JCTC Journal of Chemical Theory and Computation

Assessing the Performance of Popular Quantum Mechanics and Molecular Mechanics Methods and Revealing the Sequence-Dependent Energetic Features Using 100 Tetrapeptide Models

Jinliang Jiang,[†] Yanbo Wu,[†] Zhi-Xiang Wang,[*][,†] and Chun Wu[*][,‡]

College of Chemistry and Chemical Engineering, Graduate University of Chinese Academy of Sciences, Beijing, 100049, China, and Department of Chemistry and Biochemistry, University of California—Santa Barbara, Santa Barbara, California 93106

Abstract: A reasonable description of the conformation energies of each of the amino acids is crucial for modeling protein structures and dynamics. We here used 20 tetrapeptides (ACE-ALA-X-ALA-NME, X = one of 20 amino acids) in 5 conformations (right-handed helix (α_R), left-handed helix (α_L), β-sheet (β), antiparallel β-sheet (β_a), and polyproline II (PPII)) as structural models to investigate the relative conformation energies at the MP2/cc-pVTZ//B3LYP/6-31G** level. The results indicate that the energetic pattern (the order and the energy gap) of the five conformations bears certain resemblances among the amino acids in the same class but is quite different among the amino acids in the different classes (e.g., hydrophobic, aromatic, polar and charged classes). The MP2 energies are then used to statistically evaluate the overall performance of various methods including density functional methods (M05-2X, PBE, and B3LYP), semiempirical methods (AM1, PM3, and PM3MM), empirical polarizable force fields (AMOEBA and AMBER), additive force fields (AMBER, CHARMM, GROMOS, OPLS-AA), and united-atom force fields (AMBERUA and GROMOS). In general, M05-2X obviously outperforms PBE and B3LYP. The semiempirical methods are not able to reach the accuracy as expected. Some of the additive force fields are more accurate than the semiempirical methods. The AMOEBA polarizable force field has accuracy comparable with (or better than) the B3LYP and PBE methods. AMBER99, OPLS-AA, CHARMM27 (excluding α_L), and AMBERUA (excluding α_L) reach reasonable accuracy. However, further improvements, in particular on left-handed helical (α_L) and some residues such as Pro, Asp, and Glu, are necessary.

1. Introduction

A reliable description of conformation energy is crucial for modeling structures and dynamics of biological systems (e.g., proteins, RNA and DNA). To obtain conformation energy accurate enough for biological applications, the weak nonbonding interactions must be properly taken into account.

This requires a high degree of electron correlation energy to be accounted. The quantum mechanics (QM) CCSD(T) approach,[1] coupled-cluster with single and double and perturbative triple excitations, is often considered to be reliable in describing such weak nonbonding interactions. But CCSD(T) is extremely time-consuming at the scale of $O(N^7)$ where N is the number of basis functions, which limits its applications to large molecular systems. The second-order Møller−Plesset perturbation (MP2)[2] method is much less expensive (at a scale of $O(N^5)$) than CCSD(T) and can reach reasonable accuracy in describing the nonbonding interac-

* To whom correspondence should be addressed. E-mail: zxwang@gucas.ac.cn (Z.X.W.); cwu@chem.ucsb.edu (C.W.).

† Chinese Academy of Sciences.

‡ University of California—Santa Barbara.

tions. The methods based on density functional theory (DFT) such as the widely used B3LYP[3,4] and PBE[5] functional have a better scale of $O(N^4)$ than the former two molecular-orbital-based methods, but they are generally not reliable in accounting for nonbonding interactions. Recently, Zhao and Truhlar[6] have developed the M05-2X/M06-2X functionals that account for medium-range correlation energies and thus provide a better description of nonbonding interactions.[7,8] A further compromise between accuracy and computational cost is provided by semiempirical methods (e.g., AM1,[9] PM3, and PM3MM[10,11]). These methods are simplified versions of Hartree–Fock theory by using empirical parameters derived from experimental data, which bring the possibility to study large molecules (up to hundreds of atoms). However, the quality of the DFT and the semiempirical methods in estimating conformation energies is unclear.

