

Unfavorable Regions in the Ramachandran Plot: Is It Really Steric Hindrance? The Interacting Quantum Atoms Perspective

Peter I. Maxwell and Paul L. A. Popelier *

Accurate description of the intrinsic preferences of amino acids is important to consider when developing a biomolecular force field. In this study, we use a modern energy partitioning approach called Interacting Quantum Atoms to inspect the cause of the φ and ψ torsional preferences of three dipeptides (Gly, Val, and Ile). Repeating energy trends at each of the molecular, functional group, and atomic levels are observed across both (1) the three amino acids and (2) the φ/ψ scans in Ramachandran plots. At the molecular level, it is surprisingly electrostatic destabilization that causes the high-energy regions in the Ramachandran plot, not molecular steric hindrance (related to the intra-atomic energy). At the functional group and atomic levels, the importance of key peptide atoms

(O_{i-1} , C_i , N_i , N_{i+1}) and some sidechain hydrogen atoms (H_i) are identified as responsible for the destabilization seen in the energetically disfavored Ramachandran regions. Consistently, the O_{i-1} atoms are particularly important for the explanation of dipeptide intrinsic behavior, where electrostatic and steric destabilization unusually complement one another. The findings suggest that, at least for these dipeptides, it is the peptide group atoms that dominate the intrinsic behavior, more so than the sidechain atoms. © 2017 The Authors. Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

DOI: 10.1002/jcc.24904

Introduction

Since the original schematic published^[1] in 1963 by Ramachandran et al., few efforts have advanced the *understanding* of the commonly denoted “forbidden” and “accepted” regions of the Ramachandran φ/ψ plot. Subsequent work by Mandel et al.^[2] allowed an advanced representation of the Ramachandran φ/ψ plot to be depicted, detailing the locations of specific hard-sphere repulsions. Much more recently, we note that Mandel’s plot is still used in undergraduate biochemistry textbooks.^[3] Hence, despite the passing of almost half a century, the hard-sphere repulsion models have been accepted and incorporated in the development of many modern-day molecular force fields. Perhaps regrettably, the development of these force fields focused more on successfully parameterising torsional angles rather than on understanding the quantum mechanical nature of the interactions between the atoms involved. We believe that such a greater understanding is an important step towards simplifying the parameterization task, and especially, putting it on a firmer footing. In other words, the re-parameterization of conventional force fields typically creates new terms for experimentally observed structural effects. However, a method that directly partitions quantum mechanical information has a better chance of capturing all effects from the outset, and without extra corrections.

The conformational propensity of the 20 natural amino acids relies on three factors: intrinsic behavior, amino acid sequence, and chemical environment.^[4] Understanding the combined influence of all three aspects on a molecular system requires each factor’s individual behavior to be understood first.

Addressing the first factor, that of intrinsic behavior, many studies have shown that it causes an amino acid to show

preferences in φ/ψ space.^[4–10] One way of investigating intrinsic behavior is through the use of coil libraries. Coil libraries contain sequences of amino acids that form neither α -helical or β -sheet conformations observed in experimental X-ray crystal structures. However, the conformations collected are always influenced by the surrounding protein structure since they are simply extracted from the initial complete protein. Hence, the individual amino acids are possibly biased by the tertiary structure of a protein. Also, they are still biased by being inside a sequence of amino acids. However, it is also known that observing isolated dipeptide structures are not representative of the amino acids behavior in oligopeptide chains.^[11] Note, as a brief aside, that the often used but actually confusing name “dipeptide” refers to a *single* amino acid, flanked by a peptide bond at both termini. Here, we must ask, at what point does an oligopeptide become a sequence, resulting in the peptide’s behavior being a result of sequential effects, rather than of the intrinsic behavioral effects of its amino acids? Could it be that only the presence of intramolecular stabilization should be associated with the intrinsic behavior? Despite some

P. I. Maxwell, P. L. A. Popelier

Manchester Institute of Biotechnology (MIB), 131 Princess Street, Manchester M1 7DN, Great Britain and School of Chemistry, University of Manchester, Oxford Road, Manchester, Great Britain M13 9PL

E-mail: pla@manchester.ac.uk

Contract grant sponsor: BBSRC (P.I.M., PhD studentship) and EPSRC Established Career Fellowship (P.L.A.P., EP/K005472)

© 2017 The Authors. Journal of Computational Chemistry Published by Wiley Periodicals, Inc.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

investigations of single amino acids poorly replicating their behavior in larger systems, there are also many reports identifying the importance of their study.^[11–14]

The second factor is that of sequencing effects. As the name suggests, this is the effect of having a sequence of amino acids, that is, an oligopeptide. The formation of α -helices, β -sheets, and loops result from the specific sequence of amino acids in a chain. The α -helices, β -sheets, and loops are regularly occurring oligopeptide structural arrangements allowing amino acids to be packed closer together, typically through hydrogen bonding across different amino acids in the sequence. The formation of such secondary structure landmarks still rely partly on intrinsic propensity, but one could argue are dominated by the interatomic interactions formed between the sidechain and backbone atoms of neighboring amino acids in a sequence. Capturing the behavior of oligopeptide sequences still remains a non-trivial task. A good example is given in a study by Best et al.,^[15] which recently reported the helical character induced upon tri-, tetra-, and penta-Ala oligopeptides by many common force fields: GROMOS (53a6) $\sim 13.1\%$, CHARMM27 (with CMAP) $\sim 57.5\%$, AMBER03 $\sim 62.3\%$, AMBER99 $\sim 94.2\%$, and AMBER94 $\sim 97.6\%$. The experimental value each force field was striving to achieve was $\sim 20\%$. The helicity was excessive for all force fields other than GROMOS. Such is the motivation for many studies determining new or improved torsional potentials in all conventional force fields.^[16–23]

The third factor, chemical environment, is perhaps the most difficult to investigate due to the computational expense. Chemical environment behavior may be investigated through observing the influence of multiple sequences on a defined central sequence. However, chemical environment may also be an investigation of solvation effects. Both introduce many new intermolecular bonds to consider, and scale the system size dramatically.

Today we use quantum mechanical methods to investigate the first factor: intrinsic behavior. In order to exclusively observe the intrinsic behavior, so-called dipeptides have been chosen as a first point of investigation. We isolate the intrinsic behavior through eliminating (i) sequencing effects by working with the single amino acid blocked (or “capped”) with an acetyl group ($-\text{COCH}_3$) and an amide group ($-\text{NHCH}_3$), and (ii) chemical environment through working *in vacuo*. The additional benefit of vacuum conditions, other than computational cost, is that they are important for the study of amino acids in the hydrophobic core of folded proteins typically inaccessible to solvent.^[24] Working with gas phase *ab initio* data is also in accordance with most force field development.^[25–30] Our investigation aims to validate (or contrast, were appropriate) the long-standing interpretation of the regions in the Ramachandran plot. To do this, we will use the Interacting Quantum Atoms (IQA) energy partitioning method, an approach that falls under the “umbrella” approach of Quantum Chemical Topology (QCT), a name first coined^[31] in 2003. The IQA method allows the calculation of atomic energies, which together account for the full molecular energy. The atomic energies can be classified into both intra- and inter-atomic

components, and also by energy type, for example, electrostatic, exchange, correlation, and so forth. Strategically chosen excursions (see Figure 1) through φ/ψ conformational space are used to obtain system conformations representing multiple regions of the Ramachandran plot. Three systems are investigated: glycine (Gly), valine (Val), and isoleucine (Ile), representing a gradual increase in the number of atoms making up the aliphatic residues, going from $-\text{NH}-\text{C}_\alpha\text{H}_2-\text{CO}-$ (in Gly) over $-\text{NH}-\text{C}_\alpha\text{H}(\text{C}_\beta\text{H}(\text{C}_\gamma\text{H}_3)_2)-\text{CO}-$ (in Val) to $-\text{NH}-\text{C}_\alpha\text{H}(\text{C}_\beta\text{H}(\text{C}_\gamma\text{H}_3)(\text{C}_\delta\text{H}_2\text{C}_\delta\text{H}_3)-\text{CO}-$ (in Ile).

The investigation will allow us to identify the key atoms, both single and group thereof, that are responsible for both high- and low-energy regions in the respective Ramachandran plots, along with any global trends that consistently appear across the three systems investigated. The existence of global trends at the atomic level will indicate transferability^[32,33] within the systems, a key cornerstone of many force fields and a topic we have previously reported on for oligopeptide chain.^[34]

Methods

Dataset generation

The optimized geometries of the global energy minima of glycine (Gly) dipeptide, valine (Val) dipeptide, and isoleucine (Ile) dipeptide were taken from our previous work.^[35] The angle φ is defined as the $\text{C}_{i-1}-\text{N}_i-\text{C}_\alpha-\text{C}_i$ dihedral, and ψ as $\text{N}_i-\text{C}_\alpha-\text{C}_i-\text{N}_{i+1}$. Figure 2 shows these two angles and the nuclei involved in defining them. The generic notation used here will be explained in the next section. For each system in turn, the φ and ψ dihedral angles were rotated by 15° increments between -180° and $+180^\circ$, resulting in $24 = [180 - (-180)]/15$ geometries, additional to the global minimum. First, the ψ dihedral angle of the global minimum was frozen, while the φ angle was rotated by the increment angle over the full range ($-180^\circ \leq \varphi \leq +180^\circ$) using the GAUSSVIEW package. Once all 24 additional geometries for φ were obtained, collectively known as the *phi* (φ) scan, the procedure was repeated but now freezing the global minimum's φ angle and incrementing the ψ angle over the full range ($-180^\circ \leq \psi \leq +180^\circ$), generating the *psi* (ψ) scan. The 48 additional geometries (24 for each of the two scans in total), were then relaxed through geometry optimization but keeping both the φ and ψ dihedral angles frozen. Note that the residues were also optimized and not kept rigid relative to φ or ψ . The program GAUSSIAN09^[38] was used to perform the geometry optimization, single-point energy calculations and printing of the wavefunction for subsequent QCT analysis, for each geometry. The optimizations and calculations were also performed at B3LYP/apc-1 level,^[39] which is the same level of theory with which the optimized coordinates were originally obtained. In total, $6 (=2 \times 3)$ sets of geometries were obtained, arising from two scans carried out on each of the three capped amino acids, each set consisting of $25 (=24 + 1)$ geometries. In summary, the overall analysis of all systems is based on $147 = 6 \times 25 - 3$ IQA-partitioned wave functions, where we corrected for the fact that the three global minima are used for both φ and ψ .

After the *ab initio* calculations described above, the IQA energy partitioning calculations were performed using the AIMAll program^[40] (version 16.01.09). The non-default settings requested in AIMAll for the IQA calculations were: the “TwoE” program for the calculation of intra-atomic electron-electron repulsion energies was turned off (*-usetwoe* = 0), the target spacing between interatomic surface paths was improved from *fine* to *very fine* to ensure accurate atomic integrations (*-iasmesh* = *veryfine*), and atomic IQA energies were requested (*-encomp* = 3). The IQA energy partitioning is outlined in Section “The Interacting Quantum Atoms (IQA) Approach”, which provides only the relevant equations. In order to gauge the accuracy of AIMAll’s energy partitioning, the IQA molecular energies were compared to the *ab initio* energies obtained from GAUSSIAN09. The discrepancy between this (unpartitioned) *ab initio* molecular energy and the IQA-reconstructed molecular energy is referred to as the *IQA recovery error*. For some geometries of higher energy, the obtained IQA recovery error was considered to be too high ($1 \text{ kJ mol}^{-1} < \text{IQA recovery error} < 1.5 \text{ kJ mol}^{-1}$). Hence, the IQA energies for these geometries were recalculated using stricter conditions for the basin outer angular quadrature (*-boaq* = *skyhigh_leb* instead of the default *-boaq* = *auto*) in order to obtain better atomic integration accuracy. The best IQA energies, as determined by the IQA recovery error, were incorporated into the final dataset and will be reported in the Results section.

