Published in final edited form as:
Proteins. 2008 June ; 71(4): 1637–1646. doi:10.1002/prot.21845.

Using Quantum Mechanics to Improve Estimates of Amino Acid Side Chain Rotamer Energies

P. Douglas Renfrew, **Glenn Butterfoss**, and **Brian Kuhlman**
Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516, Email: bkuhlman@email.unc.edu

Abstract

Amino acid side chains adopt a discrete set of favorable conformations typically referred to as rotamers. The relative energies of rotamers partially determine which side chain conformations are more often observed in protein structures and accurate estimates of these energies are important for predicting protein structure and designing new proteins. Protein modelers typically calculate side chain rotamer energies by using molecular mechanics (MM) potentials or by converting rotamer probabilities from the protein database (PDB) into relative free energies. One limitation of the knowledge-based energies is that rotamer preferences observed in the PDB can reflect internal side chain energies as well as longer-range interactions with the rest of the protein. Here, we test an alternative approach for calculating rotamer energies. We use three different quantum mechanics (QM) methods (second order Moller-Plesset (MP2), density functional theory (DFT) energy calculation using the B3LYP functional, and Hartree-Fock) to calculate the energy of amino acid rotamers in a dipeptide model system, and then use these pre-calculated values in side chain placement simulations. Energies were calculated for over 35,000 different conformations of leucine, isoleucine and valine dipeptides with backbone torsion angles from the helical and strand regions of the Ramachandran plot. In a subset of cases these energies differ significantly from those calculated with standard molecular mechanics potentials or those derived from PDB statistics. We find that in these cases the energies from the QM methods result in more accurate placement of amino acid side chains in structure prediction tests.

Keywords

Computational Protein Design; Rotamers; Torsion Energies; Protein Structure Prediction

Introduction

Amino acid side chains adopt a variety of conformations. An accurate estimate of the relative energies of different side chain conformations is essential for high resolution structure prediction, protein design and modeling protein dynamics. These energies are generally calculated by using molecular mechanics (MM) potentials or by deriving energies from the probability of observing a particular side chain conformation in the PDB [1,2]. Most MM potentials use empirically derived functions to model the energetics of bond stretching, bending and torsion angle perturbation [3]. Non-bonded interactions are generally modeled with a Lennard-Jones potential and a form of Coulomb's potential to model electrostatics. MM potentials are often parameterized to match results from quantum mechanics (QM) calculations on model compounds. The advantage of MM potentials is that they are generalizable to a variety

Correspondence to: Brian Kuhlman.

of atom types, they are fast to evaluate and the same force field can be applied throughout a molecule. For instance, the same MM expressions can be used to model energetics within an amino acid side chain as between side chains. A limitation of MM potentials is that the calculated energies are sensitive to the model systems used to parameterize them [4]. Different MM potentials often give different answers when evaluating the same set of molecules, for instance, producing a MM potential that accurately represents the torsional preferences of a peptide backbone has proven to be difficult [5-7].

A common alternative to MM potentials are knowledge-based energy functions. Comparative analysis of amino acid side chains in protein structures has shown that most side chains only adopt a limited set of conformations, typically referred to as rotamers [8]. Additionally, some rotamers of an amino acid are observed more often than others, suggesting that the internal energies of the various rotamers are not equal. Using protein structures from the PDB, databases have been constructed, commonly referred to as rotamer libraries, that specify the most commonly observed torsion angles associated with each rotamer of an amino acid, and the frequency that the various rotamers are observed in the protein database, most recently reviewed by Dunbrack[9]. Because rotamer probabilities depend on the local environment of a side chain, the probabilities are often measured as a function of the backbone dihedral angles of a residue, or as a function of secondary structure [10]. Rotamer probabilities are typically converted to energy by assuming Boltzmann sampling and taking the logarithm of the probability[11]. This assumption was supported in one case by showing that the relative favorability of different methionine rotamers as determined by high level quantum mechanics simulations matches the preference of methionine to adapt a particular rotamer in the PDB[12]. Knowledge-based torsional preferences have been used with good success to predict the conformations of amino acid side chains and to design new protein structures and functions [13,14].

However, there are situations in which a knowledge-based approach may lead to an inaccurate estimate of protein energy. Particularly challenging is making sure that the knowledge-based term does not represent energies that are included in other terms in the energy function. For example, many modeling programs use a Lennard-Jones (LJ) potential evaluated between pairs of atoms to model van der Waals forces and steric repulsion. If the LJ potential is evaluated between pairs of atoms that contribute to rotamer probabilities observed in the PDB, then there will be double counting. In some cases it is clear which atom pairs to ignore to prevent double counting; if rotamer statistics are being used to evaluate side chain preferences than atom pairs within a side chain should not be considered. It is less clear if atom pair energies should be considered with backbone atoms in the neighboring residue. It will depend in part if the rotamer statistics are compiled as a function of the protein backbone dihedral angles. Because the amide group of the following residue (i+1) and the carbonyl group of the preceding residue (I -1) are determined by the phi and psi angles of the central residue (i), it can be argued that atom pair energies should not be calculated between these groups and the side chain of i. Potentially even more subtle are longer range interactions commonly observed in protein secondary structure. In a helix, the preferred side chain conformation at residue i is determined in part by interactions with the residue at positions i-3 and i-4, and therefore, the energetics of this interaction is folded into rotamer statistics of helical residues from the PDB.

Instead of using MM potentials or knowledge-based potentials to calculate protein energetics, an alternative approach is to use direct quantum mechanics (QM) calculations. Energies from QM calculations have been shown to more accurately reproduce backbone and side chain dihedral preferences in the PDB[5,12]. A crucial limitation of QM is that in general it can not be applied to full-sized proteins, and even for a single amino acid most QM simulations require on the order of minutes to perform. For this reason it is not feasible to perform a QM simulation on every structure that is created during a protein design simulation or a molecular dynamics

simulation. However, because only a limited set of side chain conformations are observed during a protein simulation, it is possible to precompute the energy of a side chain in various conformations with QM, and than use these energies during protein simulations. Here, we explore this approach by precomputing energies of ~35,000 conformations of valine, isoleucine and leucine with QM calculations, and then test these energies in side chain prediction tests on full-size proteins. We find that in situations where knowledge-based potentials are more likely to double count or miscount interactions, that the QM energies provide more accurate side chain predictions.