Large sizes of biological molecules (e.g., protein, RNA, and DNA) and long time scales of dynamic processes of biological systems (e.g., protein and RNA folding) severely limit the applications of the QM methods. Alternatively, molecular mechanics (MM) modeling provides a tractable approach to describe large biological molecules and make it possible to study the dynamics of biological processes. MM methods describe molecular systems at the atom or united-atom particle level (e.g., aliphatic hydrogen atoms are combined to the connected carbons). Instead of solving the time-consuming Schrödinger equation, MM methods simplify the total potential energy of a molecular system into the sum of several physically meaningful interaction terms (harmonic bond stretching, angle bending, Fourier series for torsion distortion, and Coulomb and Lennard-Jones terms for non-bonding interactions). Anharmonic and cross-terms may be added to improve the accuracy of the force fields.[12] The function of the potential energy and the involved parameters constitute a so-called force field. The force field is the cornerstone of any MM molecular modeling.

Considerable research efforts have been dedicated to developing reliable force fields. The conventional force fields, which have been widely used in studying biological systems, include AMBER,[13] CHARMM,[14] GROMOS,[15,16] and OPLS.[17,18] One of the major defects for the conventional force fields is using fixed partial charges to account for the electrostatic interactions, which neglects the atomic charge changes due to intra- and intermolecular polarization effects. As a consequence, developments of polarizable force fields have been pursued as the next generation of force fields. On the basis of their conventional framework, AMBER, CHARMM, GROMOS, and OPLS have further been developed to implicitly or explicitly include polarization effects.[19-26] In addition to those, the ABEEM developed by Yang's group,[27] AMOEBA developed by the Ponder group,[28] and SIBFA force field developed by Gresh et al.[29] are polarizable force fields for biological systems. In spite of the progress made in developing polarizable force fields, polarizable force fields have not been widely used for studying biological systems due to an elevated computational cost and lack of benchmarking studies to show the benefits.

Force fields were often parametrized to fit the geometric and energetic data of small model molecules from experi-

ments and QM calculations. Although the force fields developed by different groups use very similar energy functions, the parameters may differ significantly due to the different parametrization strategies. For example, the AM-BER force fields obtained atomic partial charges by fitting to the QM electrostatic fields of model molecules,[30,31] while OPLS-AA and GROMOS derived the charges by molecular dynamics (MD) simulations to reproduce the experimental data of model molecules.[15,18] Thus, the empirical nature of force fields and the variations between different force fields make it necessary to benchmark them. Although the aspects for a sufficient benchmark remain under debate (e.g., energetics versus thermodynamic properties), as an important aspect, it has been widely adopted to directly compare MM energies/structures of model molecules or larger systems with those obtained by high quality QM calculations.

During the past decades, a large number of QM calculations on the small molecules that may be regarded as model units for proteins have been reported.[32-53] Böehm et al.[54] and Gould et al.[55] independently show that AMBER force fields overestimate the stability of the C_7 conformation of alanine and glycine dipeptides when compared with their QM results. Beachy et al.[56] optimized 10 conformers of alanine tetrapeptides (ACE-(Ala)$_3$-NME) at the HF/6-31G** level, and the relative conformation energies at the level of local MP2 (LMP2) with the basis set of cc-pVTZ were used to evaluate the popular force fields AMBER (AMBER3, AMBER4.1, and AMBER94), CHARMM (CHARMM19 and CHARMM22), and OPLS (OPLS-AA(2,2), OPLS/A-UA(2,8), OPLS-UA(2,2)), and GROMOS. Their results showed that OPLS-AA(2,2) is the best force field in terms of structure and relative conformation energies. The 10 alanine tetrapeptides were then used by Gresh et al.[29] to evaluate their SIBFA force field which explicitly takes polarization into account via multipole interactions. They showed that the relative energies calculated at the LMP2/6-311G** level could be reproduced by their SIBFA force field with a root-mean-square deviation (RMS) of about 1.3 kcal/mol. Recently, Kaminsky and Jensen[57] calculated dipeptide conformational energies of four amino acids (Gly, Ala, Ser, and Cys) using different QM methods and MM force fields. They found that the B3LYP/6-31G** calculations could not reproduce all the minima found at the MP2/aug-cc-pVDZ(MP2) level, but for the minima that actually exist on the B3LYP potential energy surface, the geometries and relative energies are in good agreement with the MP2 results. For the polarizable force fields, they found that the AMOEBA polarizable force field performs as well as the B3LYP method for ~80% of the conformations but produces ~20% artificial energy minima which are not present on the MP2 energy surface. The fixed charge force fields were only able to reproduce the geometries of approximately half of the conformations, and OPLS_2005 force fields (slightly modified version of the OPLS[58] force fields in the MacroModel program) perform best among their examined force fields. Some authors also have calculated the infinite long polypeptide chain by DFT methods,[59-64] and the comparison with the force fields[65] showed that all force fields overestimate the stability of the helical conformations except for AM-

Popular QM and MM Methods

J. Chem. Theory Comput., Vol. 6, No. 4, 2010 **1201**

BER99/AMBER99SB which satisfactorily reproduce all three helical conformations (π, α, and 3_{10} helix).