Generic notation

Figure 2 illustrates the notation followed throughout this article. A generic notation is useful in the current study because it allows atoms that are present in all three amino acids to be identified using a single atom label and, thus, easily compared across the three amino acids. This notation is more concise than that of the unique atomic labels assigned by GAUSSVIEW, which naturally change with varying system size. The standard residue subscript labels are used, namely α , β , γ , and δ , and each is assigned to the covalently bonded atoms forming the residue. Both the carbonyl and amino groups at either side of the C_α are labelled with label “*i*,” which refers to the central residue as a subscript. Either side of these groups, the adjacent carbonyl and amino groups are labelled as “*i* - 1” and “*i* + 1,” respectively. So, an increasing index refers to a move towards the NHCH_3 terminus (by convention on the right), while a decreasing index refers to moving in the opposite direction, towards the acetyl C(=O)CH_3 terminus (on the left).

The Interacting Quantum Atoms approach

IQA^[41] is a topological approach that sits alongside the Quantum Theory of Atoms in Molecules (QTAIM)^[42–44] and the Electron Localization Function (ELF)^[45] under the collective header of QCT.^[46,47] All three share the central idea of using the gradient vector field to extract chemical information from a system. QTAIM and IQA both share the presence of topological atoms. Topological atoms, such as those seen in Figure 2 for isoleucine dipeptide, are finite-volume three-dimensional fragments of space representing a single atomic basin, determined by

the gradient paths of a system’s electron density. These atomic basins (i.e., atoms) are well-defined even when molecules are compressed (short range van der Waals complexes), and they are space-filling (i.e., non-overlapping and gapless). The latter feature ensures that in an analysis of properties derived from the electron density (such as atomic energies) no part of the system is unaccounted for. This hallmark is an important advantage^[48] of QCT, particularly when applications are expanded to interactions between ligands and proteins^[49] where currently classically standardized van der Waals radii are used leaving areas of space unattributed to either the ligand or protein. The previously entitled Quantum Chemical Topological Force Field,^[50] but recently renamed to FFLUX,^[51] is a force field currently being developed with topological atoms at its heart. FFLUX features a novel design, unlike the classical designs used in other popular force fields such as AMBER and CHARMM. FFLUX maps geometrical change to a change in atomic energy through a machine learning method known as kriging.^[52] Two recent publications^[50,53] describe its architecture and the process of model building in detail. FFLUX uses four primary energies to describe a molecule (or any system). The energies are obtained via the IQA energy partitioning, and include the intra-atomic energy, the classical electrostatic energy, the exchange energy and the correlation energy. Each will be introduced in turn and described through the following equations.

IQA partitions a molecule’s energy, $E_{\text{IQA}}^{\text{Mol}}$, into a sum of atomic energies, E_{IQA}^A , which in turn are composed of intra-atomic and inter-atomic energy components:

$$E_{\text{IQA}}^{\text{Mol}} = \sum_A E_{\text{IQA}}^A = \sum_A E_{\text{intra}}^A + \frac{1}{2} \sum_A \sum_{B \neq A} V_{\text{inter}}^{\text{AB}} = \sum_A \left[E_{\text{intra}}^A + \frac{1}{2} \sum_{B \neq A} V_{\text{inter}}^{\text{AB}} \right] \quad (1)$$

where A and B represent atoms, the superscript denotes the atoms the energy is associated with and the subscript denotes the type of energy, a format that applies to all subsequent equations.

The intra-atomic energy can be divided into its kinetic, T , and potential, V , energy contributions as follows:

$$E_{\text{intra}}^A = T^A + V_{\text{ee}}^{\text{AA}} + V_{\text{en}}^{\text{AA}} \quad (2)$$

where T^A represents the kinetic energy of atom A , $V_{\text{ee}}^{\text{AA}}$ is the (repulsive) potential energy between the electrons within atom A , and $V_{\text{en}}^{\text{AA}}$ is the (attractive) potential energy between the electrons and nucleus of atom A .

Similarly, the interatomic energy can be divided into its potential energy contributions (there is no kinetic contribution this time):

$$V_{\text{inter}}^{\text{AB}} = (V_{\text{nn}}^{\text{AB}} + V_{\text{en}}^{\text{AB}} + V_{\text{ne}}^{\text{AB}}) + V_{\text{ee}}^{\text{AB}} \quad (3)$$

where $V_{\text{en}}^{\text{AB}}$, $V_{\text{ne}}^{\text{AB}}$, and $V_{\text{ee}}^{\text{AB}}$ follow the same format as described earlier. This time the superscript and subscript ordering playing a more important role. For example, $V_{\text{en}}^{\text{AB}}$ refers to the electrons of A and the nucleus of B . Additionally, $V_{\text{nn}}^{\text{AB}}$ is the (repulsive) potential energy between the nuclei of A and B . The first three

terms are bracketed to illustrate their connection to forming the “classical” electrostatic energy. To complete the electrostatic energy, V_{ee}^{AB} must be expanded to:

$$V_{ee}^{AB} = V_{Coul}^{AB} + V_x^{AB} + V_{corr}^{AB} \quad (4)$$

Here, “Coul” refers to the Coulombic interaction between the electrons, “x” represents the exchange energy, and “corr” the correlation energy. Now that the Coulombic energy has been separated from V_{ee}^{AB} , the classical electrostatic energy V_{cl}^{AB} can be represented as:

$$V_{cl}^{AB} = (V_{nn}^{AB} + V_{en}^{AB} + V_{ne}^{AB}) + V_{Coul}^{AB} \quad (5)$$

allowing the interatomic interaction energy to be rearranged to

$$V_{inter}^{AB} = V_{cl}^{AB} + V_x^{AB} + V_{corr}^{AB} = V_{cl}^{AB} + V_{xc}^{AB} \quad (6)$$

This arrangement is intuitive: the classical electrostatic energy can be identified separately from the exchange and correlation energies, which together, can be thought of as the covalent contribution within an interaction.

A recent FFLUX publication^[53] introduced the use of interatomic energies designated by AA' instead of AB . Here A' represents every other atom in the molecular system except A . Thus, the notation AA' denotes the interatomic energy between an atom A and its surrounding environment A' , such that

$$\sum_A V^{AA'} \cong \sum_A \sum_{B \neq A} V^{AB} \quad (7)$$

The energies in eq. (7) are only approximately equivalent because they use two separate algorithms for calculation, one analytical (left term) and one numerical (right term), naturally resulting in some minor differences between the values.

In this investigation, we will also study the IQA energies at the molecular level, as well as at the atomic level and at (functional) group level (more precisely, at the level of a meaningful collection of atoms). In order to define the “molecular energies,” we observe that:

$$E_i^{Mol} = \sum_A E_i^A \quad (8)$$

where “i” may be substituted for the IQA energy type of choice (i.e., *intra*, *IQA*, *cl* or *xc*). Note that for the latter two subscripts, E is replaced by V ($E_{cl}^A \equiv V_{cl}^{AA'}$ and $E_{xc}^A \equiv V_{xc}^{AA'}$). A particular type of molecular energy (e.g., electrostatic or exchange) is then obtained from a simple summation of the respective energy type over every atom A . Similarly, a particular energy of a (functional) group is obtained by energy summation over every atom belonging to the (functional) group. As a result, a hierarchical search for chemical insight can be carried out whereby first the total energy profile itself is studied (E_{IQA}^{Mol}), next the various energy types at the molecular level

(E_i^{Mol}), at the (functional) group level (E_i^G) (where G is any meaningful collection of atoms), and finally the various energy types at the atomic level (E_i^A).

The IQA approach has been used to study many different chemical systems such as the interactions of Zn(II) complexes,^[54] organoselenium molecules,^[55] halogen-trinitromethanes,^[56] halogen bonding,^[57,58] and hydrogen bonding.^[59,60] IQA has also been used to shed light on chemical phenomena such as steric repulsion,^[61–63] hyperconjugation,^[64] reactions,^[65] and transferability.^[34] The broad applicability of IQA, and its well-defined and robust quantitative nature make it ideal for the current investigation. We note that IQA does not suffer from a list of conceptual and numerical problems plaguing the older and more traditional energy decomposition analysis, the many variants of which have recently been reviewed and critically discussed.^[66]

The next point to highlight regards IQA’s compatibility limitations, in particular the lack of affordable correlation. Until recently, IQA was incompatible at theory levels other than Hartree–Fock, full configuration interaction, configuration interaction with single and double excitations, and complete active space. This is due to perturbation theory remaining computationally very expensive even for small systems, and standard density functional theory (DFT) not providing a well-defined second-order reduced density matrix. However, recent developments have managed to expand IQA’s application to include at least some correlation through B3LYP^[67–69] and M06–2X level DFT, and the direct correlation through coupled cluster with single and double excitations^[70,71] level. In 2016, MPn-IQA ($n = 2, 3, \text{ or } 4$) also became possible.^[72] The inclusion of correlation is anticipated to have important consequences in the investigation of systems driven by dispersion energy. For further details on the expansion of IQA, the reader is directed to the respective references. Accordingly, for a more complete description of the IQA approach, the original paper of Blanco et al.^[41] should be consulted.

A final point discusses a potential concern in connection with the validity of the atomic virial theorem. Although at the root of QAIM, this theorem is actually irrelevant for both IQA atomic energies and molecular energies, in the sense that IQA does not assume nor use the virial theorem (either atomic or molecular) in any way. We also note that self-consistent virial scaling (SCVS) is not applicable anyway to DFT methods, such as B3LYP, so the energy partitioning in this paper could not benefit from such a correction in the first place. It is debatable whether virial-based atomic energies are useful, in practice, even with SCVS to satisfy the molecular virial theorem. One could go as far as to state that virial-based energies are given in AIMAll basically for historical reasons. In summary, our results are not affected by the concern raised above.

Results and Discussion

Preliminary analysis

The aforementioned IQA recovery error is due to the integration error, $L(\Omega)$, that accompanies each atomic integration. For

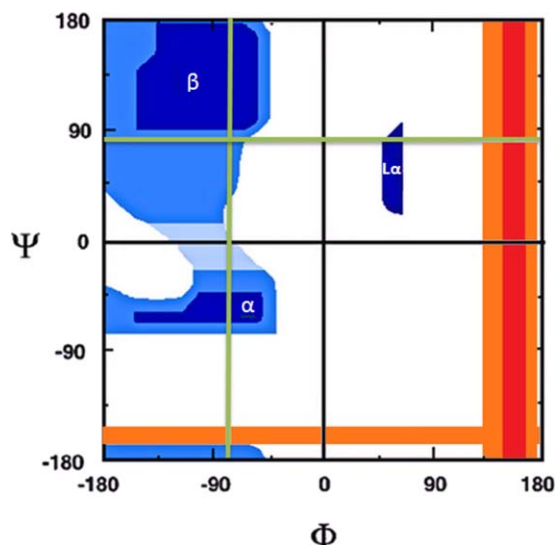


Figure 1. Schematic Ramachandran plot indicating the positions of β -turns (marked as β), right-handed helices (α), and left-handed helices ($L\alpha$). Trajectories across the φ (fixed ψ) and ψ (fixed φ) torsional angles are indicated in green. The exact positions of the trajectories will vary slightly depending on the global minimum φ/ψ angles for each system (Gly, Ile, and Val). The crossing-point of the green trajectories indicates the φ/ψ angles of the global minimum. The area of the Ramachandran plot is shaded according to various degrees of energetic stability: very favorable (dark blue), favorable (blue), slightly favorable (light blue), slightly unfavorable (white), unfavorable (orange) and very unfavorable (red). [Color figure can be viewed at wileyonlinelibrary.com]

our systems, the mean absolute IQA recovery errors for the φ scans were 0.39, 0.60, and 0.95 kJ mol^{-1} for Gly, Val, and Ile, respectively. For the ψ scans, they were 0.35, 0.58, and 0.57 kJ mol^{-1} , respectively. With observed relative energy barriers of up to $\sim 64 \text{ kJ mol}^{-1}$ and mean absolute IQA recovery errors of up to 0.95 kJ mol^{-1} , the maximum percentage error of the values becomes $(0.95/64) \times 100 = 1.5\%$. We conclude that all

effects seen and discussed are far above integration noise. The energy profile with the highest energy range (Ile- φ) also has the highest mean absolute IQA recovery error. Larger atomic integration errors are typically observed for atoms in more complex geometries, for example, in molecules energetically far from the global energy minimum or in molecules with unusual topology.