Materials and Methods

Dipeptides

To calculate the internal energies of amino acid side chain rotamers QM and MM calculations were performed on amino acid dipeptides (ACE-X-NME, where X is the amino acid being tested) (figure 1). The dipeptide is commonly used to probe side chain energetics because the relative positions of all the atoms in a dipeptide are primarily determined by phi, psi and the side chain chi angle of a single residue. Backbone and side chain dihedral angles were fixed to their desired values during the calculations. Backbone dihedral angles were sampled combinatorially in regions of phi/psi space that correspond to α-helical and β-strand conformations (α: phi = -70 to -40 and psi = -50 to -20, β: phi = -110 to -160 and psi = 110 to 160) in ten degree intervals. Chi angles were sampled at their canonical angles (-60, 60, 180) and ±10, ±20, and ±30 degrees; resulting in 336 valine-α structures, 7056 isoleucine-α/leucine-α structures, 504 valine-β structures, and 10584 isoleucine-β/leucine-β structures. When referring to the various rotamers we use the nomenclature established in Lovell *et al.* [10] where “m” is minus gauche (-g, ~-60), “p” is plus gauche (+g, ~+60), and “t” is trans (t, ~180).

MM

Molecular mechanics simulations on the constrained dipeptides were carried out using the CHARMM force field (version 22) [15] and Cedar molecular mechanics force fields as implemented in the molecular mechanics package Sigma [16,17]. The phi, psi, and chi dihedral angles of each dipeptide were constrained using a 1000 kcal / mol force. To optimize bond angles, bond lengths, and unconstrained dihedrals, structures were put through 2000 rounds of conjugate gradient minimization. Amber version 9 with the FF99 force field was also used to evaluate the energies of the dipeptides. As with the Cedar and CHARMM methods, the phi, psi, and chi dihedrals angles were constrained using a 5000 kcal / mol force and structure optimization was done with 5000 cycles of conjugate gradient minimization. The dihedral constraint energy was not included in the final calculated energies.

QM

Quantum mechanics calculations were carried out using Gaussian03 from Gaussian Inc.[18]. Energies were calculated by first performing a Hartree Fock (HF) minimization followed by a second order Moller-Plesset (MP2) energy calculation and a density functional theory (DFT) energy calculation using the B3LYP functional. In addition to the MP2 and DFT energies the final energy from the HF minimization was also used in the tests described below. All calculations were performed with the 6-31G(d) basis set except where noted. The HF minimization is the slowest step in this process. To shorten the time of the calculations and to prevent large clashes that can occur when the dihedral angles of a starting structure are rotated and fixed, the starting structure used for each minimization varied by only ten degrees in either phi, psi or one of the chi dihedrals from the target set of angles. For example a minimized valine dipeptide with phi = -60, psi = -40, chi1 = -60, would be allowed to serve as a starting structure for ((-70 or -50), -40, -60), (-60, (-50 or -30), -60), and (-60, -40, (-70 or -50)).

Each class of calculations (valine-α, valine-β, isoleucine-α, isoleucine-β, leucine-α, leucine-β) had one set of phi/psi values tested with a larger basis (6-31+G(d), 6-311+G(d)) set to see if increasing the size of the basis set lead to improvements in the rotamer prediction benchmarks. Only the leucine α class of dipeptides showed improvement with an increased basis set (6-31+G(d)) and the entire phi/psi range was rerun using this larger basis set.

Calculations were performed on either a IBM P690 Model 681 running AIX or an SGI Altix 3700bx2 running RedHat Enterprise Linux 3 maintained by the UNC Information Technology Services (http://its.unc.edu) (both) or the National Center for Supercomputing Applications (IBM P690). Calculations on either machine take ~1 hour of CPU time per structure with the 6-31G(d) basis set and ~3 hours using the 6-31+G(d) basis set.

Knowledge-based rotamer energies

Knowledge-based rotamer energies were computed using the protein modeling program Rosetta:

$$E_{\text{rotamer}}(\text{rot},\text{aa},\varphi,\psi,\vec{\chi}) = -RT\ln\left[P_{\text{chi}}(\vec{\chi}\,|\text{rot},\text{aa},\varphi,\psi) * P_{\text{rot}}(\text{rot}|\text{aa},\varphi,\psi)\right] \quad (1)$$

where P_{chi} is the probability that a particular rotamer will have a certain set of chi angles ($\vec{\chi}$), P_{rot} is the probability that, given phi and psi, a particular amino acid (*aa*) will adopt a particular rotamer (*rot*). P_{rot} is taken directly from Dunbrack's most recent rotamer library (http://dunbrack.fccc.edu/). P_{chi} is determined using the standard deviations included in Dunbrack's library assuming each side chain torsion angle (χ1, χ2, …) is independent of the other torsion angles:

$$P_{\text{chi}}(\vec{\chi},\text{rot},\text{aa},\varphi,\psi) = \prod_{i}^{\max\_\chi} P(\chi(i)|\text{rot},\text{aa},\varphi,\psi) \quad (2)$$

$$P\left(\chi(i)|\text{rot},\text{aa},\varphi,\psi\right) = \left(\frac{1}{\sqrt{2\pi}\sigma_{\chi}}\right)\exp\left(\frac{(\chi(i)-\bar{\chi}(i,\text{rot},\text{aa},\varphi,\psi))^2}{2*\sigma_{\chi}^2}\right) \quad (3)$$

Where χ (*i*) is the torsion angle for the *i*th chi angle, $\bar{\chi}$, is the average value for that chi angle for a particular rotamer, and σ_{χ} is the standard deviation for that chi angle for the same rotamer.

Side chain prediction tests

To determine the usefulness of the different methods for calculating the relative energies of amino acid rotamers, we tested them to see how accurately they could reproduce native side chain conformations from a set of ~2800 protein structures with a resolution not higher than 2.0 angstroms [19]. In these tests the side chain of a residue was removed and rebuilt with Rosetta in the context of the whole protein. Neighboring residues were held fixed. The energy of each rotamer in the context of the whole protein was calculated by adding the intrinsic energy of the rotamer, as determined by the theoretical calculations on the dipeptides, to the standard Rosetta energy. Because the theoretical calculations were performed for 10 degree increments of phi, psi and chi angles, linear interpolation was used to estimate the energy for a specific set of torsion angles. The knowledge-based term usually used to evaluate internal rotamer

preferences was removed from the Rosetta energy function except in the cases in which it was being tested. The lowest energy rotamer in the context of the whole protein was taken as the Rosetta prediction. The test was performed for all valine, isoleucine and leucine residues with phi and psi angles in the range covered by the QM simulations. Each side chain was sampled at its most probable chi angle (as given by Dunbrack's backbone dependent rotamer library) as well a chi angles that varied ±0.5, ±1, ±1.5, and ±2 standard deviations away from the mean (again as given by Dunbrack's backbone dependent library). This results in 27 rotamers for each valine and 729 rotamers for isoleucine and leucine.