These benchmarking studies have contributed greatly to force field developments. Nevertheless, the following limitations motivated the present study: (i) The validations were limited to a few amino acids in the limited secondary structure types. In a force field, three sets of main chain torsion parameters are often used: glycine and proline have their own main chain parameters; the parameters obtained by fitting to the potential energy surfaces of alanine dipeptide or analogues[66,67] are often extended to all remaining amino acids under the assumption of transferability. Yet, the transferability remains elusive, and the errors in these torsion parameters could be one of reasons leading to an imbalance of force fields over the major secondary structure types.[68,69] In addition to the common right-handed helix (α_R) and β secondary structures, a peptide can also adopt a left-handed helix (α_L) and polyproline II conformation (PPII), while the latter conformations are rarely considered in the force field parametrization and assessment. Therefore, evaluations over the complete chemical space of 20 amino acids and over the major secondary structure types could bring us a better understanding of the sequence-dependent conformational energetics and the overall behavior of force fields. (ii) Most of the evaluations were limited to alanine and glycine dipeptides. However, this leads to an additional problem: even if a force field can reproduce the relative energies of dipeptides at different conformations, it does not necessarily imply that the force field is adequate for longer polypeptides because the long-range nonadditive interactions in larger systems that are not present in the dipeptides may play an important role in determining the conformation energy of the longer polypeptides. For example, dipeptides are not able to form intramolecular hydrogen bonds (H-bond) like those in helix secondary structures.

In this study, using tetrapeptides as models, we perform a systemic study to investigate the energetic features of the major conformations of amino acids in the common protein secondary structures at the MP2/cc-pVTZ//B3LYP/6-31G** level and then use the MP2 energies as a "standard" to examine the accuracy of various QM and MM methods (see below for the examined methods). Tetrapeptides are the smallest peptides that can contain H-bonds similar to those in the helix secondary structures. We focus on the five major secondary structures of peptides (i.e the right-handed helix (α_R), left-handed helix (α_L), β-sheet (β), anti-β sheet (β_a), and polyproline II conformation (PPII)). In total, 100 tetrapeptide structures (20 amino acids × 5 conformations) were used in this study. To our knowledge, this is the first time a study of the energetic features of all amino acids in the major secondary structures has been done at levels ranging from MP2/cc-pVTZ to the MM-based molecular mechanics model. These results could provide invaluable information to both method developers and users for future development and method selection.
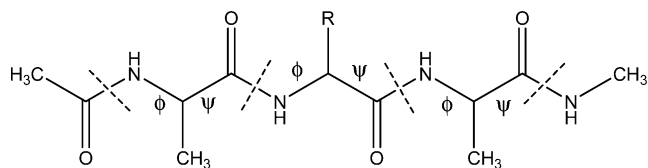

Figure 1. Tetrapeptide model (R = side chain of 20 amino acids).

2. Computational Methods

For a tetrapeptide (Figure 1), ACE-ALA-X-ALA-NME (X is one of the 20 amino acids and ACE and NME are respectively acetyl and methylamide groups which cap the tetrapeptide; in comparison with dipeptide models, the tetrapeptide models using two additional ALAs can reduce the errors due to terminal groups in force field assessment), the five typical conformations have the backbone (ϕ/Ψ) and side chain dihedral angles defined as below. The backbone dihedral angles of the five conformations are the right-handed helix (α_R; $\phi = -57.0°$, $\Psi = -47°$), left-handed helix (α_L; $\phi = 57.0°$, $\Psi = 47°$), β-sheet (β; $\phi = -119.0°$, $\Psi = 113.0°$), anti-β sheet (β_a; $\phi = -140.0°$, $\Psi = 135°$), and polyproline II conformation (PPII; $\phi = -79.0°$, $\Psi = 150.0°$). These (ϕ/Ψ) angles are applied to the three sets of backbone ϕ/Ψ pairs of the tetrapeptides. The rotamer library developed by Dunbrack's group,[70] was used to determine the side chain dihedral angles (except for χ_3 and χ_4 of proline, which are not available from the rotamer library and were obtained from the geometry optimization at B3LYP/6-31G** with χ_1 and χ_2 fixed to the library values). Given the main-chain dihedral angles, the side chain dihedral angles were chosen to be the values in the most populated rotamers. The dihedral angles for the five conformations of each amino acid are provided in the Supporting Information (Table S1 in Supporting Information A (SIA)). All the QM and MM geometric optimizations in the gas phase were carried out with the backbone and side chain dihedral angles fixed to the predefined values. The reason for using these restraints is to prevent the geometric optimizations from producing structures that rarely exist in the peptide/protein structures in aqueous solution (for example, the C_{7eq} conformation is the most stable conformation of alanine dipeptide, but it is rarely seen in protein structures) or from producing divergent structures under different methods, which make the comparisons of conformational energies inconsistent. In other words, the restraints of dihedral angles make it possible to focus our benchmarking on the common protein secondary structures for different methods. The solvation effect is critical in determining protein structures;[71] its influence on the benchmark is under investigation and will be reported in the future.