Analysis at molecular level

Figure 3 plots the $E_{\text{IQA}}^{\text{Mol}}$ energy profiles for each system, and for both the φ and ψ scans. The colored regions depict the similarity between Gly and Val/Ile energies: (1) brown indicates confluence between all three dipeptides, (2) navy indicates the appearance of an additional maximum in Val/Ile, not seen for Gly, and (3) orange indicates a change in position of a maximum seen for all three dipeptides. The first point to note is the similarity between the Val and Ile energy profiles throughout both scans. Using the Pearson correlation coefficient r , where $r = 1$ indicates a perfectly correlated dataset, values of $r = 0.996$ and $r = 0.997$ are obtained between the Val and Ile energy profiles within the φ and ψ scans, respectively. The striking similarity between the profile of Val and Ile is not surprising given that their side chains only differ by a methylene group. In contrast, the energy profile of Gly is less correlated to that of Val, for example, with $r = 0.857$ and $r = 0.833$ for the φ and ψ scans, respectively. However, in the φ scan, it is clear that the common backbone structure between Gly and Val/Ile is accountable for the molecular barrier interval of $-150^\circ \leq \varphi \leq +15^\circ$ (brown area in Fig. 3, panel φ) where very similar energy profiles are observed across all three systems. Outside of this interval, we deduce that the sidechain must influence the energy profile and cause the maximal φ torsional barrier at $+165^\circ$ for Val and Ile, which is absent in Gly (navy area in Fig. 3, panel φ). In the ψ scans, the Gly and Val/Ile

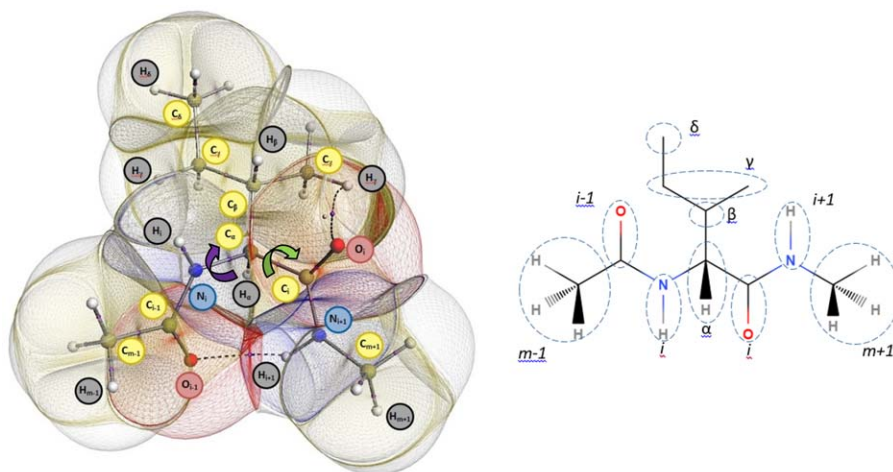


Figure 2. (left) Topological atoms occurring in the global energy minimum of the isoleucine dipeptide (Ile) with atom generically labelled. The dihedral angles φ and ψ are marked by a purple and green arrow, respectively. The atoms are space-filling: they do not overlap and leave no gaps between them. Note that not all methylene or methyl hydrogen atoms are labeled in order to avoid cluttering the figure. This emblematic figure was generated by the in-house program IRIS, which is based on previously published^[36,37] algorithms. The following fragmentation will prove to make sense later in this article: $\text{CH}_3[\text{C}(=\text{O})-\text{N}(\text{H})\text{C}_2\text{HR}[\text{C}(=\text{O})-\text{N}(\text{H})\text{CH}_3]$, where each fragment is flanked by two vertical bars and consists of 4, 4, 15, 4, and 4 atoms, respectively, totaling 31 atoms; (right) schematic clarifying the notation in this paper (for the i th amino acid, which is isoleucine in this case). [Color figure can be viewed at wileyonlinelibrary.com]

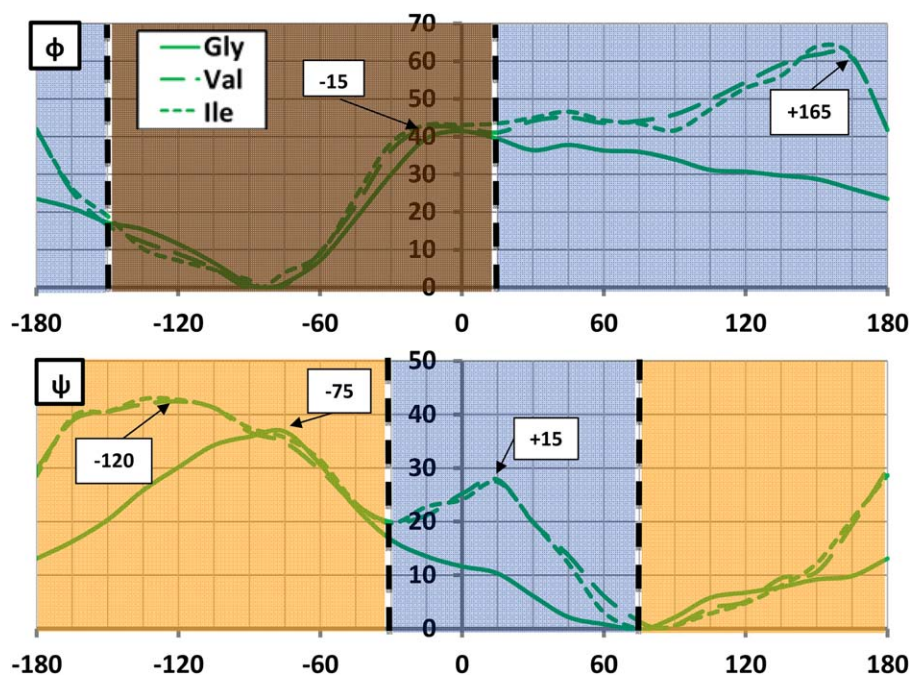


Figure 3. $\Delta E_{\text{IQA}}^{\text{Mol}}$ (in kJ mol^{-1}) scanned across ϕ (top) and ψ (bottom), for the Gly (solid), Val (large dashed), and Ile (small dashed) systems. All ϕ energies are relative to the optimized global minima at $\phi = -81.9^\circ$, -84.4° , and -84.1° and all ψ energies are relative to the optimized global minima at $\psi = +69.8^\circ$, $+83.7^\circ$, and $+82.5^\circ$, respectively. The relevant energy maxima discussed in the main text are marked here for convenience. The brown area in the ϕ scan marks a region of high confluence between Gly, Val, Ile where the influence of the side chain is minimal. The navy areas in the ϕ and ψ scans mark regions where an extra maximum appears due to the presence of a side chain. The orange areas in the ψ scan mark regions where no new maxima appear but existing maxima are shifted. [Color figure can be viewed at wileyonlinelibrary.com]

energy profiles are less correlated according to r , and indeed turn out to be more different visually. The additional local maximum at $\psi = +15^\circ$ (navy area in Fig. 3, panel ψ), and the translation of the maximum barrier in ψ , from -75° in Gly, to -120° (in Val/Ile) (orange area in Fig. 3, panel ψ) suggests a broader influence of the sidechain throughout the dihedral angles. In summary, the sidechain influences the energy profile more in ψ than in ϕ , because the former lacks the brown area of high energy profile confluence. To aid the interpretation of Figure 3, Figure 4 shows the molecular graphs of the two energy maxima in ϕ scan (-15° and $+165^\circ$), and two energy maxima in the ψ scan (-120° and $+15^\circ$).

So far, we have presented an unpartitioned perspective based on geometrical differences between Gly, Val, and Ile, allowing us to comment on the general influence of the sidechain (and its size) on the molecular energy. The current literature states that the barrier at $\phi = +165^\circ$ is a result of the β -branching on the residue causing hard-sphere steric clashes^[73] between O_{i-1} and C_β (at B3LYP/ANO-L-VDZP level), whereas the barrier at $\psi = -120^\circ$ relates to clashes^[74] between C_β and N_{i+1} . However, around $\psi = +15^\circ$, where our earlier observations would suggest a sidechain-related destabilization, the literature reports the region as being “sterically allowed.”^[74] This region will be investigated further later. Finally, the barrier at $\phi = -15^\circ$, which occurs in all three systems, is reportedly due to two sets of backbone clashes, one between O_{i-1} and C_α^2 , and a second^[74] between O_{i-1} and N_{i+1} . Figure 4 is meant to help in visualising the atomic clashes mentioned above but a careful inspection may leave the impression that a more

thorough atomic analysis, in the spirit of topological atoms, is needed.

To this point, we only commented on the general regions where one expects the sidechain atoms to influence (navy/orange in Fig. 3), or not (brown), the molecular energies, given the structural differences between Gly and Val/Ile. To comment further on the agreement between literature and the IQA

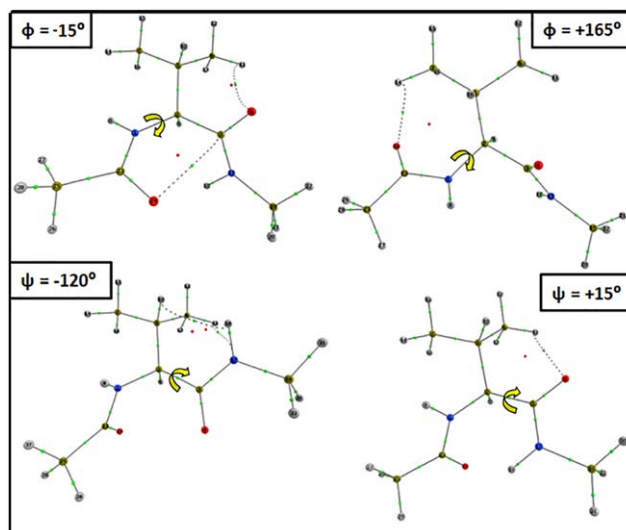


Figure 4. Conformations of Val at each of the maximum energy torsion angles of $\phi = -15^\circ$, $\phi = +165^\circ$, $\psi = -120^\circ$, and $\psi = +15^\circ$ shown in Figure 3. Note that the backbone geometries also refer to the energy maxima in Ile and Gly, with the exception of the $\psi = -75^\circ$ energy maximum unique for Gly. [Color figure can be viewed at wileyonlinelibrary.com]