To insure that the position of the side chain was well-defined in the crystal structure, residues were only used for the side-chain replacement test if all atoms of the side chain in the crystal structure had B-factors less than 20. In the case of leucine rotamers, if the native conformation in the crystal structure was one of the commonly mistakenly assigned mp* and tt* rotamers [10], the position was omitted.

Analysis of side chain prediction results

A number of statistics were gathered to determine how well side chain conformations were predicted.

Percent Total correct—The percent of residue positions where the side chain conformation was correctly predicted. A prediction was considered correct if all chi angles in the predicted side chain were within the same torsional basin as the native side chain.

Percent Correct and Chi Free—The percent of residue positions where the rotamer was correctly predicted given that the position was “free.” A residue is considered free if the preferred side chain conformation is not primarily determined by repulsive interactions with neighboring residues. We define a position to be free if the repulsive energy as computed by Rosetta between the side chain and neighboring residues is less than 0.5 kcal / mol for at least two alternate side chain conformations, where the conformations differ by more than 60 degrees in at least one of their chi angles. For Percent Correct and Chi1 Free, a position is only considered free if 2 conformations with low repulsive energies have chi1 angles that differ by more than 60 degrees.

Percent Minimum Energy and Closest Chi—The percent of positions where the dihedral angle with the lowest energy is the closest to that of the native angle of all the dihedrals tested.

Standard Rosetta Energy Function

The Rosetta energy function has been described previously [20]. Directly relevant to this study is the 12-6 Lennard-Jones potential that is used to evaluate van der Waals forces and steric repulsion. This potential is evaluated between most pairs of atoms in the protein. It is not evaluated between atoms within a residue. In addition it is not evaluated between the amide group and c_α of residue i+1 and the atoms in i, and it is not evaluated between the carbonyl group and c_α of the preceding residue (i-1) and the atoms in i. These interactions are left out because these interactions should be accounted for by backbone dependent rotamer energies derived from PDB statistics. These interactions with the neighboring backbone atoms will also contribute to the energies calculated for the dipeptides, and therefore it is appropriate that these energies are not included in the Rosetta Lennard-Jones calculations. Explicit hydrogens are modeled on all atoms, but they are only used to check for steric overlap and only contribute to the energy of the protein when they have Lennard-Jones energies that are greater than zero. The van der Waals radii and Lennard-Jones well depths have been described previously [20].

Results

QM and MM energy calculations were performed on dipeptides of valine, isoleucine and leucine with a variety of side chain conformations and phi and psi angles from either the α-helical or β-strand regions of the Ramachandran plot. In many cases, the QM energies from the final step of the HF minimization or the MP2 or DFT energy calculations were significantly different from those calculated with the CHARMM22, Cedar or Amber force fields. For example, for a valine dipeptide with a phi of -60° and a psi of -40° the m (chi1 ~ -60°) rotamer is predicted by the CHARMM22 force field to be -0.8 kcal / mol more favorable than the p rotamer (chi 1 ~ 60°) (figure 2). QM calculations at the MP2 level predict the opposite; the p rotamer is predicted to be -0.6 kcal / mol more favorable than the m rotamer. These are significant differences when one considers that proteins are only stable by a few kcals / mol. In some cases, the most preferred chi angle for each rotamer also differed between the QM and MM simulations. For the valine dipeptide the QM calculations of the HF energy preferred a chi1 near 170° for the t rotamer while the CHARMM22 force field favors a chi1 near 190° (figure 2). A complete list of calculated energies is provided in the supplementary material.

Energy calculations with dipeptides that have different phi and psi angles highlight the importance of interactions between the side chain and the local backbone. The relative energies of the rotamers often shift dramatically with just small changes in one of the backbone dihedral angles: for valine with a phi -50° and a psi of -30° the QM calculations (MP2) predict that the t rotamer is 0.8 kcal / mol less favorable than the m rotamer, when psi is shifted to -50° the situation is reversed and the t rotamer is predicted to be 1.1 kcal / mol more favorable than the m rotamer (figure 3, table 1). This dramatic change with such a small change in psi reflects interactions between the backbone carbonyl oxygen and the side chain methyl groups on valine. In general when the backbone torsion angles are varied, the energies calculated with the 3 MM potentials follow the same trends observed with the 3 QM calculations. The strong dependence of rotamer energies on phi and psi indicates that if precomputed rotamer energies are to be used during protein simulations, they should be calculated as a function of phi and psi, and phi and psi should be sampled at least every 10 degrees.

To compare the QM and MM energies with rotamer statistics from the PDB, the energies were converted to rotamer probabilities assuming a Boltzmann distribution and a temperature of 298 K (figure 4). Overall agreement between two methods for a single amino acid and backbone conformation was measured by computing the root mean square deviation between the probabilities of observing each rotamer (table II). The biggest differences between the PDB statistics and the theoretical methods occur for valine and isoleucine with helical phi and psi angles. Unlike most amino acids, valine and isoleucine are β-branched, i.e. there are two non-hydrogen side chain atoms bonded to the C_β atom. When a valine or isoleucine is in a α-helix there is only one chi 1 rotamer it can adopt and avoid a clash between its C_β groups and the carbonyl oxygen on residue i-3. This restriction on chi 1 is evident in the PDB statistics (figure 4), but is absent from the theoretical calculations that were performed in the context of a dipeptide. This provides a clear example of a case where double counting will occur if PDB statistics are used in combination with Lennard-Jones energies when calculating the energy of residues i and i-3. Tables showing the rotamer probabilities binned by phi and psi dihedral angles for all of the theoretical methods are provided in the supplementary material.

Aside from valine and isoleucine in the helical region, the energies calculated with QM and MM match reasonably with those derived from PDB statistics, although there are some specific cases where the MM potentials deviate significantly. The Amber potential favors the TP rotamer over the MT rotamer for leucine when it has helical torsion angles, but the MT rotamer is more commonly observed in the PDB. The Cedar potential strongly favors the trans rotamer for valine when it has β backbone angles, but this is the least common rotamer in the PDB. Not

surprisingly, the aforementioned potentials perform poorly in side chain prediction tests for the regions of Ramachandran space in which they deviated from the QM and the PDB statistics.