The structures of the tetrapeptides were optimized at either B3LYP/6-31G** (the optimized structures are drawn in Figure S1 of SIA) or M05-2X/6-31G** levels. The single point energies were then obtained at MP2/cc-pVTZ. Because the two sets of MP2 energies are very close, which is due to the restraints used in the geometry optimizations leading to very similar structures, we only present the data set with the B3LYP/6-31G** optimized structures in the main text,

and the other set of results are provided in Supporting Information B (SIB).

The MP2/cc-pVTZ energies are used as the "standard" values to evaluate the performance of all other methods. The examined QM methods include M05-2X/cc-pVTZ//M05-2X/6-31G**, M05-2X/6-31G**//M05-2X/6-31G**, PBE/cc-pVTZ//PBE/6-31G**, PBE/6-31G**//PBE/6-31G**, B3LYP/cc-pVTZ//B3LYP/6-31G**, B3LYP/6-31G**//B3LYP/6-31G**, AM1//AM1, PM3//PM3, and PM3MM//PM3MM (PM3 + the optional molecular mechanics correction for HCON linkages), where the calculation levels behind "//" indicate the levels used in the structural optimizations. All QM calculations were carried out by using the Gaussian 03 program.[72] Note that the M06-2X functional is not available in Gaussian 03 and should be studied in future work. MM calculations were carried out with the Tinker program[73] using different force fields, including AMOEBA,[28] AMBER94,[13] AMBER96,[74] AMBER99,[30] CHARMM27,[14] OPLS-AA,[17] and OPLS-AA/L.[18] AMBER03,[31] AMBER99SB,[75] AMBEREP,[19] AMBERPOL,[20] and AMBERUA[76] were carried out using the Amber 9 package.[77] GROMOS96 force fields (for versions G43b1, G45a3,[15] and G53a6,[16] COOH was used as C-terminal due to lack of NME) were calculated using the GROMACS 3.3 simulation package.[78] The MM energies were obtained on the basis of the reoptimized structures at the corresponding level. A dielectric constant of 1.0 and an infinite cutoff for Lennard-Jones interactions was used in MM calculations. All conformation energies relative to the α_R conformation are provided in Tables S2 of SIA and the Table S1 of SIB).

To statistically evaluate the performance of the examined methods, two types of root-mean-square deviations (RMS) of conformation energies were calculated either for each amino acid, averaged over the five conformations (RMS), or for each conformation type, averaged over 20 amino acids (RMS-C). The first type of RMS is calculated by using eqs 1−3:

$$ \text{RMS} = \sqrt{\frac{\sum_{i=1}^{n} (\text{error})^2}{n}} \tag{1} $$

$$ \text{error} = E_{ai} - E_{bi} + E_c \tag{2} $$

$$ E_c = \frac{\sum_{i=1}^{n} (E_{bi} - E_{ai})}{n} \tag{3} $$

where n is the total number of the conformations (i.e., $n = 5$), E_{bi} and E_{ai} are respectively the relative energies (i.e., setting the energy of α_R to be zero) of the reference method (i.e., MP2) and a given method, error is a signed error using the MP2 energy as the "true" value, and E_c is a constant to minimize the rms for each amino acid type, which fixes the issue that, if the relative energies are defined relative to a given conformation, the rms values will depend on the reference conformation. As indicated by eqs 2 and 3, the use of E_c is actually equivalent to using the mean value of all conformations as the reference. The second type of rms