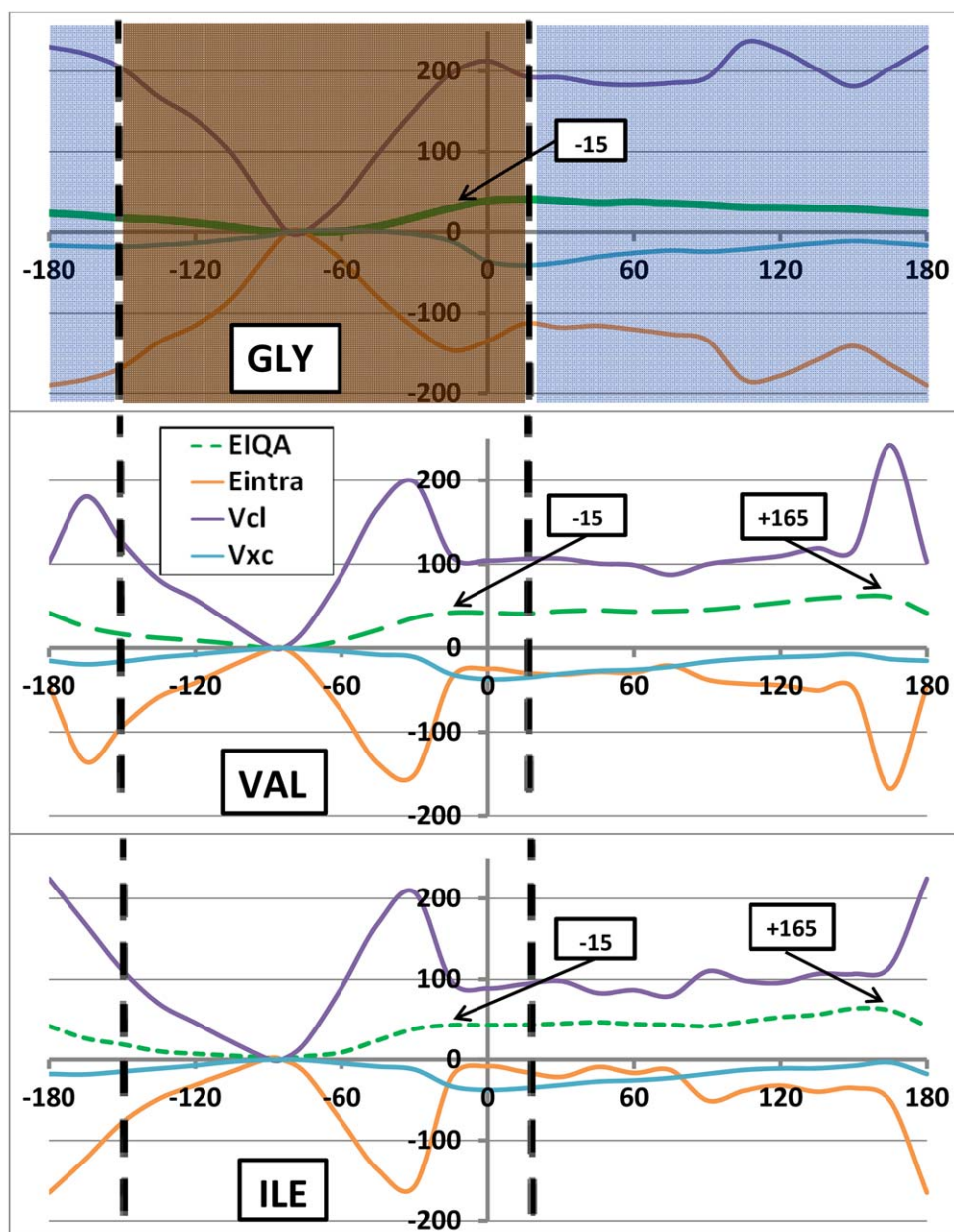


Figure 5. Breakdown of $\Delta E_{\text{IQA}}^{\text{Mol}}$ relative energies (in kJ mol^{-1}) (green) for the φ scan into $\Delta V_{\text{cl}}^{\text{Mol}}$ (purple), $\Delta E_{\text{intra}}^{\text{Mol}}$ (orange), and $\Delta V_{\text{xc}}^{\text{Mol}}$ (turquoise) components for (a) Gly, (b) Val, and (c) Ile. The same convention of the green line types (solid, small dash, large dash) has been applied throughout the paper, starting in Figure 3. [Color figure can be viewed at wileyonlinelibrary.com]

perspective, it is necessary to partition the molecule into fragments. The IQA interpretation for the causes of the observed maxima (and minima) will now be investigated, for each system, at three partitioning levels: molecular, functional group (or collections of atoms) and atomic (i.e., single atom).

We now analyze the overall trends of the molecular IQA energies (relative to the global minimum) illustrated for the φ and ψ scans in Figures 5 and 6, respectively. These figures show, for each of the three systems, the profile of each IQA molecular energy contribution to $\Delta E_{\text{IQA}}^{\text{Mol}}$, that is $\Delta V_{\text{cl}}^{\text{Mol}}$, $\Delta E_{\text{intra}}^{\text{Mol}}$, and $\Delta V_{\text{xc}}^{\text{Mol}}$ (see eq. (8)). For convenience, $\Delta E_{\text{IQA}}^{\text{Mol}}$ is plotted again for each system, repeating what was already shown in Figure 3. It is clear that the energy scale between Figure 3 and Figure

5 (or Fig. 6) differs by about an order of magnitude. This scale difference explains why the energy barriers look less pronounced at the scale of hundreds of kJ mol^{-1} , which is necessary though to show the behavior of the three types of molecular energy contributions to the total molecular energy. The difference in energy scales also suggest immediately that substantial energy cancellation must take place.

Indeed, $\Delta V_{\text{cl}}^{\text{Mol}}$ and $\Delta E_{\text{intra}}^{\text{Mol}}$ broadly mirror each other, at either side of the zero energy line, and thereby more or less cancel each other. Meanwhile, $\Delta V_{\text{xc}}^{\text{Mol}}$ acts as a spectator since the absolute magnitude of its values stays of the order of tens of kJ mol^{-1} (peaking at 41 kJ mol^{-1}). This also means that a curve representing $\Delta V_{\text{cl}}^{\text{Mol}} + \Delta E_{\text{intra}}^{\text{Mol}}$ (not shown in Fig. 5) would

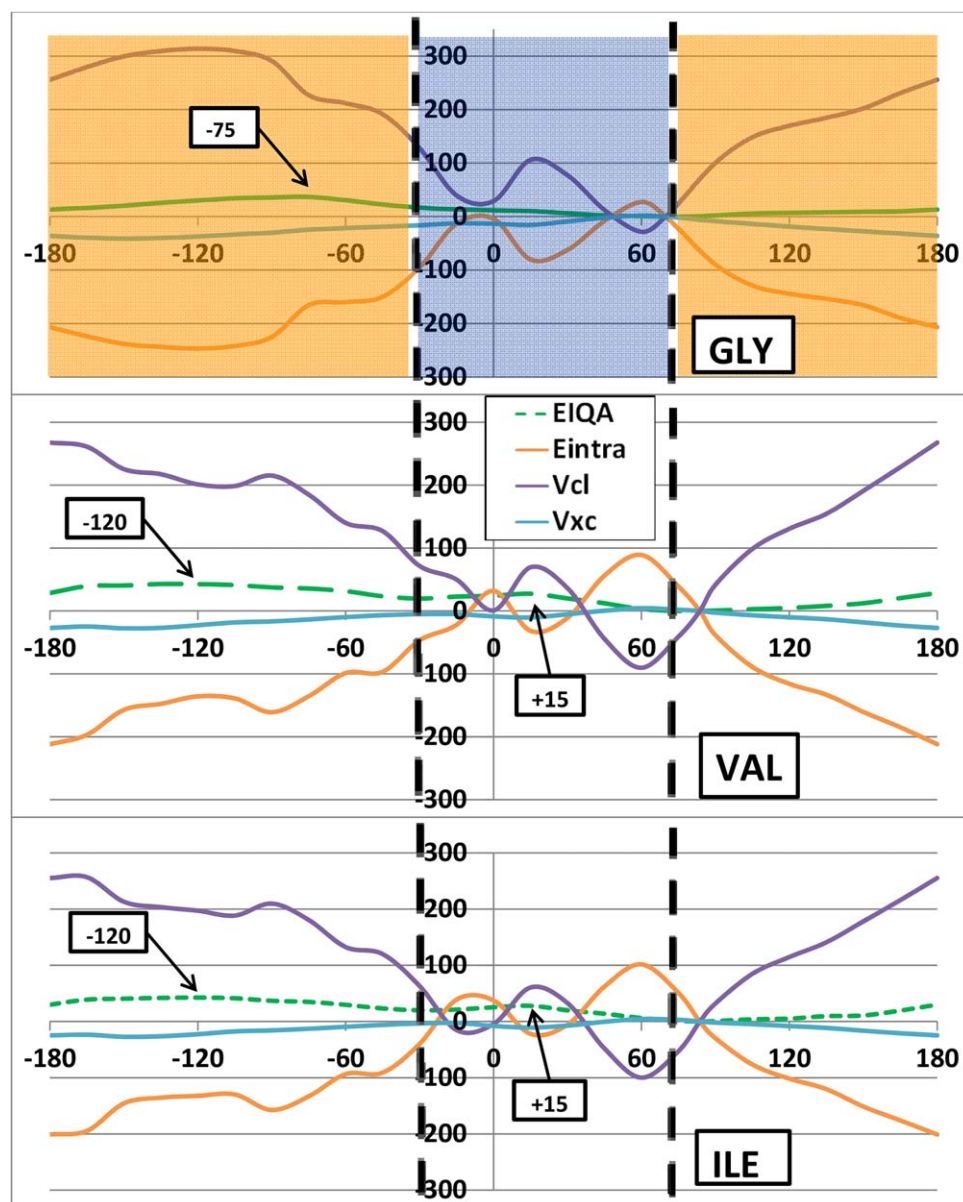


Figure 6. Breakdown of $\Delta E_{\text{IOA}}^{\text{Mol}}$ relative energies (in kJ mol^{-1}) (green) into $\Delta V_{\text{cl}}^{\text{Mol}}$ (purple), $\Delta E_{\text{intra}}^{\text{Mol}}$ (orange), and $\Delta V_{\text{xc}}^{\text{Mol}}$ (turquoise) components for (a) Gly, (b) Val, and (c) Ile, for the ψ scan. The same convention of the green line types (solid, small dash, large dash) has been applied throughout the paper, starting in Figure 3. [Color figure can be viewed at wileyonlinelibrary.com]

be quite similar to that of $\Delta E_{\text{IOA}}^{\text{Mol}}$. The mirroring of electrostatics by $\Delta E_{\text{intra}}^{\text{Mol}}$ is not that surprising in the light of earlier work by Martín Pendás et al.: charging an atom leaves behind a very important energetic fingerprint in the atomic self-energies. *In vacuo*, the latter would increase linearly through the appropriate ionization potentials, or decrease (becoming more negative) with electron affinities (see, for instance, Ref. 75).

We learn that the cause of the energy barriers is consistent across the φ scan for each of the three systems: the classical electrostatic energy $\Delta V_{\text{cl}}^{\text{Mol}}$ (purple) is destabilized in the barrier regions, relative to the global minimum. Correspondingly, the height of the barriers at 0° and $+165^\circ$ (Val/Ile only) is dampened through stabilization of $\Delta E_{\text{intra}}^{\text{Mol}}$ (orange) and $\Delta V_{\text{xc}}^{\text{Mol}}$ (turquoise). It is important to note that $V_{\text{cl}}^{\text{Mol}}$ is always negative (attractive) across all systems and for all φ/ψ combinations.

Thus, the positive *relative* (Δ) energies should be interpreted as a destabilization (i.e. a lack of stabilization) with respect to the global minimum rather than as a repulsive energy. Both $E_{\text{intra}}^{\text{Mol}}$ and $V_{\text{xc}}^{\text{Mol}}$ energies are also always negative in value, and should thus not be mistaken to be repulsive energies either.

Within the ψ scan, the destabilization of $\Delta V_{\text{cl}}^{\text{Mol}}$ is again the cause of the barrier within the marked orange region but both $\Delta V_{\text{cl}}^{\text{Mol}}$ and $\Delta E_{\text{intra}}^{\text{Mol}}$ fluctuate between being stabilising and destabilising in the navy region. The fluctuations result from the formation of two intramolecular hydrogen bonds causing sterically destabilising 1,5 ($\psi = -15^\circ$) and 1,7 ($\psi = +60^\circ$) intramolecular rings in the backbones of the dipeptides. The hydrogen bonds involve the $N_{i-1} \cdots H_{i+1}-N_{i+1}$ atoms (1,5 ring) and the $O_{i-1} \cdots H_{i+1}-N_{i+1}$ atoms (1,7 ring), and are illustrated in Figure 7 for the Val system at conformations $\psi = 0^\circ$ (top) and $\psi = +60^\circ$

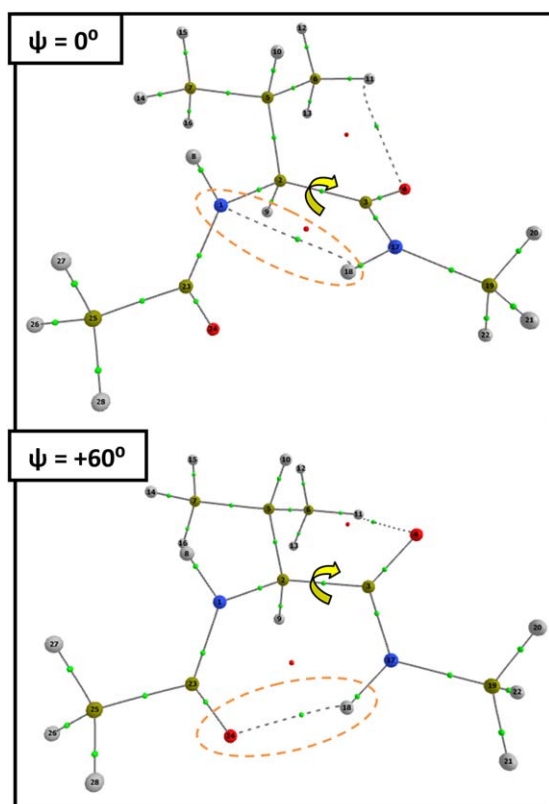


Figure 7. Conformations of Val illustrating the presence of (top) a 1,5 cyclic intramolecular hydrogen bond containing $N_i \dots H_{i+1} - N_{i+1}$ (orange) for $\psi = 0^\circ$ and (bottom) a 1,7 hydrogen-bonded ring containing $O_{i-1} \dots H_{i+1} - N_{i+1}$ (orange) for $\psi = +60^\circ$. [Color figure can be viewed at wileyonlinelibrary.com]

(bottom). Accompanying the destabilising intra-atomic energy is a stabilization of the electrostatic energy within these atoms, which is expected during the formation of a hydrogen bond.