Side chain prediction tests

We have shown several examples that demonstrate that the three different approaches, QM, MM and knowledge-based, give significantly different energies for many side chain rotamers. To determine which of these potentials more accurately represents the internal energy of amino acid residues, we performed side chain prediction tests with the Rosetta protein modeling program. In these tests a single side chain was removed from a residue in a protein, and Rosetta was used to predict the conformation of the removed side chain. The prediction was performed by cycling through all rotamers and sub-rotamers of the missing amino acid and choosing the one with the lowest energy. The energy function was a linear sum of the internal energy of the rotamer, as calculated by the QM, MM or knowledge-based potential, and long range interactions between the rotamer and its neighbors calculated with the standard Rosetta energy function. Neighboring residues were held fixed in this test because QM energies are only available for valine, isoleucine and leucine. As a control, tests were also performed in which each rotamer of an amino acid was assumed to have equal internal energy (flat). The side chain prediction test was performed on 5360 valine-α, 6377 valine-β, 7569 leucine-α, 2546 leucine-β, 4278 isoleucine-α and 3928 isoleucine-β positions in over 2800 proteins. The predictions were analyzed to determine how often the correct rotamers were predicted (i.e. the correct torsional wells), and how close the chi angles were to the native chi angles as described in the materials and methods section.

Overall, all of the methods do well in the side chain prediction test; all of them predict the correct rotamer at more than 90% of the positions. This result was expected because at most sequence positions only one rotamer can fit without clashing with the neighboring residues, and the energy from a clash will overwhelm the internal energies of the amino acids. Indeed, in the tests without any internal energy for the side chain the correct rotamer was predicted over 85% of the time. This does not indicate that the internal rotamer energies are unimportant. This test is artificial in that we are keeping all the neighbors fixed as well as the protein backbone. In a full protein simulation all backbone positions and side chains are free to vary and changes in 1 kcal / mol as a side chain moves to a new rotamer are certainly important. To make the test more discriminatory, we focused on sequence positions at which the correct rotamer was not specified by simply looking for clashes with neighboring residues. If a side chain could adopt two rotamers that had a predicted clash score of less than 0.5 kcal / mol and differed by more than 60° at chi 1, than that position was included in our refined test. Because isoleucine, valine, and leucine are often found in the interior of a protein, this filter removed a large number of sequence positions from our test. The filter reduced the number of test positions to 118 valine-α, 549 valine-β, 842 leucine-α, 761 leucine-β, 2949 isoleucine-α and 477 isoleucine-β.

In the filtered side chain prediction test there are notable differences between the three methods, reflecting the different energies the methods give for the internal energy of rotamers. The largest differences are seen for isoleucines and valines with helical phi and psi angles. Rosetta's knowledge-based potential, which is based on Dunbrack's backbone dependent rotamer library, only picks the correct rotamer 53% of the time for valine and 41% for isoleucine (table III). The QM calculations with the HF energy predict the correct rotamer 67% of the time for valine and isoleucine. The prediction accuracy with the MM potentials vary significantly; Cedar only places 35% of the valine side chains accurately while CHARMM22 places 55% correctly. These results confirm that for isoleucine and valine with helical torsion angles that the knowledge-based potential does not accurately reflect the internal energy of isoleucine and

valine, but rather the potential is dominated by interactions that isoleucine and valine make with neighboring residues in a helix.

For residues with phi and psi angles in the β-strand region of the Ramachandran plot the QM and knowledge-based potentials do equally well. This suggests that for these residues that the knowledge-based potential is a fairly accurate measure of the internal energy of a side chain. The results with the MM potentials are more varied, and no single potential performs as well as the QM potential or the knowledge-based potential. The complete results table is available in the supplementary information.

Discussion

Accurate estimates for the relative energies of amino side chain conformations are important for protein structure prediction, protein design and drug design. Here, we have shown that various approaches for calculating these energies, molecular mechanics potentials, quantum mechanics calculations and knowledge-based potentials, can give significantly different results, in some cases on the order of 1 kcal / mol per side chain. In general, the QM and knowledge-based energies are more similar with each other than with the results from the molecular mechanics potentials. To evaluate which potentials were most accurate we performed side chain prediction tests. In particular, we examined residues in proteins for which the correct side chain conformation could not be predicted by searching for clashes with neighboring residues. In most scenarios the QM potentials and the knowledge-based potential performed equally well. The exceptions were valines and isoleucines with backbone torsion angles from the helical region of the Ramachandran plot. In these cases the QM potential significantly outperformed the knowledge-based potential because the knowledge-based potential is not an accurate representation of the internal energy of the side chains in this situation, but rather also represents energetics terms derived from being in a helix.

The discrepancy between the QM and knowledge-based energies for β-branched amino acids with helical torsion angles, highlights one of the potential pitfalls of using knowledge-based potentials. The physical basis for preferences observed in the protein database may not always be cleanly assigned to a single energetic effect. For instance, the common hydrogen bond geometries and distances observed in the backbone of an α-helix represent more than the relative energy of different hydrogen bond configurations, they also reflect all the other energetic terms that go in to determining the optimal conformation for a helix. In other words, when knowledge-based potentials are combined with each other or with molecular mechanics potentials, there is a possibility of double counting.

The MM potentials gave fairly erratic results: performing well in some cases but poorly in others. The overall success of the QM energies in side chain prediction tests suggest that they could be used as a benchmark for improving the MM potentials [21,22]. QM simulations on dipeptides have played extensive roles in the parameterization of molecular mechanics potentials from the beginning. Recently there have been attempts by the developers of the CHARMM (version 31) [4,23,24] and ECEPP (version 5) [25] suites to improve the modeling of the protein backbone using QM simulations similar to those conducted here. Both groups sampled either the complete or selected regions of phi/psi space of alanine, glycine, and proline dipeptides. The ECEPP group refit the parameters used to compute backbone torsional energy while the CHARMM group refit its torsional backbone parameters as well as created a 2D grid correction scheme. Both groups have shown improved modeling of the protein backbone [26].