is the signed rms-C of each conformation type averaged over 20 amino acids for a given method and calculated by using eqs 4 and 5:

$$ \text{RMS-C} = \sqrt{\frac{\sum_{j=1}^{m} (E_{aj} - E_{bj} + E_c)^2}{m}} \tag{4} $$

$$ \text{SIGN} = \text{sign}\left(\frac{\sum_{j=1}^{m} E_{aj} - E_{bj} + E_c}{m}\right) \tag{5} $$

where m is the total number of the amino acid type (i.e., $m = 20$), E_{bj} and E_{aj} are respectively the relative energies (i.e., setting the energy of α_R to be zero) of the MP2 and a given method for a particular conformation, and E_c is the energy offset obtained from eq 2 in minimizing the RMS for each amino acid type. The reason for not directly using the E_c obtained from eq 4 in minimizing RMS-C for each conformation type (using 20 amino acids) is that such an energy reference E_c should not depend on conformation type. SIGN is determined from eq 5, which determine the sign of the averaged signed error over the 20 amino acids. On the basis of the definition of RMS/RMS-C, one can see that the RMS/RMS-C can provide statistical information of the performance of a given method on an individual amino acid over all five conformations/on individual conformations over all 20 amino acids, respectively.

In addition, RMS and RMS-C are also calculated on the basis of four conformations (i.e., excluding the left-handed helix conformation), because the α_L conformation is only adopted by short peptides and is rarely presented in protein structure modeling. To distinguish them from those calculated over all five conformations, we refer to them as RMS-Nα_L and RMS-C-Nα_L, where Nα_L is the abbreviation for "not including α_L".

To evaluate the overall performance of the examined methods, the means (μ) of RMS/RMS-Nα_L were calculated as the averages over 20 amino acids by taking five/four conformations into account. The means (μ) of unsigned RMS-C/RMS-C-Nα_L were calculated as the averages over five/four conformations by taking 20 amino acids into account. The standard deviations (σ) of RMS/RMS-Nα_L and unsigned RMS-C/RMS-C-Nα_L were calculated correspondingly, which provide information on whether a given method has a balanced performance on 20 amino acids or on five/four conformations.

3. Results and Discussion

For brevity, we use the amino acid name to refer the whole tetrapeptide hereafter. The energies at the MP2/cc-pVTZ//B3LYP/6-31G** level relative to the α-helix conformation are listed in Table 1 and are plotted in Figure 2.

From left to right in Figure 2, we order the results by following the common classifications of 20 amino acids: hydrophobic (Pro-Met), aromatic (Phe-Trp), polar (Cys-Gln), and charged (Asp-Arg) classes. As expected, the pattern in terms of energy order and gaps of the five conformations

Popular QM and MM Methods

J. Chem. Theory Comput., Vol. 6, No. 4, 2010 **1203**

Table 1. Relative Energies (kcal/mol; with reference to the α_R conformation) of the 100 Tetrapeptide Structures at the MP2/cc-pVTZ//B3LYP/6-31G** Level

	α_R	α_L	β	β_a	PPII
Pro	0.00	2.24	4.26	5.29	1.76
Gly	0.00	1.55	3.67	1.74	2.85
Ala	0.00	2.23	3.44	1.01	2.91
Val	0.00	3.61	2.15	0.18	3.17
Leu	0.00	3.24	3.37	0.79	3.01
Ile	0.00	4.15	1.97	0.13	2.27
Met	0.00	−0.12	3.05	0.87	1.53
Phe	0.00	−0.62	0.55	−0.91	2.69
Tyr	0.00	−0.43	0.37	−0.97	2.51
His	0.00	0.05	1.28	0.13	2.32
Trp	0.00	−0.19	2.54	−0.27	2.40
Cys	0.00	0.95	1.78	1.15	4.06
Ser	0.00	1.33	3.90	1.13	3.62
Thr	0.00	1.99	5.44	4.22	4.00
Asn	0.00	−0.84	1.36	1.53	3.52
Gln	0.00	2.06	2.63	0.43	−1.74
Asp	0.00	16.70	−0.44	7.01	12.97
Glu	0.00	5.34	−0.63	2.91	7.20
Lys	0.00	2.92	3.51	0.42	0.78
Arg	0.00	−4.06	−1.33	−18.48	−4.90