The fact that $V_{\text{cl}}^{\text{Mol}}$ causes its own high-energy regions is interesting, surprising, and perhaps even controversial. To explain why, it is necessary to refer to a recent study completed within our group where fluctuations in E_{intra} were observed to mimic a Buckingham-type potential and hence significantly contribute towards steric hindrance.^[76] As a result, *the behavior of ΔE_{intra} can be viewed as a measure of the steric hindrance.* However, this conclusion is only based on a set of observed correlations, and statistically robust fits to classical Buckingham-type (exponential) potentials, successfully obtained for many small van der Waals complexes. Unpublished results (involving an oligopeptide intra-atomic energy analysis) have also shown that E_{intra} often correlates with the atomic volume of an atom: where energy is stabilized, the atomic volume increases. Thus, the general stabilization of $\Delta E_{\text{intra}}^{\text{Mol}}$ across the energy profile should be interpreted as due to expanding atoms (i.e., relaxing) when the backbone is extending itself. Such extension occurs when φ or ψ values move towards -180° or 180° , that is, further away from the global minimum. To be more specific, of the 49 (=24 + 24 + 1) conformations (see Section 2.1) studied for each dipeptide, only four are sterically destabilized relative to the global minimum

(deducible from Fig. 6, Val/Ile orange curve). At this minimum, there is an electrostatically stabilising intramolecular hydrogen bond, which explains why $\Delta E_{\text{cl}}^{\text{Mol}}$ stabilizes at nearby torsional angles (Fig. 6, Val/Ile purple curve at $\psi = 0^\circ$ and $\psi = 60^\circ$). From this reasoning, we can initially conclude that it is generally not steric hindrance (through hard-sphere clashes) causing the molecule to be less stable in many regions of the Ramachandran plot. Instead, the high energies are caused by a lack of electrostatic stability.

Figures 5 and 6 also show how Gly is more electrostatically destabilized than Val and Ile for many dihedral angles (both φ and ψ) on the energy profile. However, the greater electrostatic destabilization is also accompanied by a greater $\Delta E_{\text{intra}}^{\text{Mol}}$ stabilization, resulting in Val and Ile having the higher $\Delta E_{\text{IQA}}^{\text{Mol}}$, and therefore barriers, at such φ/ψ angles. Hence, IQA confirms that Gly, due to the absence of a side chain, is conformationally less restricted than other amino acids, which is expressed through greater relative stabilization via $\Delta E_{\text{intra}}^{\text{Mol}}$. These results are another example of the prominent relationship between ΔE_{intra} and ΔV_{cl} energies: as one becomes stabilized, the other becomes typically destabilized. This counterbalancing effect is elaborated upon in our recent work^[77] on large water clusters.

Analysis at (functional) group level

Next we observe the functional group behavior. As mentioned in the caption of Figure 1 the following partition will prove useful: $\text{CH}_3|\text{C}(=\text{O}) - \text{N}(\text{H})|\text{C}_\alpha\text{HR}|\text{C}(=\text{O}) - \text{N}(\text{H})|\text{CH}_3$. We introduce the following notation to describe these fragments: (i) the methyl groups are combined and this collection of eight atoms is called "Caps," (ii) the peptide group at the C-terminus (i.e., left, involving the $O_{i-1} - C_{i-1} - N_i - H_i$) is called "Pep-," (iii) the peptide group at the N-terminus (i.e., right, involving the $O_i - C_i - N_{i+1} - H_{i+1}$) is called "Pep+," (iv) the pivotal C_α atom (and one H_α (for Ile/Val) or two H_α (for Gly) atoms bonded to it) is called $(\text{CH})_\alpha$ and finally (v) the sidechain atoms (full chain for Val/Ile only, called "Sidechain"). The energies associated with the five atom groups defined above are denoted respectively: $\Delta E_{\text{IQA}}^{\text{Pep-}}$, $\Delta E_{\text{IQA}}^{\text{Pep+}}$, $\Delta E_{\text{IQA}}^{\text{Caps}}$, $\Delta E_{\text{IQA}}^{(\text{CH})_\alpha}$, and $\Delta E_{\text{IQA}}^{\text{Sidechain}}$. We then sum the $\Delta E_{\text{IQA}}^{\text{A}}$ contribution of each atom forming these given groups to recover the respective group energies. We additionally sum (a) the two peptide groups denoted $\Delta E_{\text{IQA}}^{\text{Peps}\pm}$ and (b) the caps and α -atoms into a single term denoted $\Delta E_{\text{IQA}}^{\text{Caps},(\text{CH})_\alpha}$. This grouping offers an even coarser point of view: peptides, α -pivot and sidechains or $\Delta E_{\text{IQA}}^{\text{Peps}\pm}$, $\Delta E_{\text{IQA}}^{\text{Caps},(\text{CH})_\alpha}$, and $\Delta E_{\text{IQA}}^{\text{Sidechain}}$. Figures 8 and 9 plot the functional group analysis for the φ and ψ scans, respectively.

Figures 8 and 9 illustrate a few interesting points. First, through the newly available functional group energy profiles we can now establish a high degree of transferability, that is, a high similarity between the various energy profiles in both φ and ψ scans. This analysis allows us to dissect the consistency in the molecular energy trends seen in Figures 5 and 6. Second, the α -atoms fluctuate very little across φ and ψ scans within $\pm 15 \text{ kJ mol}^{-1}$. These atoms are at the pivot point of the dihedral rotations and link the backbone to the sidechain.

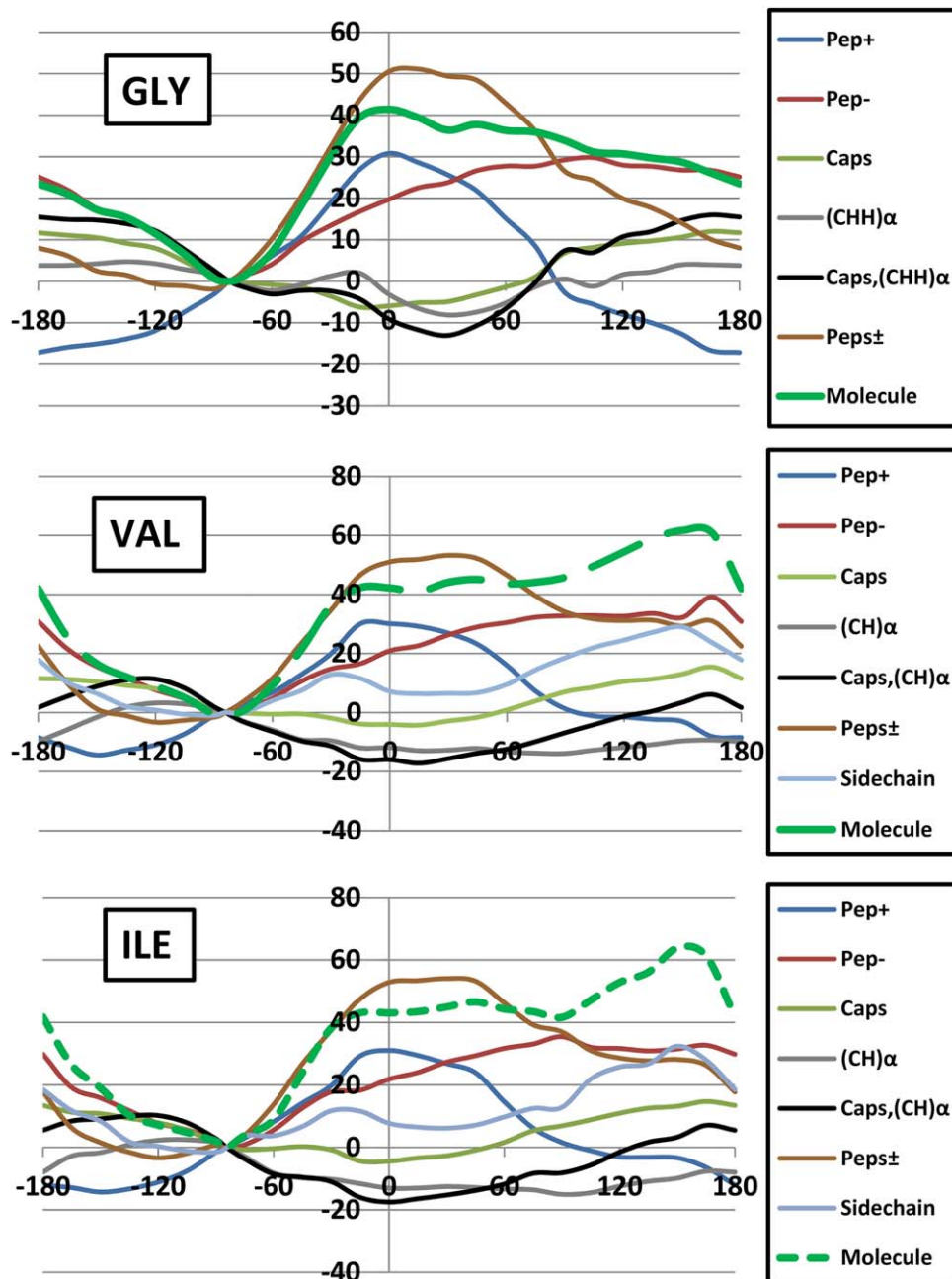


Figure 8. Breakdown of $\Delta E_{\text{IOA}}^{\text{Mol}}$ relative energies (in kJ mol^{-1}) (green) into fragment energies: $\Delta E_{\text{IOA}}^{\text{Pep}^+}$ (blue), $\Delta E_{\text{IOA}}^{\text{Pep}^-}$ (red), $\Delta E_{\text{IOA}}^{\text{Caps}}$ (light green), $\Delta E_{\text{IOA}}^{(\text{CH})\alpha}$ (gray), $\Delta E_{\text{IOA}}^{\text{Caps},(\text{CH})\alpha}$ (black), $\Delta E_{\text{IOA}}^{\text{Peps}\pm}$ (brown), and $\Delta E_{\text{IOA}}^{\text{sidechain}}$ (light blue) components for (a) Gly, (b) Val, and (c) Ile, for the φ scan. The same convention of the green line types (solid, small dash, large dash) has been applied throughout the paper, starting in Figure 3. [Color figure can be viewed at wileyonlinelibrary.com]

One would expect them to be energetically sensitive but this is not the case. Third, regions within the φ and ψ scans can be attributed to certain functional groups. As a remarkable example, we see that *the peptide groups only are responsible for the barriers in the φ scan within an interval approximately stretching from the global minimum to $\varphi = -15^\circ$* . Outside of this interval the barrier results from both the peptide groups and the sidechain atoms. Within the ψ scan, again only the peptide groups are responsible for the barrier right of the global minimum ($\psi > +75^\circ$) in all systems (allowing for discrepancies up to $\sim 5 \text{ kJ mol}^{-1}$). For glycine, this remarkable match extends from

$\psi = -180^\circ$ to $\psi = -45^\circ$. Again, outside these areas, the α -atoms and the sidechain (for Val/Ile) atoms make significant contributions towards the barrier seen at $\psi = +15^\circ$. Collectively, the barriers at $\varphi = +165^\circ$ and $\psi = +15^\circ$ result from important sidechain contributions, confirming our earlier hypothesis on the rationale behind each of these maxima for Val and Ile.