In this study we have restricted our tests to hydrophobic amino acids that do not have the potential to form strong electrostatic interactions between the side chain and the polar atoms in the backbone. In vacuum QM simulations with dipeptides will not be as useful for

determining the rotamer preferences of polar side chains. An alternative approach is to perform QM/MM simulations where the dipeptide is treated by QM and explicit solvent is modeled with a MM forcefield. This type of approach has been used by Hermans and co-workers to map out the conformational preferences of solvated peptides [5]. The peptides intramolecular energies were calculated with the self-consistent charge density functional tight binding method (SCCDFTB) and the solvent was represented by either the SPC or TIP3P models. The distribution of backbone torsion angles obtained with the QM/MM approach more closely matched distributions from high-resolution protein structures than did distrubutions obtained using only MM potentials. Our results suggest that before performing computationally intensive QM/MM simulations with polar side chains, it will be prudent to test our knowledge-based potential in side chain prediction test with polar amino acids. The QM/MM simulations will be most useful for conformations for which the knowledge-based potential is not an accurate reflection of the internal energy of the residue, but rather reflects longer range interactions from the protein. In conclusion, our results indicate that calculating the relative energies of side chain rotamers is still a difficult problem, and combining QM calculations with knowledge-based scores may be the best way to generate an accurate potential.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

UNC Information Technology Services

This work was partially supported by the National Center for Supercomputing Applications under grant MCB040053, and utilized the IBM pSeries 690 systems. This research was supported by an award from the W.M. Keck foundation and the grant GM073960 from the National Institutes of Health.

Bibliography

1. Gordon DB, Marshall SA, Mayo SL. Energy functions for protein design. Curr Opin Struct Biol 1999;9 (4):509–513. [PubMed: 10449371]
2. Moult J. Comparison of database potentials and molecular mechanics force fields. Curr Opin Struct Biol 1997;7(2):194–199. [PubMed: 9094335]
3. McCammon, JA.; Harvey, SC. Dynamics of Proteins and Nucleic Acids. Cambridge, UK: Cambridge University Press; 1987.
4. Mackerell AD. Empirical force fields for biological macromolecules: Overview and issues. Journal of Computational Chemistry 2004;25(13):1584–1604. [PubMed: 15264253]
5. Hu H, Elstner M, Hermans J. Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine “dipeptides” (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution. Proteins 2003;50 (3):451–463. [PubMed: 12557187]
6. Feig M, MacKerell AD, Brooks CL. Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations. Journal of Physical Chemistry B 2003;107(12):2831–2836.
7. Beachy MD, Chasman D, Murphy RB, Halgren TA, Friesner RA. Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. Journal of the American Chemical Society 1997;119(25):5908–5920.
8. Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol 1987;193(4):775–791. [PubMed: 2441069]
9. Dunbrack RL Jr. Rotamer libraries in the 21st century. Curr Opin Struct Biol 2002;12(4):431–440. [PubMed: 12163064]

10. Lovell SC, Word JM, Richardson JS, Richardson DC. The penultimate rotamer library. Proteins-Structure Function and Genetics 2000;40(3):389–408.
11. Pohl FM. Empirical Protein Energy Maps. Nature-New Biology 1971;234(52):277. [PubMed: 5289442]
12. Butterfoss GL, Hermans J. Boltzmann-type distribution of side-chain conformation in proteins. Protein Sci 2003;12(12):2719–2731. [PubMed: 14627733]
13. Butterfoss GL, Kuhlman B. Computer-based design of novel protein structures. Annu Rev Biophys Biomol Struct 2006;35:49–65. [PubMed: 16689627]
14. Dunbrack RL Jr. Comparative modeling of CASP3 targets using PSI-BLAST and SCWRL. Proteins 1999:81–87. [PubMed: 10526356]
15. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. Journal of Computational Chemistry 1983;4(2):187–217.
16. Hermans J, Berendsen HJC, Vangunsteren WF, Postma JPM. A Consistent Empirical Potential for Water-Protein Interactions. Biopolymers 1984;23(8):1513–1518.
17. Ferro DR, Mcqueen JE, Mccown JT, Hermans J. Energy Minimizations of Rubredoxin. Journal of Molecular Biology 1980;136(1):1–18. [PubMed: 7365789]
18. Frisch GWT, MJ.; Schlegel, HB.; Scuseria, GE.; Robb, MA.; Cheeseman, JR.; Montgomery, JA., Jr; Vreven, T.; Kudin, KN.; Burant, JC.; Millam, JM.; Iyengar, SS.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, GA.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, JE.; Hratchian, HP.; Cross, JB.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, RE.; Yazyev, O.; Austin, AJ.; Cammi, R.; Pomelli, C.; Ochterski, JW.; Ayala, PY.; Morokuma, K.; Voth, GA.; Salvador, P.; Dannenberg, JJ.; Zakrzewski, VG.; Dapprich, S.; Daniels, AD.; Strain, MC.; Farkas, O.; Malick, DK.; Rabuck, AD.; Raghavachari, K.; Foresman, JB.; Ortiz, JV.; Cui, Q.; Baboul, AG.; Clifford, S.; Cioslowski, J.; Stefanov, BB.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, RL.; Fox, DJ.; Keith, T.; Al-Laham, MA.; Peng, CY.; Nanayakkara, A.; Challacombe, M.; Gill, PMW.; Johnson, B.; Chen, W.; Wong, MW.; Gonzalez, C.; Pople, JA. Gaussian 03, Revision C.02. Wallingford CT: Gaussian, Inc.; 2004.
19. Wang GL, Dunbrack RL. PISCES: a protein sequence culling server. Bioinformatics 2003;19(12): 1589–1591. [PubMed: 12912846]
20. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. Science 2003;302(5649):1364–1368. [PubMed: 14631033]
21. Petrella RJ, Lazaridis T, Karplus M. Protein sidechain conformer prediction: a test of the energy function. Folding & Design 1998;3(5):353–377. [PubMed: 9806937]
22. Jacobson MP, Kaminski GA, Friesner RA, Rapp CS. Force field validation using protein side chain prediction. Journal of Physical Chemistry B 2002;106(44):11673–11680.
23. Mackerell AD, Feig M, Brooks CL. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. Journal of Computational Chemistry 2004;25(11): 1400–1415. [PubMed: 15185334]
24. MacKerell AD, Feig M, Brooks CL. Improved treatment of the protein backbone in empirical force fields. Journal of the American Chemical Society 2004;126(3):698–699. [PubMed: 14733527]
25. Arnautova YA, Jagielska A, Scheraga HA. A new force field (ECEPP-05) for peptides, proteins, and organic molecules. Journal of Physical Chemistry B 2006;110(10):5025–5044.
26. Buck M, Bouguet-Bonnet S, Pastor RW, MacKerell AD. Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. Biophysical Journal 2006;90(4):L36–L38. [PubMed: 16361340]