for each amino acid shows certain similarity within the class (e.g., Gly vs Ala, Phe vs Tyr) and more obvious differences between the classes (e.g., aromatic and charged classes vs hydrophobic and polar classes). The main features are summarized as below: (1) α_R conformations are the lowest in the hydrophobic (except for Met) and polar classes (except for Asn and Gln), but this is not the case in the charged (except for Lys) and aromatic classes. This data may indicate the following: when the interaction between the side chain and backbone is weak, as in the hydrophobic and polar classes, the backbone interactions (e.g., H-bond) dictate the conformation energy, but when the side chain−backbone interaction is strong, as in the charged and aromatic classes, it may overtake the backbone interactions and change the energetic pattern. (2) In the aromatic class, the energy of α_L is close to or less than that of α_R (−0.62, −0.43, −0.19, and 0.05 kcal/mol energy difference for Phe, Tyr, Trp, and His, respectively). This might be caused by the ring−backbone interactions. (3) In the charged class, the energy gaps between different conformations become large except for Lys. For example, the energy gaps between the lowest energy

Figure 2. Relative energies of five conformations of 20 amino acids using MP2/cc-pVTZ//B3LYP/6-31G**.

conformation and the highest energy conformation is 7.83, 17.14, and 18.48 kcal/mol for Glu, Asp, and Arg, respectively. In addition, Arg has an extremely favorable β_a conformation, at least 13.58 kcal/mol lower than other conformations; this is caused by H-bond interaction between the N−H of the side chain and the C=O of the backbone for β_a (Figure S1 in SIA). (4) Although α_L appears to be "mirror" image of α_R in a reduced ribbon representation, their energies are not close to each other. In fact, α_L has a higher energy than α_R for 13 amino acids (except for the whole aromatic class, Met, Asn, and Arg) by at least 1.0 kcal/mol; this can be attributed to the stronger steric effect between C=O groups and side chains in α_L than that between N−H groups and side chains in α_R. However, in the case of Arg, α_L has a lower energy than α_R by −4.06 kcal/mol, which might stem from the different side chain torsions in the two conformations. In short, no two amino acids have the exact same conformation energy profile, highlighting the sequence dependent features. From the perspective of secondary structure propensity, the conformation energy profile determines the intrinsic preference toward certain secondary structure types (extended conformation versus helical conformation) for each amino acid. Thus, it is critical for lower level methods to reproduce these energetic signatures of each amino acid as accurately as possible. The errors may lead to the wrong secondary structure propensity for each amino acid.

In the following, we use the MP2 energies as the "standard" to assess the other methods. It should be pointed out that the MP2 energies may have a deviation of about 0.5 kcal/mol with respect to the "true" values calculated at the more sophisticated QM level. For brevity, we mainly discuss the overall performance of the examined methods and place the details in the Supporting Information. Tables 2 and 3 list only the signed RMS-C values and the RMS-Nα_L values, respectively. The means (μ) and standard deviations (σ) of the unsigned RMS-C's and the RMS-Nα_L's are plotted in Figures 3 and 4 for visualization, respectively. The RMS-C-Nα_L values are given in Table S4c of SIA and included in Figure 3. The RMS values are given in Table S3d of SIA but not included in Figure 4 for brevity.

The RMS-C values in Table 2 and its mean in Figure 3 demonstrate the M05-2X functional obviously outperforms the PBE and B3LYP functional in predicting the relative conformation energies. The mean RMS-C values (μ) of M05-2X/cc-pVTZ (0.79 kcal/mol) and M05-2X/6-31G** (0.84 kcal/mol) are substantially less than those of PBE/cc-pVTZ (2.60 kcal/mol), PBE/6-31G** (1.45 kcal/mol), B3LYP/cc-pVTZ (3.24 kcal/mol), and B3LYP/6-31G** (1.90 kcal/mol). This can be attributed to the better description of nonbonding interactions by the M05-2X functional. As shown in Table 2, at both levels of PBE (PBE/6-31G** and PBE/cc-pVTZ), B3LYP (B3LYP/6-31G** and B3LYP/cc-pVTZ), and M05-2X/cc-pVTZ, the α_R and α_L conformations which contain intramolecular H-bonds have positive RMS-C values, while the RMS-C's of extended conformations (β, β_a, and PPII) which do not have intramolecular H-bonds are negative. In other words, these DFT methods overestimate the energies of the compact helical conformations but underestimate the
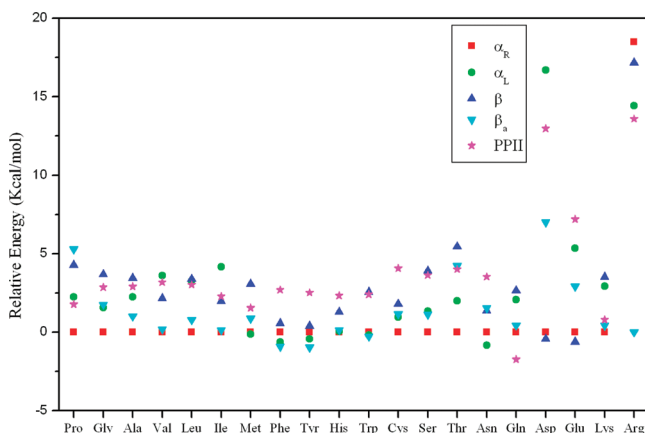
Table 2. Signed RMS-C (kcal/mol) of Each Conformation over 20 Amino Acids for All Considered Methods