The behavior of each functional group energy (ΔE_{intra} , ΔV_{cl} and ΔV_{xc}), composing the total energy of the fragment (ΔE_{IOA}) may be seen in the Supporting Information in Figures S1 and S3 for φ , and Figures S4 to S6 for ψ . From these plots, it is clear that the peptide groups experience very large

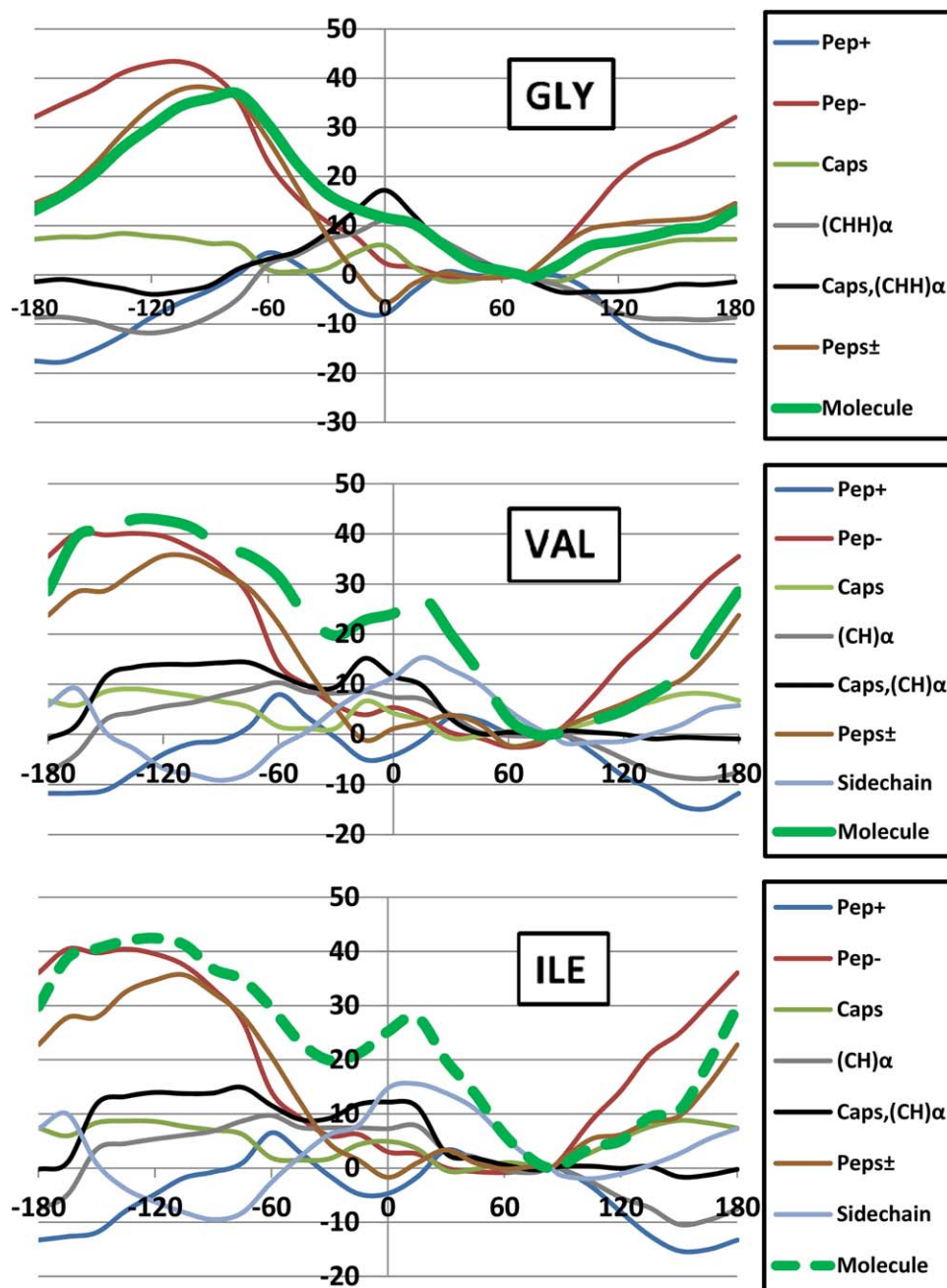


Figure 9. Breakdown of $\Delta E_{\text{IOA}}^{\text{Mol}}$ relative energies (in kJ mol^{-1}) (green) into fragment energies: $\Delta E_{\text{IOA}}^{\text{Pep}^+}$ (blue), $\Delta E_{\text{IOA}}^{\text{Pep}^-}$ (red), $\Delta E_{\text{IOA}}^{\text{Caps}}$ (light green), $\Delta E_{\text{IOA}}^{(\text{CH})\alpha}$ (gray), $\Delta E_{\text{IOA}}^{\text{Caps},(\text{CH})\alpha}$ (black), $\Delta E_{\text{IOA}}^{\text{Peps}\pm}$ (brown), and $\Delta E_{\text{IOA}}^{\text{sidechain}}$ (light blue) components for (a) Gly, (b) Val, and (c) Ile, for the ψ scan. The same convention of the green line types (solid, small dash, large dash) has been applied throughout the paper, starting in Figure 3. [Color figure can be viewed at wileyonlinelibrary.com]

electrostatic and steric fluctuations across both φ and ψ scans. The remaining atom groups fluctuate less dramatically, both electrostatically and sterically. As a result, $\Delta E_{\text{intra}}^{\text{Mol}}$ and $\Delta V_{\text{cl}}^{\text{Mol}}$ energy profiles are accurately (remarkably within a few kJ mol^{-1}) described by profiles of the eight peptide atoms alone where: $\varphi < -60^\circ$ and $\varphi > +150^\circ$, and $-105^\circ < \psi < -15^\circ$ and $\psi > +30^\circ$. However, since $\Delta E_{\text{intra}}^{\text{Pep}\pm}$ and $\Delta V_{\text{cl}}^{\text{Pep}\pm}$ cancel to a large degree, their combined contribution to ΔE_{IOA} is much lower and of a comparable magnitude to that of the other group ΔE_{intra} and ΔV_{cl} contributions and that of $\Delta V_{\text{xc}}^{\text{Pep}\pm}$.

In summary and broadly speaking, *the molecular intra-atomic and electrostatics are remarkably well described by the peptide atoms alone*. However, such remarkable behavior is lost when the intra-atomic and electrostatics are added resulting in a more constrained relative energy range. In other words, the resultant cancellation and concomitant intricate interplay, leads to energy magnitudes similar to those of the remaining atomic groups ($\Delta E_{\text{intra}}^{\text{Mol}}$ and $\Delta V_{\text{cl}}^{\text{Mol}}$) and exchange energy in general (ΔV_{xc}).

Supporting Information Figure S3 shows how ΔV_{xc} is dominated by only one peptide group (Pep+), which is significantly

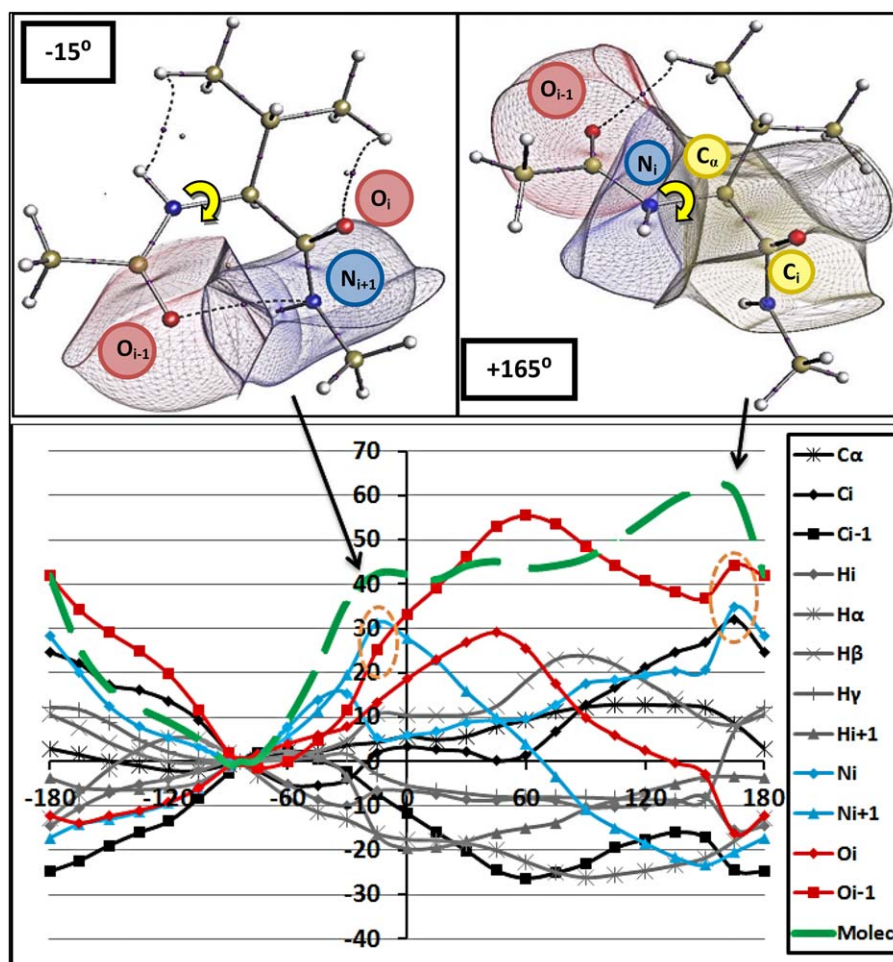


Figure 10. Val with $\varphi = -15^\circ$ (top left) and $\varphi = +165^\circ$ (top right) with key atomic basins depicted and the φ angles marked in yellow. $\Delta E_{\text{QA}}^{\text{A}}$ relative atomic energies (in kJ mol^{-1}) for the φ scan (Bottom) for Valine with only atoms with significant energy fluctuations plotted. Traditional element colors are used for lines and symbols to distinguish element type: carbons (dark gray), hydrogens (light gray), nitrogen (blue), and oxygen (red). Symbols are indicative of the subscript of the element indicating their position in the molecule. $\Delta E_{\text{QA}}^{\text{Mol}}$ energy is given in green. Orange circles depict most destabilized atoms in each barrier region. The same convention of the green line types (solid, small dash, large dash) has been applied throughout the paper, starting in Figure 3. [Color figure can be viewed at wileyonlinelibrary.com]

stabilized throughout each system and in φ scan. Supporting Information Figure S6 shows the same effect for the ψ scan, although not so pronounced. For clarity, Pep+ corresponds to the peptide group with H_{i+1} forming an intra-molecular hydrogen bond with the O_{i-1} and N_i atoms (see Fig. 7).

Analysis at atomic level

As a result of the group partitioning, molecular behavior has been localized to, for example, peptide atoms for certain torsional intervals. The energy profiles have also been rationalized by their electrostatic, steric, and exchange origins. Next, we take our analysis one partitioning step further and observe the energies at the atomic level. At the atomic level, we aim to isolate individual atoms causing the barriers observed within the φ/ψ scans. So far, we have learnt about the consistency of the molecular and group energy profiles across each system. At the more-refined atomic level, we also now expect to see this consistency.

Figures 10 and 11 plot the $\Delta E_{\text{QA}}^{\text{A}}$ energy profiles for key atoms in the Val φ and ψ scans, respectively. Supporting Information Figures S7 and S8 plot the same type of information for the remaining two systems: Gly and Ile. Here, we only report the Val results because it is clear that the atomic trends are very similar throughout each system, within each scan. To clarify, where backbone atomic destabilization is observed within Gly, is it equally present within Val. In addition, by comparing Figures 10 and 11 to Supporting Information Figures S7 and S8, it is clear the Val and Ile plots are almost identical for every torsional angle. Hence, we focus on the general trends using Val as the example.