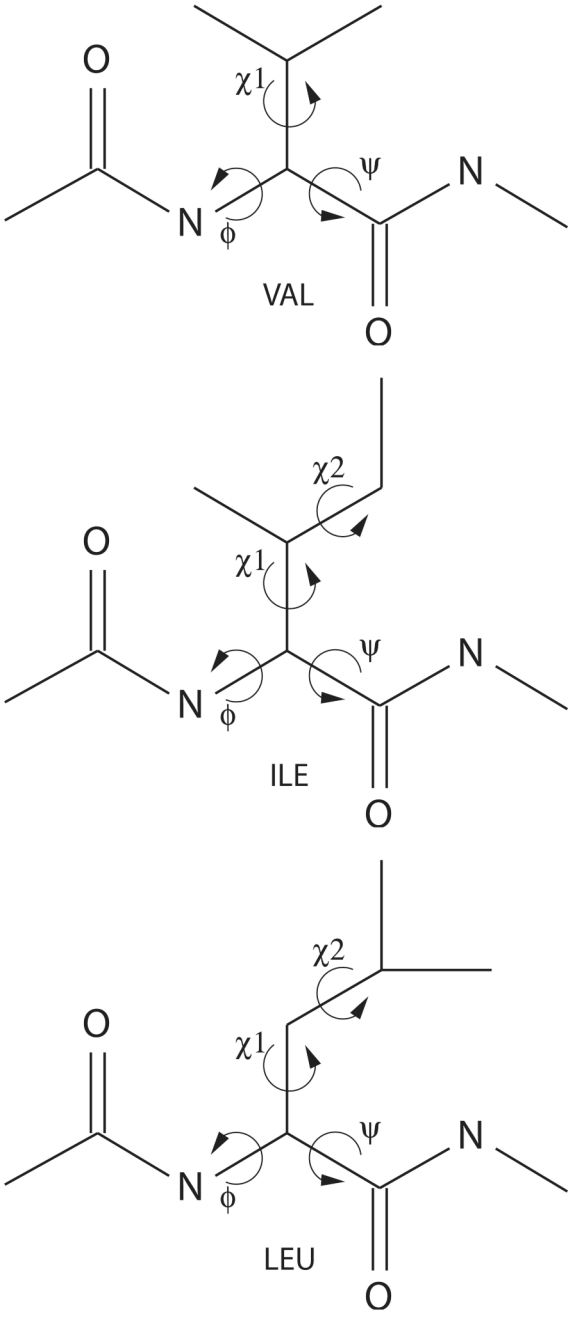

Figure 1.
Diagrams of the (top) valine, (middle) isoleucine, and (bottom) leucine dipeptides showing backbone and side chain torsion angles.

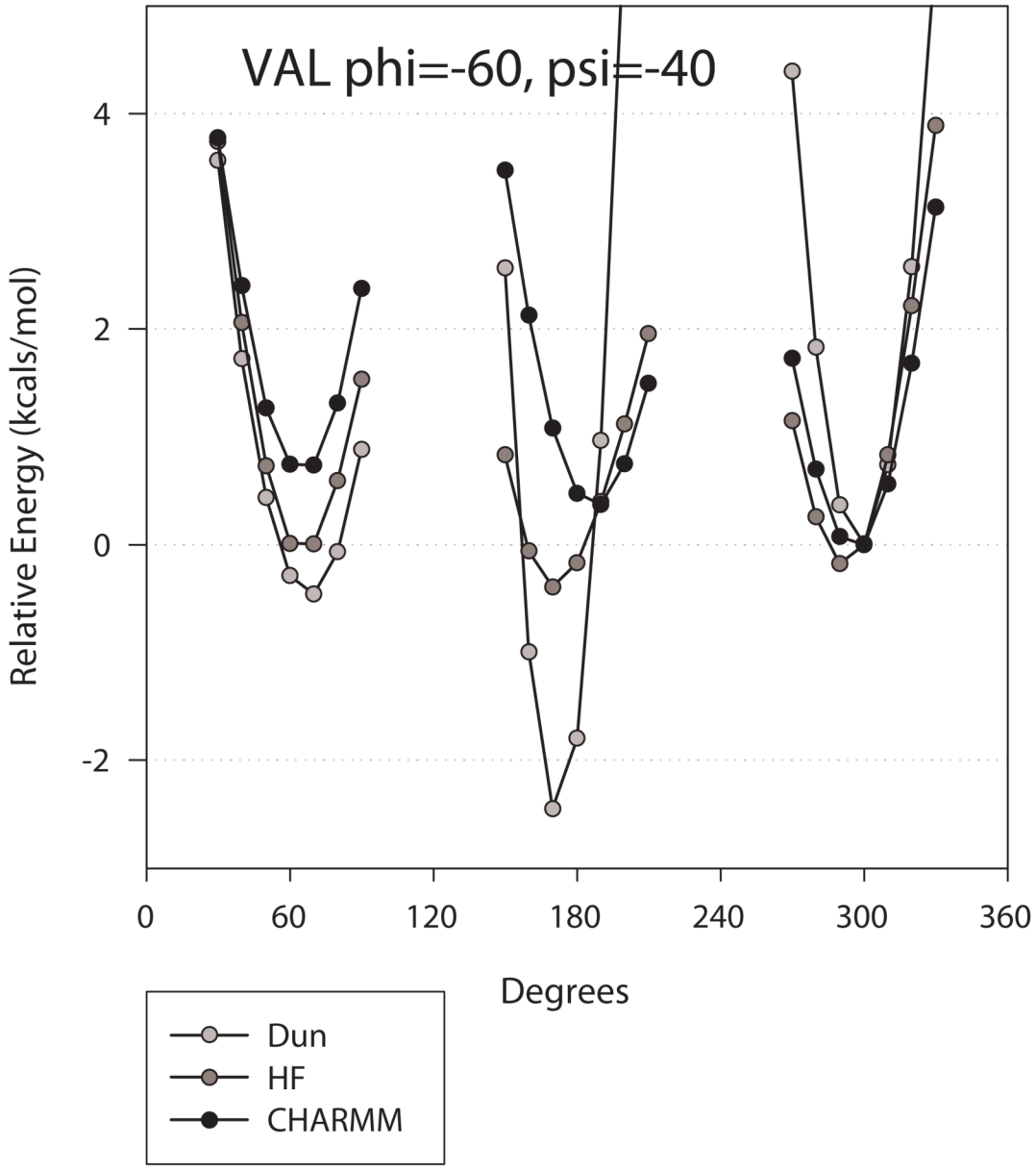

Figure 2.
Comparison between the relative energy differences of (black) CHARMM22 MM potential, (dark grey) Dunbrack rotamer library, or (light grey) energies from the final step of HF minimization for valine in the α-helical region (phi = -60, psi = -40). Probabilities from the Dunbrack library were converted to energies using equations 2 and 3 from the text. Energies for each method were set equal to a value of 0 at a chi of -60 to allow for comparison.