	α_R	α_L	β	β_a	PPII	mean (μ)	[1]SD (σ)
M05-2X/cc-pVTZ	1.22	0.62	−0.52	−0.79	−0.78	0.79	0.24
M05-2X-D[a]/cc-pVTZ	−1.09	0.86	0.92	0.89	0.54	0.86	0.18
M05-2X/6-31G**	0.48	−1.79	0.93	0.31	0.70	0.84	0.52
M05-2X-D[a]/6-31G**	−1.81	−1.97	1.54	1.24	1.59	1.63	0.25
PBE/cc-pVTZ	3.22	3.18	−3.02	−2.66	−0.91	2.60	0.87
PBE-D[a]/cc-pVTZ	1.21	3.16	−2.55	−2.05	0.51	1.90	0.94
PBE/6-31G**	2.19	0.95	−1.85	−1.69	0.58	1.45	0.60
PBE-D[a]/6-31G**	0.48	1.08	−1.39	−1.19	1.31	1.09	0.32
B3LYP/cc-pVTZ	3.89	4.12	−3.30	−3.06	−1.83	3.24	0.80
B3LYP-D[a]/cc-pVTZ	1.85	4.09	−2.83	−2.42	−1.04	2.45	1.02
B3LYP/6-31G**	2.75	1.87	−2.11	−2.08	−0.68	1.90	0.68
B3LYP-D[a]/6-31G**	0.81	1.91	−1.63	−1.50	0.56	1.28	0.51
AM1	2.11	4.68	−3.20	−1.74	−2.83	2.91	1.02
AM1-D[a]	−1.17	4.70	−2.81	1.74	−1.98	2.48	1.23
PM3	5.54	8.01	−3.93	−3.01	−7.27	5.55	1.90
PM3-D[a]	3.48	7.99	−3.54	−2.44	−6.39	4.77	2.08
PM3MM	5.73	8.74	−4.98	−3.45	−6.71	5.92	1.77
PM3MM-D[a]	3.67	8.69	−4.56	−2.76	−5.83	5.10	2.06
AMOEBA	1.21	−1.61	1.67	1.78	−1.54	1.56	0.19
AMBEREP	−4.11	10.24	−1.89	−2.59	−2.68	4.30	3.06
AMBERPOL	−3.17	5.91	1.29	−1.45	−2.54	2.87	1.67
AMBER94	−4.60	4.80	1.45	1.81	−2.22	2.98	1.43
AMBER96	1.04	9.55	−3.15	−4.20	−3.09	4.21	2.86
AMBER99	−2.86	3.78	1.90	−2.28	−1.02	2.37	0.93
AMBER99SB	2.72	2.98	−1.25	−2.16	−3.53	2.53	0.78
AMBER03	−1.89	9.11	−1.83	−3.28	−4.79	4.18	2.69
CHARMM27	−2.99	14.55	−4.10	−2.87	−6.02	6.11	4.37
OPLS-AA	2.60	4.58	−3.20	−2.57	−2.96	3.18	0.74
OPLS-AA/L	2.54	5.70	−2.31	−3.02	−3.97	3.51	1.23
AMBERUA	−3.50	13.86	−2.54	−3.32	−5.54	5.75	4.17
GROMOS(G43b1)	4.33	10.65	−5.63	−3.79	−4.53	5.79	2.50
GROMOS(G45a3)	3.59	9.60	−5.68	−3.66	−4.05	5.32	2.27
GROMOS(G53a6)	5.14	10.36	−6.51	−4.06	−4.87	6.19	2.23

[a] AMBER 99 dispersion energies are applied (see text for details).

energies of the extended conformations with respect to the corresponding MP2 energies. This can be attributed to the fact that the DFT methods (in particular the B3LYP functional) are not able to account for the nonbonding interactions properly (e.g., underestimation of the dispersion and H-bonding interactions in the compact conformations). Similarly, Table 2 can be used to examine the performance of other methods on the individual conformations.

