinf.* **2005**, *58*, 571−588.
(20) Zamora, M. A.; Baldoni, H. A.; Bombasaro, J. A.; Mak, M. L.; Perczel, A.; Farkas, O.; Enriz, R. D. An exploratory ab initio study of the full conformational space of N-acetyl-l-cysteine-N-methylamide. *J. Mol. Struct.: THEOCHEM* **2001**, *540*, 271−283.
(21) Boeckx, B.; Ramaekers, R.; Maes, G. A theoretical and matrix-isolation FT-IR investigation of the conformational landscape of N-acetylcysteine. *J. Mol. Spectrosc.* **2010**, *261*, 73−81.
(22) Dobrowolski, J. C.; Jamroz, M. H.; Kołos, R.; Rode, J. E.; Sadlej, J. Theoretical Prediction and the First IR Matrix Observation of Several l-Cysteine Molecule Conformers. *ChemPhysChem* **2007**, *8*, 1085−1094.
(23) Baldoni, H. A.; Rodriguez, A. M.; Zamora, M. A.; Zamarbide, G. N.; Enriz, R. D.; Farkas, Ö.; Csàszàr, P.; Torday, L. L.; Sosa, C. P.; Jàkli, I.; Perczel, A.; Papp, J. G.; Hollosi, M.; Csizmadia, I. G. Peptide models XXIV: An ab initio study on N-formyl-L-prolinamide with trans peptide bond. The existence or non-existence of αL and ϵL conformations. *J. Mol. Struct.: THEOCHEM* **1999**, *465*, 79−91.
(24) Koo, J. C. P.; Chass, G. A.; Perczel, A.; Farkas, Ö.; Torday, L. L.; Varro, A.; Papp, J. G.; Csizmadia, I. G. Exploration of the Four-Dimensional-Conformational Potential Energy Hypersurface of N-Acetyl-l-aspartic Acid N′-Methylamide with Its Internally Hydrogen Bonded Side-Chain Orientation. *J. Phys. Chem. A* **2002**, *106*, 6999−7009.
(25) Hudáky, P.; Beke, T.; Perczel, A. Peptide models XXXIV. Side-chain conformational potential energy surfaces associated with all major backbone folds of neutral tautomers of N- and C-protected l-histidine. An ab initio study on ethylimidazole and N-formyl-l-histidinamide. *J. Mol. Struct.: THEOCHEM* **2002**, *583*, 117−135.
(26) Smolyar, A.; Wong, C. F. Theoretical studies of the spectroscopic properties of tryptamine, tryptophan and tyrosine. *J. Mol. Struct.: THEOCHEM* **1999**, *488*, 51−67.
(27) Kaczor, A.; Reva, I. D.; Proniewicz, L. M.; Fausto, R. Matrix-Isolated Monomeric Tryptophan: Electrostatic Interactions as Non-trivial Factors Stabilizing Conformers. *J. Phys. Chem. A* **2007**, *111*, 2957−2965.
(28) Huang, Z.; Lin, Z. Detailed Ab Initio Studies of the Conformers and Conformational Distributions of Gaseous Tryptophan. *J. Phys. Chem.A* **2005**, *109*, 2656−2659.
(29) Lakard, B. Ab initio study of amino acids containing hydroxy groups (serine, threonine and tyrosine). *J. Mol. Struct.: THEOCHEM* **2004**, *681*, 183−189.
(30) Ramaekers, R.; Pajak, J.; Rospenk, M.; Maes, G. Matrix-isolation FT-IR spectroscopic study and theoretical DFT(B3LYP)/6-31 ++ G** calculations of the vibrational and conformational properties of tyrosine. *Spectrochim. Acta, Part A: Mol. Biomol. Spectrosc.* **2005**, *61*, 1347−1356.
(31) Sapse, A.-M.; Mezei, M.; Jain, D. C.; Unson, C. Ab initio study of aspartic and glutamic acid: supplementary evidence for structural requirements at position 9 for glucagon activity. *J. Mol. Struct.: THEOCHEM* **1994**, *306*, 225−233.
(32) Stepanian, S. G.; Ivanov, A. Y.; Adamowicz, L. Conformational composition of neutral leucine. Matrix isolation infrared and ab initio study. *Chem. Phys.* **2013**, *423*, 20−29.