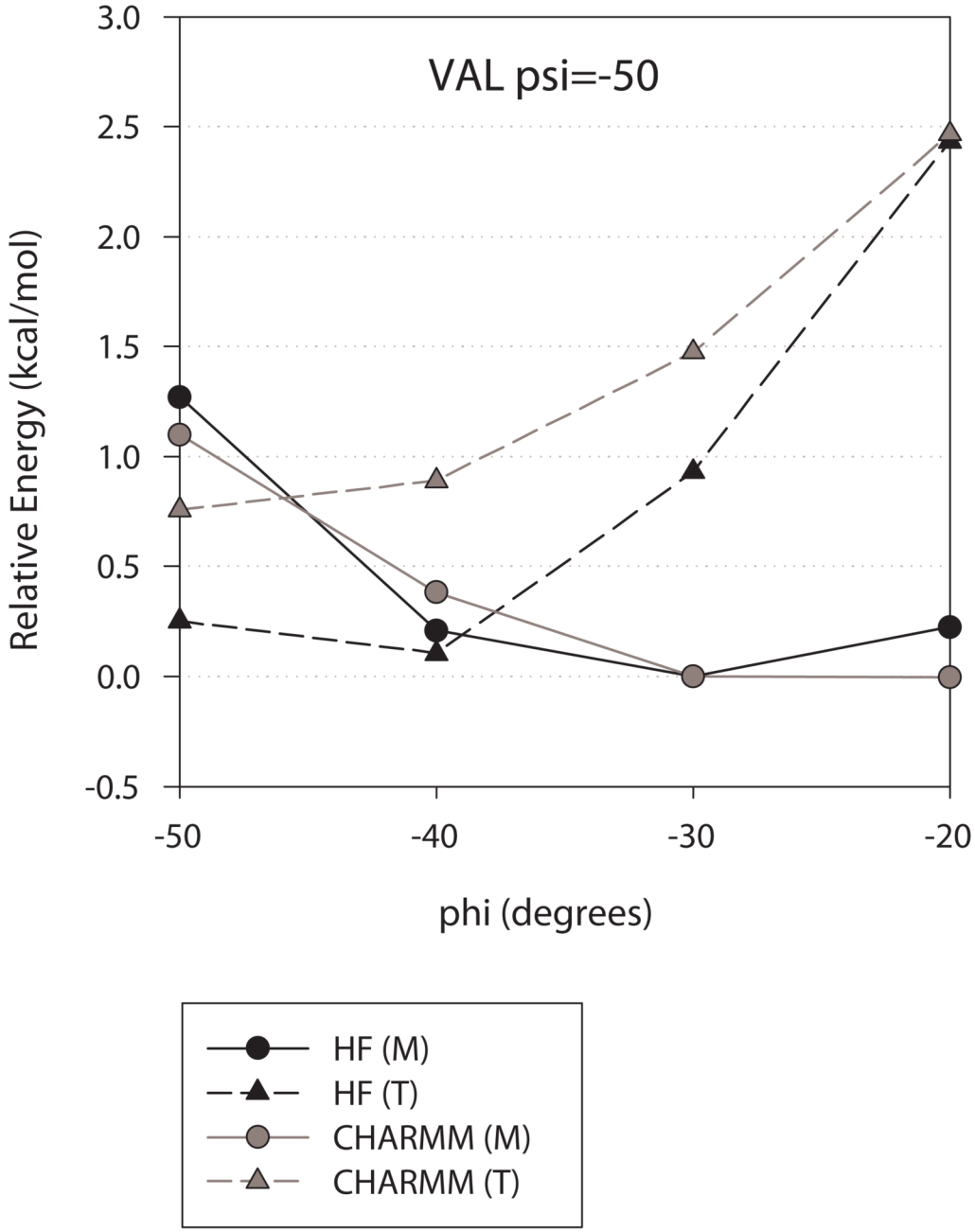

Figure 3.
Relative energy versus psi angle for the M (solid line, circles) and T (dashed line, triangles) rotamers of valine dipeptides using the HF (black) and CHARMM (dark grey) methods. Energies shown are for the phi and psi angle shown and the chi angle that had the minimum energy for that rotamer bin relative to the calculated energy of the M rotamer minimum for each method at a phi of -50, and psi of -30. Psi and Chi angles are as follows HF (M): -50/-70, -40/-70, -30/-60, -20/-60; HF (T) -50/170, -40/170, -30/170, -20/170; CHARMM (M) -50/-70, -40/-60, -30/-60, -20/-60; CHARMM (T) -50/190, -40/190, -30/190, -20/190. See methods for rotamer labeling.