In Figures 10 and 11, we plot only the key atoms with energy fluctuations greater than $\pm 10 \text{ kJ mol}^{-1}$. Many side-chain and methyl-cap atoms fell below this threshold and are hence not included in these figures. In fact, very few sidechain atoms fluctuate with any significant energy deviations, except for two H_γ atoms. In addition, the C_α atoms also fluctuate very little ($< \pm 8 \text{ kJ mol}^{-1}$) but are included in the plot to demonstrate this key point. In the group analysis, we already

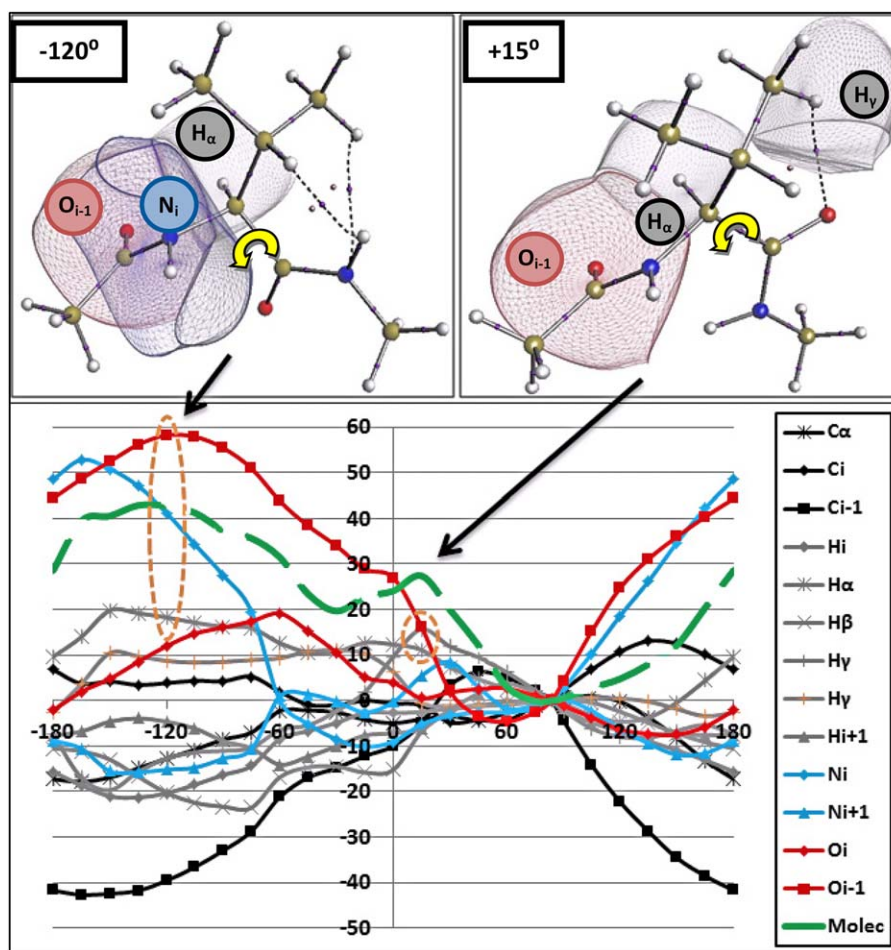


Figure 11. Val with $\psi = -120^\circ$ (top left) and $\psi = +15^\circ$ (top right) with key atomic basins depicted and the ψ angles marked in yellow. $\Delta E_{\text{IOA}}^{\text{A}}$ relative atomic energies (in kJ mol^{-1}) for the ψ scan (Bottom) for Valine with only atoms with significant energy fluctuations plotted. Traditional element colors are used for lines and symbols to distinguish element type: carbons (dark gray), hydrogens (light gray), nitrogen (blue), and oxygen (red). Symbols are indicative of the subscript of the element indicating their position in the molecule. $\Delta E_{\text{IOA}}^{\text{Mol}}$ energy is given in green. Orange circles depict most destabilized atoms in each barrier region. The same convention of the green line types (solid, small dash, large dash) has been applied throughout the paper, starting in Figure 3. [Color figure can be viewed at wileyonlinelibrary.com]

established a lack of energy fluctuation for C_α atoms but we reiterate this surprising result considering C_α 's key bridging role. The O_{i-1} atoms are the most perturbed atoms across all six torsional scans, indicating their importance to the overall molecular stability. In contrast to the destabilizing behavior of the O_{i-1} atoms, the vicinal C_{i-1} atoms significantly stabilize throughout. Within the φ scan of Figure 10, we also see N_i becoming the most destabilized atom when $-90^\circ < \varphi < 0^\circ$ and. In the ψ scan of Figure 11, N_i starts to match O_{i-1} in terms of destabilization magnitude in the vicinity of the $\psi = -120^\circ$ barrier. We learn that the O_{i-1} and N_i atoms dominate the destabilization within each system, across both scans. Many of the remaining atoms fluctuate with some preference towards stabilization or destabilization, or oscillate around the zero-energy line (given by the global minimum).

Figures 10 and 11 also plot the atomic basins for the most destabilizing atoms present at the barriers at $\varphi = -15^\circ$, $\varphi = +165^\circ$, and $\psi = -120^\circ$, $\psi = +15^\circ$, respectively. The literature reasoning behind each barrier will now be further compared with the IQA-based reasoning. Both Mandel et al.^[2] and Ho

et al.^[74] state that the clash between O_{i-1} and N_{i+1} contributes to the barrier at $\varphi = -15^\circ$. The $\Delta E_{\text{IOA}}^{\text{A}}$ analysis shown in Figure 10 confirms this because the largest destabilizing (i.e., positive energy) contributors to this barrier are indeed O_{i-1} and N_{i+1} . However, Mandel et al.^[2] quotes the clash between O_{i-1} and C_i as an extra contributor to this barrier, which we cannot confirm because the positive $\Delta E_{\text{IOA}}^{\text{A}}$ for C_i is an order of magnitude smaller than that of N_{i+1} .

We now analyze the barrier at $\varphi = +165^\circ$ in a similar way. Ho et al.^[74] suggest that a clash between O_{i-1} and C_β causes this barrier. Our analysis confirms that O_{i-1} is indeed a major factor of destabilization (large positive $\Delta E_{\text{IOA}}^{\text{A}}$ value) but C_β is not at all (in fact, because it is always smaller than 4 kJ mol^{-1} is not even shown in Fig. 10). However, if our analysis is forced to point out destabilizing atoms from the side chain then one H_β and one H_γ atom emerge. Much more significant destabilization originates from N_i and C_i . We are now in a position to refine our earlier observation in the molecular energy analysis (see Section "Analysis at molecular level"). Although the side-chain causes the $\varphi = +165^\circ$ barrier, it results from three

Table 1. Summary of depicted atomic basins and their origin of destabilization.

Barrier	Atom	Origin of destabilization
$\varphi = -15^\circ$	O_{i-1}	Steric
	N_{i+1}	Electrostatic
$\varphi = +165^\circ$	O_{i-1}	Steric and Electrostatic
	N_i	Electrostatic and Exchange
	C_i	Steric and Electrostatic
	H_β and $H_\gamma^{[a]}$	Steric
$\Psi = -120^\circ$	O_{i-1}	Steric and Electrostatic
	N_i	Electrostatic
$\Psi = +15^\circ$	O_{i-1}	Electrostatic and Exchange
	H_z	Steric and Exchange
	H_γ	Steric

[a] Their atomic basins are not drawn in Figure 10 but collectively contribute around the barrier.

peptide atoms (O_{i-1} , C_i and N_i) being destabilized alongside two sidechain hydrogen atoms but not sidechain carbons.

Within the ψ scans, we do not see the reported^[74] clash between N_{i+1} and C_β when $\psi = -120^\circ$. Instead, we observe that N_i and O_{i-1} are most destabilized alongside H_z (see Fig. 11). Moreover, the suggested N_{i+1} is actually stabilising at $\psi = -120^\circ$, according to our findings. For the $\psi = +15^\circ$ barrier, which is known to be sterically allowed but without specific clashing atoms identified, we discover that the sidechain (within Val/Ile) is destabilized through H_γ . Indeed, in Figure 11 we see a clear peak at $\psi = +15^\circ$ for H_γ . In addition to H_γ being destabilized, H_z and O_{i-1} are also destabilising when $\psi = +15^\circ$.

Overall, some of our atomic interpretations of energy barriers are quite different to those in previous literature. However, our energies are more complete than those represented by, for example, the hard-sphere model, which only considers the steric-like behavior of an atom. To better understand the nature of the destabilization of each atom, it is necessary to observe the causal energies ($\Delta E_{\text{intra}}^A$, $\Delta V_{\text{cl}}^{AA'}$, and $\Delta V_{\text{xc}}^{AA'}$) composing ΔE_{IQA}^A . The Supporting Information reports each of these three atomic energy profiles in Figures S9 to S11 for the φ scan, and again in Supporting Information Figures S12 to S14 for the ψ scan. Collectively, Supporting Information Figures S9 to S14 allow us to identify the source of destabilization for every atom known to be significantly destabilized (through ΔE_{IQA}^A) at the barrier peaks. The results are summarized in Table 1.

We note an unusual result for two cases: O_{i-1} ($\varphi = +165^\circ$ and $\Psi = -120^\circ$) and C_i ($\varphi = +165^\circ$), where we observe the anomalous combination of both significantly destabilising steric (intra-atomic) and destabilising electrostatic energies. For O_{i-1} in particular, the anomalous lack of cancellation causes the atom to be the most destabilized (through ΔE_{IQA}^A) atom across each of all six torsional scans.

Energy profile consistency has been identified within both the molecular energy analysis and the (functional) group analysis. When a system is partitioned, consistency of energy trends is commonly known as “transferability,” which is a key topic in force field design. If atomic energies (group or single

atoms) are identified as being consistent across systems, then such atomic energies are said to be transferable. Categorising such transferable atoms should become a significant topic within computational chemistry itself.

Finally, we comment further on the transferable energy trends. Some weak trends occur in the backbone atoms when comparing Gly with Val/Ile but they are strengthened when comparing Val and Ile. Within the atomic analysis, Gly and Val atomic energies are similar to within 8 kJ mol^{-1} (while $-165^\circ < \varphi < +15^\circ$) and within 9 kJ mol^{-1} (across all ψ angles). The Val and Ile atomic energies are even closer, within 5 kJ mol^{-1} (while $-165^\circ < \varphi < +165^\circ$) and within 3 kJ mol^{-1} (when $\psi < 0^\circ$ or $\psi > 0^\circ$). The minimal energy discrepancies across Gly, Val, and Ile corroborate fragment transferability, which force field developers need in their atom typing. The results presented also support some of our other work^[34] on IQA and transferability.

Conclusions

In this study, three dipeptides (Gly, Val, and Ile) were investigated to gain a better understanding of the intrinsic behavior of amino acids at three successive levels of detail: molecular, (functional) group, and atomic. The topological energy partitioning method called IQA provided four types of energy to achieve this goal: intra-atomic (self) energy (E_{intra}), electrostatic energy (V_{cl}), exchange(-correlation) energy (V_{xc}), and the sum of all three (E_{IQA}). We determined the causes of the high-energy regions at relevant combinations of φ/ψ in the Ramachandran plots.

At molecular level, a destabilization of the electrostatic energy is the cause of the barrier regions across both φ and ψ scans, and across each dipeptide system. However, each electrostatic barrier is dampened by counter-stabilization from $\Delta E_{\text{intra}}^{\text{Mol}}$ and $\Delta V_{\text{xc}}^{\text{Mol}}$. Electrostatics dictating the barriers is an unexpected conclusion given the prevailing view that steric hindrance can explain the Ramachandran regions.

At atom-group level, the peptide groups are consistently the cause of the barriers at $\varphi = -15^\circ$ and $\psi = -120^\circ$, with the barriers at $\varphi = +165^\circ$ and at $\psi = +15^\circ$ arising as the result of both the peptides and sidechain groups becoming destabilized, cooperatively.

At atomic level (A), the aforementioned group trends were reflected in destabilized ΔE_{IQA}^A energies for key peptide atoms (O_{i-1} , C_i , N_i , and N_{i+1}) and some sidechain hydrogen atoms (H_β and H_γ).