(33) Boeckx, B.; Maes, G. Experimental and Theoretical Observation of Different Intramolecular H-bonds in Lysine Conformations. *J. Phys. Chem. B* **2012**, *116*, 12441−12449.
(34) Ling, S.; Yu, W.; Huang, Z.; Lin, Z.; Harañczyk, M.; Gutowski, M. Gaseous Arginine Conformers and Their Unique Intramolecular Interactions. *J. Phys. Chem. A* **2006**, *110*, 12282−12291.
(35) Meng, L.; Wu, W.; Zhang, J. Gas Phase Conformations of Selenocysteine and Related Ions: A Comprehensive Theoretical Study. *J. Phys. Chem. A* **2014**, *118*, 1684−1696.
(36) Toroz, D.; Van Mourik, T. The structure of the gas-phase tyrosine-glycine dipeptide. *Mol. Phys.* **2006**, *104*, 559−570.
(37) Toroz, D.; Van Mourik, T. The structure of the gas-phase tyrosine-glycine-glycine tripeptide. *Mol. Phys.* **2007**, *105*, 209−220.
(38) Toroz, D.; van Mourik, T. Structure of the gas-phase glycine tripeptide. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3463−3473.
(39) Lorenzo, L.; Moa, M. J. G.; Mandado, M.; Mosquera, R. A. Do the neighboring residues in a polypeptide affect the electron distribution of an amino acid significantly? A quantitative study using the quantum theory of atoms in molecules (QTAIM). *J. Chem. Inf. Model.* **2006**, *46*, 2056−2065.
(40) Ho, B. K.; Brasseur, R. The Ramachandran plots of glycine and pre-proline. *BMS Struct. Biol.* **2005**, *5*, 14−24.
(41) Kosov, D. S.; Popelier, P. L. A. Convergence of the multipole expansion for electrostatic potentials of finite topological atoms. *J. Chem. Phys.* **2000**, *113*, 3969−3974.
(42) Popelier, P. L. A.; Joubert, L.; Kosov, D. S. Convergence of the Electrostatic Interaction Based on Topological Atoms. *J. Phys. Chem. A* **2001**, *105*, 8254−8261.

(43) Rafat, M.; Popelier, P. L. A. A convergent multipole expansion for 1,3 and 1,4 Coulomb interactions. *J. Chem. Phys.* **2006**, *124*, 144102−1−7.
(44) Rafat, M.; Shaik, M.; Popelier, P. L. A. Transferability of quantum topological atoms in terms of electrostatic interaction energy. *J. Phys. Chem. A* **2006**, *110*, 13578−13583.
(45) Yuan, Y.; Mills, M. J. L.; Popelier, P. L. A. Multipolar Electrostatics for Proteins: Atom-Atom Electrostatic Energies in Crambin. *J. Comput. Chem.* **2014**, *35*, 343−359.
(46) Cardamone, S.; Hughes, T. J.; Popelier, P. L. A. Multipolar Electrostatics. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10367−10387.
(47) Stone, A. J.; Tough, R. J. A. Spherical Tensor Theory of Long-Range Intermolecular Forces. *Chem. Phys. Lett.* **1984**, *110*, 123−129.
(48) Bader, R. F. W. *Atoms in Molecules. A Quantum Theory*; Oxford Univ. Press: Oxford, 1990.
(49) Popelier, P. L. A. *Atoms in Molecules. An Introduction*; Pearson Education: London, 2000.
(50) Popelier, P. L. A. Quantum Chemical Topology: on Bonds and Potentials. In *Structure and Bonding. Intermolecular Forces and Clusters*; Wales, D. J., Eds.; Springer: Heidelberg, Germany, 2005; Vol. *115*; pp 1−56.
(51) Popelier, P. L. A.; Brémond, É. A. G. Geometrically faithful homeomorphisms between the electron density and the bare nuclear potential. *Int. J. Quantum Chem.* **2009**, *109*, 2542−2553.
(52) Bader, R. F. W.; Popelier, P. L. A.; Chang, C. Similarity and complementarity in chemistry. *J. Mol. Struct.: THEOCHEM* **1992**, *255*, 145−171.
(53) Popelier, P. L. A.; Bader, R. F. W. Effect of twisting a polypeptide on its geometry and electron distribution. *J. Phys. Chem.* **1994**, *98*, 4473−4481.
(54) Matta, C. F.; Bader, R. F. W. An Atoms-In-Molecules study of the genetically-encoded amino acids: I. Effects of conformation and of tautomerization on geometric, atomic, and bond properties. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 310−329.