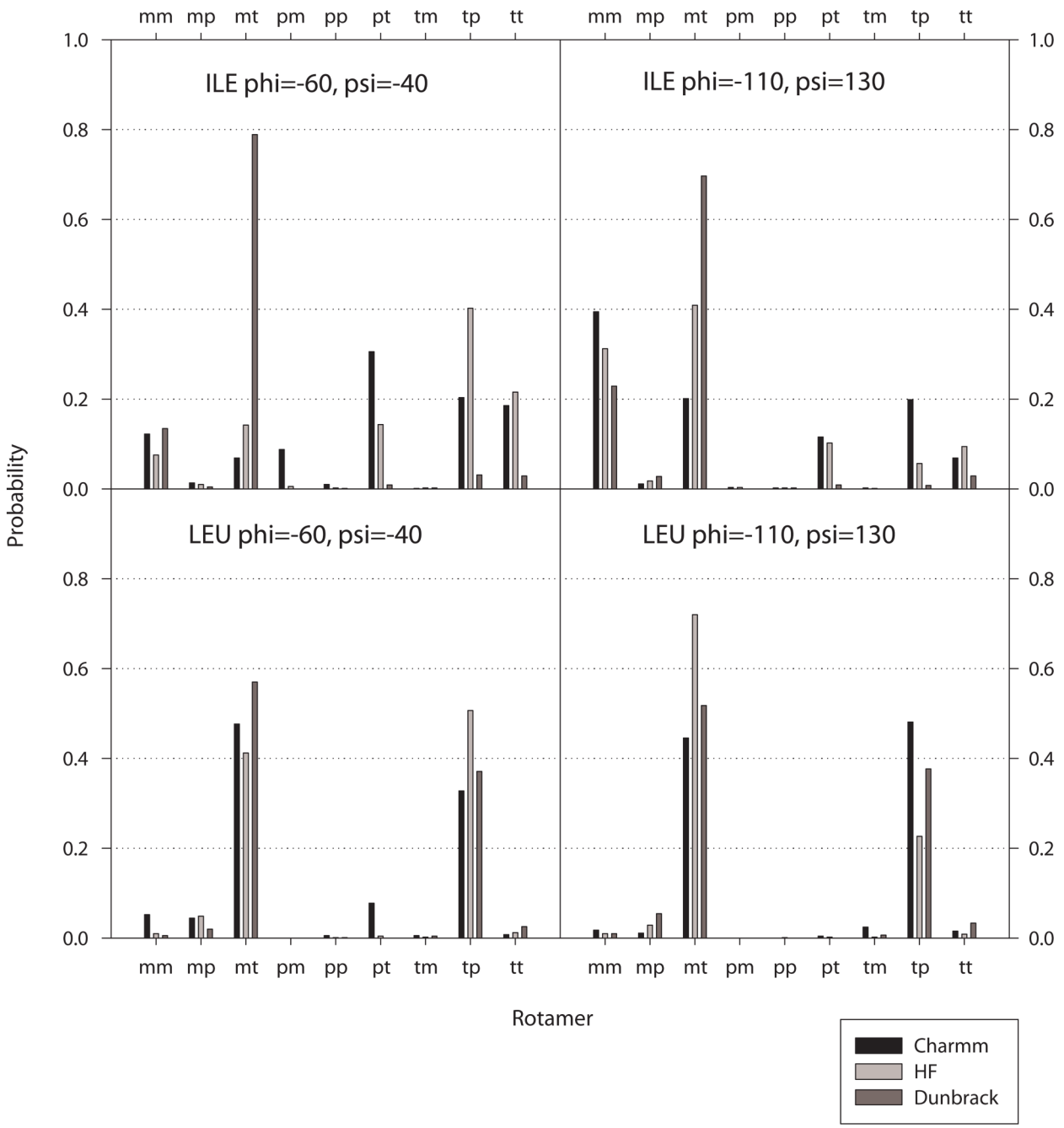

Figure 4.
Probability of choosing a particular rotamer according to (black) CHARMM22 MM potential, (dark grey) Dunbrack rotamer library, or (light grey) HF QM potential for isoleucine and leucine in the canonical α-helical (phi = -60, psi = -40) and β-strand (phi = -110, psi = 130) region. Log probabilities were calculated from energies and normalized to 1 (P = exp(-E/RT)). See methods for rotamer labeling.

Table I

Table of rotamer probabilities[4] for valine dipeptides with α-helical phi and psi.

			psi (degrees), rotamer											
			-50			-40			-30			-20		
			M	T	P	M	T	P	M	T	P	M	T	P
phi (degrees), method	-70	CHARMM[1]	23	14	63	49	19	32	74	14	12	89	8	3
		HF[2]	9	13	78	30	20	50	57	26	17	77	20	4
		DUN[3]	1	1	99	4	4	93	19	17	64	52	30	18
	-60	CHARMM	27	15	58	55	16	29	77	13	9	91	7	2
		HF	11	16	73	32	23	45	57	28	15	79	18	3
		DUN	1	1	98	3	5	92	16	22	62	45	39	16
	-50	CHARMM	30	16	54	59	16	25	82	11	7	94	5	2
		HF	12	20	68	34	27	40	59	29	12	81	17	2
		DUN	1	3	97	3	11	87	13	38	50	28	58	15
	-40	CHARMM	36	15	50	64	15	21	85	9	6	95	4	1
		HF	15	24	62	36	31	33	63	29	8	82	17	1
		DUN	3	7	89	8	29	63	15	61	24	39	29	33

[1] Rotamer probabilities for valine dipeptides as calculated with the CHARMM22 energy function.

[2] Rotamer probabilities for valine dipeptides as calculated with Hartree-Fock, 6-31G(d), Gaussian Inc.

[3] Rotamer probabilities in Dunbrack's backbone dependent rotamer library.

[4] For both QM and MM methods, log probabilities were calculated from the energy of the rotamer that had the phi and psi angle shown and the chi angle that had the minimum energy for that rotamer bin and then normalized to 1 (P = exp(-E/RT)).

Table II

Root mean square deviations between knowledge-base and theoretical rotamer preferences[1].

method	Valine phi = -60 psi = -40	Valine phi = -110 psi = 130	Isoleucine phi = -60 psi = -40	Isoleucine phi = -110 psi = 130	Leucine phi = -60 psi = -40	Leucine phi = -110 psi = 130
CHARMM22 (MM)	47%	22%	27%	19%	5%	5%
Amber-FF99 (MM)	63%	32%	31%	14%	24%	14%
Cedar (MM)	54%	47%	28%	23%	10%	6%
HF (QM)	33%	14%	26%	11%	7%	9%
DFT (QM)	33%	15%	27%	13%	5%	9%
MP2 (QM)	36%	12%	31%	20%	15%	10%

[1]Root mean square (RMS) deviation between rotamer probabilities calculated with QM or MM with probabilities observed in the protein database (Dunbrack's backbone dependent library). A lower value indicates more similarity between the calculated and the PDB probabilities. For both QM and MM methods, log probabilities were calculated from the energy of the rotamer that had the phi and psi angle shown and the chi angle that had the minimum energy for that rotamer bin and then normalized to 1 (P = exp(-E/RT)). Dunbrack probabilities were pulled directly from the library. RMS deviations were calculated between the calculated log probabilities and the probabilities from the Dunbrack library for the same phi, psi, rotamer combinations.

Table III

Side chain prediction tests for “free” positions[1]

	Valine α-helical region	Valine β-strand region	Isoleucine α-helical region	Isoleucine β-strand region	Leucine α-helical region	Leucine β-strand region
Flat	70%	63%	38%	69%	62%	64%
Knowledge-based	53%	92%	41%	86%	86%	91%
CHARMM22 (MM)	55%	89%	40%	78%	87%	78%
Amber-FF99 (MM)	31%	86%	16%	68%	45%	78%
Cedar (MM)	35%	58%	49%	61%	88%	85%
HF (QM)	67%	92%	67%	87%	91%	87%
DFT (QM)	67%	92%	64%	86%	90%	86%
M22 (QM)	68%	92%	54%	81%	77%	86%

[1] Side chain prediction tests were carried out on positions with backbone dihedral angles that correspond to the ranges used in the QM and MM calculations, see methods. A residue position is considered free if the side chain can adopt at least two rotamers that do not have a clash score of greater than 0.5 kcal / mol with neighboring atoms.