The origin of the atomic destabilization was also clarified through the analysis of the $\Delta E_{\text{intra}}^A$, $\Delta V_{\text{cl}}^{AA'}$, and $\Delta V_{\text{xc}}^{AA'}$ energies (A' is the atomic environment of A), confirming some steric destabilization within the O_{i-1} , C_i , H_z and sidechain H_γ and H_β atoms at barrier peaks. Surprisingly and interestingly, the energies of the sidechain carbon atoms (C_β , C_γ , C_δ), and more importantly C_α , remained relatively unperturbed throughout.

Finally, some very promising results regarding transferability were observed where absolute values of atomic energies are smaller than 9 kJ mol^{-1} between Gly and Val/Ile, and smaller


than 5 kJ mol⁻¹ between Val and Ile, for the majority of torsional angles across both φ/ψ scans.

Acknowledgment

We thank Dr Todd Keith for useful comments on virial-based atomic energies.

Keywords: Quantum Chemical Topology (QCT) · QTAIM · rotation barrier · peptides · IQA

How to cite this article: P. I. Maxwell, P. L. A. Popelier. *J. Comput. Chem.* **2017**, *38*, 2459–2474. DOI: 10.1002/jcc.24904

 Additional Supporting Information may be found in the online version of this article.

- [1] G. N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, *J. Mol. Biol.* **1963**, *7*, 95.
- [2] N. Mandel, G. Mandel, B. L. Trus, J. Rosenberg, G. Carlson, R. E. Dickerson, *J. Biol. Chem.* **1977**, *252*, 4619.
- [3] C. Halkides, *J. Chem. Educ.* **2013**, *90*, 760.
- [4] M. C. Childers, C. L. Towse, V. Daggett, *Protein Eng. Des. Sel.* **2016**, *29*, 271.
- [5] D. A. C. Beck, D. O. V. Alonso, D. Inoyama, V. Daggett, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 12259.
- [6] R. Schweitzer-Stenner, *Mol. BioSyst.* **2012**, *8*, 122.
- [7] C. L. Towse, G. Hopping, I. Vulovic, V. Daggett, *Protein Eng. Des. Sel.* **2014**, *27*, 447.
- [8] R. Schweitzer-Stenner, *J. Phys. Chem. B* **2009**, *113*, 2922.
- [9] R. Schweitzer-Stenner, S. E. Toal, *Phys. Chem. Chem. Phys.* **2014**, *16*, 22527.
- [10] S. Toal, R. Schweitzer-Stenner, *Biomolecules* **2014**, *4*, 725.
- [11] M. Feig, *J. Chem. Theory Comput.* **2008**, *4*, 1555.
- [12] F. Jiang, W. Han, Y.-D. Wu, *Phys. Chem. Chem. Phys.* **2013**, *15*, 3413.
- [13] X. Zhu, P. E. M. Lopes, J. Shim, A. D. J. MacKerell, *J. Chem. Inf. Model.* **2012**, *52*, 1559.
- [14] M. Iwaoaka, D. Yosida, N. Kimura, *J. Phys. Chem. B* **2006**, *110*, 14475.
- [15] R. G. Best, N. V. Buchete, G. Hummer, *Biophys. J.* **2008**, *95*, L07.
- [16] P. Krupa, A. K. Sieradzian, S. Rackovsky, M. Baranowski, S. Oldziej, H. A. Scheraga, A. Liwo, C. Czaplowski, *J. Chem. Theory Comput.* **2013**, *9*, 4620.
- [17] F. Jiang, C.-Y. Zhou, Y.-D. Wu, *J. Phys. Chem. B* **2014**, *118*, 6983.
- [18] A. K. Sieradzian, P. Krupa, H. A. Scheraga, A. Liwo, C. Czaplowski, *J. Chem. Theory Comput.* **2015**, *11*, 817.
- [19] M. J. Robertson, J. Tirado-Rives, W. L. Jorgensen, *J. Chem. Theory Comput.* **2015**, *11*, 3499.
- [20] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins Struct. Funct. Bioinformatics* **2010**, *78*, 1950.
- [21] Y. Gao, Y. Li, L. Mou, B. Lin, J. Z. H. Zhang, Y. Mei, *Sci. Rep.* **2015**, *5*, 10359.
- [22] S. E. Feller, A. D. MacKerell, *J. Phys. Chem. B* **2000**, *104*, 7510.
- [23] Y. Sakae, Y. Okamoto, *J. Phys. Soc. Jpn.* **2006**, *75*, 054802.
- [24] D. Toroz, T. van Mourik, *Phys. Chem. Chem. Phys.* **2010**, *12*, 3463.
- [25] M. D. Beachy, D. Chasman, R. B. Murphy, T. A. Halgren, R. A. Friesner, *J. Am. Chem. Soc.* **1997**, *119*, 5908.
- [26] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr.; D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- [27] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, *J. Comput. Chem.* **2003**, *24*, 1999.
- [28] G. A. Kaminsky, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem. B* **2001**, *105*, 6474.
- [29] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. I. I., W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
- [30] A. D. MacKerell, M. Feig, C. L. Brooks, III, *J. Am. Chem. Soc.* **2004**, *126*, 698.
- [31] N. O. J. Malcolm, P. L. A. Popelier, *Faraday Discuss.* **2003**, *124*, 353.
- [32] T. L. Fletcher, P. L. A. Popelier, *J. Comput. Chem.* **2017**, *38*, 336.
- [33] T. L. Fletcher, P. L. A. Popelier, *Theor. Chem. Acc.* **2015**, *134*, 1.
- [34] P. Maxwell, P. L. A. Popelier, *Mol. Phys.* **2016**, *114*, 1304.
- [35] Y. Yuan, M. J. L. Mills, P. L. A. Popelier, F. Jensen, *J. Phys. Chem. A* **2014**, *118*, 7876.
- [36] M. Rafat, M. Devereux, P. L. A. Popelier, *J. Mol. Graph. Modell.* **2005**, *24*, 111.
- [37] M. Rafat, P. L. A. Popelier, *J. Comput. Chem.* **2007**, *28*, 2602.
- [38] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09; Gaussian, Inc.:* Wallingford, CT, **2009**.
- [39] F. Jensen, *J. Chem. Phys.* **2002**, *117*, 9234.
- [40] T. A. Keith, AIMAll; TK Gristmill Software: Overland Park, KS, **2016**. Available at: aim.tkgristmill.com. Accessed: 1 April 2017.
- [41] M. A. Blanco, A. Martín Pendas, E. Francisco, *J. Chem. Theory Comput.* **2005**, *1*, 1096.
- [42] R. F. W. Bader, *Atoms in Molecules. A Quantum Theory*; Oxford University Press: Oxford, Great Britain, **1990**.
- [43] P. L. A. Popelier, *Atoms in Molecules. An Introduction*; Pearson Education: London, Great Britain, **2000**.
- [44] P. L. A. Popelier, In *The Nature of the Chemical Bond Revisited*; G. Frenking, S. Shaik, Eds.; Weinheim, Germany: Wiley-VCH, **2014**; Chapter 8, p. 271.
- [45] B. Silvi, A. Savin, *Nature (London)* **1994**, *371*, 683.
- [46] P. L. A. Popelier, In *Structure and Bonding. Intermolecular Forces and Clusters*, Vol. 115; D. J. Wales, Ed.; Springer: Heidelberg, Germany, **2005**; p. 1.
- [47] P. L. A. Popelier, In *The Chemical Bond - 100 Years Old and Getting Stronger*; Mingos, M., Ed.; Springer: Switzerland, **2016**; p. 71.
- [48] P. L. A. Popelier, *Curr. Top. Med. Chem.* **2012**, *12*, 1924.
- [49] N. Schneider, G. Lange, S. Hindle, R. Klein, M. Rarey, *J. Comput. Aided Mol. Des.* **2013**, *27*, 15.
- [50] P. L. A. Popelier, *Int. J. Quantum Chem.* **2015**, *115*, 1005.
- [51] P. L. A. Popelier, *Phys. Scr.* **2016**, *91*, 033007.
- [52] A. Bartok, M. C. Payne, R. Kondor, G. Csanyi, *Phys. Rev. Lett.* **2010**, *104*, 136403.
- [53] P. Maxwell, N. di Pasquale, S. Cardamone, P. L. A. Popelier, *Theor. Chem. Acc.* **2016**, *135*, 195.
- [54] I. Cukrowski, J. H. de Lange, M. Mitoraj, *J. Phys. Chem. A* **2014**, *118*, 623.
- [55] T. I. Madzhidov, G. A. Chmutova, A. M. Pendas, *J. Phys. Chem. A* **2011**, *115*, 10069.
- [56] T. M. Klapoetke, B. Krumm, R. Moll, S. F. Rest, Y. V. Vishnevskiy, C. Reuter, H.-G. Stammer, N. W. Mitzel, *Chem. Eur. J.* **2014**, *20*, 1.
- [57] K. Eskandari, M. Lesani, *Chem. Eur. J.* **2015**, *21*, 4739.
- [58] O. A. Syzgantseva, V. Tognetti, L. Joubert, *J. Phys. Chem. A* **2013**, *117*, 8969.
- [59] J. M. Guevara-Vela, R. Chavez-Calvillo, M. Garcia-Revilla, J. Hernandez-Trujillo, O. Christiansen, E. Francisco, A. M. Pendas, T. Rocha-Rinza, *Chem. Eur. J.* **2013**, *19*, 14304.
- [60] A. Martín Pendas, M. A. Blanco, E. Francisco, *J. Chem. Phys.* **2006**, *125*, 184112.
- [61] K. Eskandari, C. Van Alsenoy, *J. Comput. Chem.* **2014**, *35*, 1883.

- [62] J. Dillen, *Int. J. Quantum Chem.* **2013**, *113*, 2143.
- [63] A. M. Pendas, M. A. Blanco, E. Francisco, *J. Comput. Chem.* **2009**, *30*, 98.
- [64] D. Ferro-Costas, R. A. Mosquera, *Phys. Chem. Chem. Phys.* **2015**, *17*, 7424.
- [65] R. Inostroza-Rivera, M. Yahia-Ouahmed, V. Tognetti, L. Joubert, B. Herrera, A. Toro-Labbe, *Phys. Chem. Chem. Phys.* **2015**, *17*, 17797.
- [66] M. J. S. Phipps, T. Fox, C. S. Tautermann, C.-K. Skylaris, *Chem. Soc. Rev.* **2015**, *44*, 3177.
- [67] P. Maxwell, A. Martin Pendas, P. L. A. Popelier, *Phys. Chem. Chem. Phys.* **2016**, *18*, 20986.
- [68] V. Tognetti, L. Joubert, *Phys. Chem. Chem. Phys.* **2014**, *16*, 14539.
- [69] E. Francisco, J. L. Casals-Sainz, T. Rocha-Rinza, A. M. Pendas, *Theor. Chem. Acc.* **2016**, *135*, 1.
- [70] F. J. Holguin-Gallego, R. Chavez-Calvillo, M. Garcia-Revilla, E. Francisco, A. M. Pendas, T. Rocha-Rinza, *J. Comput. Chem.* **2016**, *37*, 1753.
- [71] R. Chávez-Calvillo, M. García-Revilla, E. Francisco, A. Martín Pendás, T. Rocha-Rinza, *Comput. Theor. Chem.* **2015**, *1053*, 90.
- [72] J. L. McDonagh, M. A. Vincent, P. L. A. Popelier, *Chem. Phys. Lett.* **2016**, *662*, 228.
- [73] R. P. Bywater, V. Veryazov, *Springerplus* **2015**, *4*, 668.
- [74] B. K. Ho, A. Thomas, R. Brasseur, *Protein Sci.* **2003**, *12*, 2508.
- [75] A. M. Pendas, E. Francisco, M. A. Blanco, *Faraday Discuss.* **2007**, *135*, 423.
- [76] A. Wilson, P. L. A. Popelier, *J. Phys. Chem. A* **2016**, *120*, 9647.
- [77] S. J. Davie, P. I. Maxwell, P. L. A. Popelier, *Phys. Chem. Chem. Phys.* **2017**, *19*, 20941.

Received: 12 April 2017

Revised: 10 July 2017

Accepted: 18 July 2017

Published online on 25 August 2017