(55) Matta, C. F.; Bader, R. F. W. Atoms-in-Molecules Study of the Genetically Encoded Amino Acids. II. Computational Study of Molecular Geometries. *Proteins: Struct., Funct., Genet.* **2002**, *48*, 519−538.
(56) Matta, C. F.; Bader, R. F. W. Atoms-in-Molecules Study of the Genetically Encoded Amino Acids. III. Bond and Atomic Properties and their Correlations with Experiment including Mutation-Induced Changes in Protein Stability and Genetic Coding. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 360−399.
(57) Jensen, F. Polarization consistent basis sets. III. The importance of diffuse functions. *J. Chem. Phys.* **2002**, *117*, 9234−9240.
(58) Dunning, T. H. J. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(59) Popelier, P. L. A.; Logothetis, G. Characterization of an agostic bond on the basis of the electron density. *J. Organomet. Chem.* **1998**, *555*, 101−111.
(60) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94*, 8897−8909.
(61) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
(62) Hughes, T. J.; Popelier, P. L. A. Where does charge reside in amino acids? The effect of side-chain protonation state on the atomic charges of Asp, Glu, Lys, His and Arg. *Comput. Theor. Chem.* **2014**.
(63) Sagui, C.; Pedersen, L. G.; Darden, T. A. Towards an accurate representation of electrostatics in classical force fields: Efficient implementation of multipolar interactions in biomolecular simulations. *J. Chem. Phys.* **2004**, *120*, 73−87.
(64) Popelier, P. L. A. MORPHY, a program for an automated "atoms in molecules" analysis. *Comput. Phys. Commun.* **1996**, *93*, 212−240.
(65) Popelier, P. L. A. A method to integrate an atom in a molecule without explicit representation of the interatomic surface. *Comput. Phys. Commun.* **1998**, *108*, 180−190.
(66) Császár, A. G.; Perczel, A. Ab initio characterization of building units in peptides and proteins. *Prog. Biophys. Mol. Biol.* **1999**, *71*, 243−309.
(67) Kaminsky, J.; Jensen, F. Force Field Modeling of Amino Acid Conformational Energies. *J. Chem. Theory Comput.* **2007**, *3*, 1774−1788.
(68) Shirazian, S.; Gronert, S. The gas-phase conformations of valine: an ab initio study. *J. Mol. Struct.: THEOCHEM* **1997**, *397*, 107−112.
(69) Stepanian, S. G.; Reva, I. D.; Radchenko, E. D.; Adamowicz, L. Combined Matrix-Isolation Infrared and Theoretical DFT and ab Initio Study of the Nonionized Valine Conformers. *J. Phys. Chem. A* **1999**, *103*, 4404−4412.
(70) Rai, A. K.; Song, C.; Lin, Z. An exploration of conformational search of leucine molecule and their vibrational spectra in gas phase using ab initio methods. *Spectrochim. Acta, Part A* **2009**, *73*, 865−870.
(71) Dokmaisrijan, S.; Lee, V. S.; Nimmanpipug, P. The gas phase conformers and vibrational spectra of valine, leucine and isoleucine: an ab initio study. *J. Mol. Struct.: THEOCHEM* **2010**, *953*, 28−38.
(72) Matta, C. F.; Hernandez-Trujillo, J.; Tang, T.-H.; Bader, R. F. W. Hydrogen-hydrogen bonding: a stabilizing interaction in molecules and crystals. *Chem.—Eur. J.* **2003**, *9*, 1940−1951.
(73) Hofinger, S.; Wendland, M. Method/basis set dependence of the traceless quadrupole moment calculation for N-2, CO_2, SO_2, HCl, CO, NH_3, PH_3, HF, and H_2O. *Int. J. Quantum Chem.* **2002**, *86*, 199−217.
(74) Mills, M. J. L.; Popelier, P. L. A. Polarisable multipolar electrostatics from the machine learning method Kriging: an application to alanine. *Theor. Chem. Acc.* **2012**, *131*, 1137−1153.
(75) Kandathil, S. M.; Fletcher, T. L.; Yuan, Y.; Knowles, J.; Popelier, P. L. A. Accuracy and Tractability of a Kriging Model of Intramolecular Polarizable Multipolar Electrostatics and Its Application to Histidine. *J. Comput. Chem.* **2013**, *34*, 1850−1861.