Because the MP2 energies were computed using the cc-pVTZ basis set, one may assume that the cc-pVTZ basis set could give better agreement than the 6-31G** basis set. However, the mean RMS-C values indicate that the cc-pVTZ basis set only marginally improves the agreement of the M05-2X functional from 0.84 kcal/mol (6-31G** basis set) to 0.79 kcal/mol, but it even worsens the agreement of the B3LYP (and PBE) functional from 1.90 (and 1.45) kcal/mol (6-31G** basis sets) to 3.24 (and 2.60) kcal/mol. This can be attributed to the larger basis set superposition error (BSSE) of the 6-31G** basis set than that of the cc-pVTZ basis set. With respect to the cc-pVTZ basis set, the 6-31G** basis set leads to larger BSSE values for more compact conformations than for the extended conformations, which compensates more for the dispersion that is intrinsically underestimated by DFT methods in the compact conformations than in the extended conformations. Because of the defect of the PBE and B3LYP functionals in accounting for the nonbonding interactions, we simply added the MM dispersion energies obtained from AMBER99 calculations to the PBE and B3LYP energies (denoted by adding suffix

"D" in the tables and figures); the agreements of the PBE and B3LYP functionals are improved by about 0.4−0.7 kcal/mol; the mean RMS-C's of the cc-pVTZ basis set are reduced from 2.60 and 3.24 kcal/mol to 1.90 and 2.45 kcal/mol for PBE and B3LYP, respectively; the mean RMS-C's of the 6-31G** basis set are decreased from 1.45 and 1.90 to 1.09 and 1.28 kcal/mol for PBE and B3LYP, respectively. This implies that the PBE and B3LYP functional can be moderately improved by adding the Lennard-Jones potential, indicating that the accuracy of DFT methods can be further improved by treating dispersion interactions in a more systematic way, as exemplified by the M05-2X and M06-2X density functionals. Due to the double counting of the medium-range dispersion effect, such corrections worsen the M05-2X performance by about 0.1−0.8 kcal/mol. The standard RMS-C deviations (σ) pronounce that the M05-2X/cc-pVTZ method ($\sigma = 0.24$ kcal/mol) has more consistent descriptions of the five conformations than do PBE/cc-pVTZ ($\sigma = 0.87$ kcal/mol) and B3LYP/cc-pVTZ ($\sigma = 0.80$ kcal/mol), which is in agreement with the above discussion.
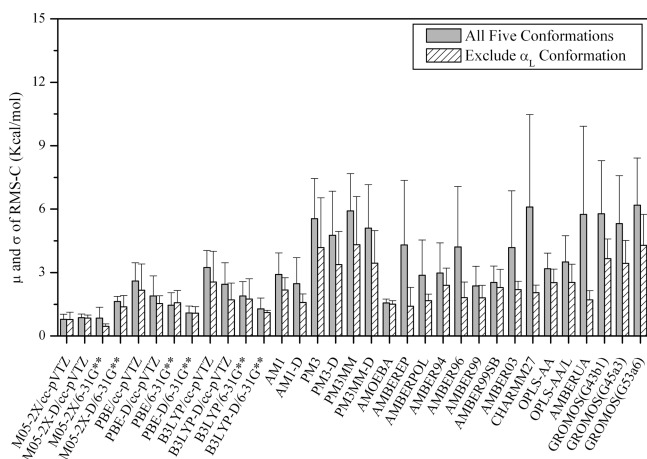
The examined semiempirical methods are less accurate than both DFT methods. Their mean RMS-C's are larger than those of DFT methods (see Table 2). AM1 ($\mu = 2.91$ kcal/mol) outperforms PM3 ($\mu = 5.55$ kcal/mol). The signed RMS-C values of AM1 and PM3 indicate that the semiempirical methods may share the same reasons for their poor performance as the B3LYP methods but with larger errors. The poor performance of PM3 cannot be rescued by adding a MM correction of the pyramidalization of the amide

Popular QM and MM Methods

J. Chem. Theory Comput., Vol. 6, No. 4, 2010 **1205**

Table 3. RMS-Nα_L (kcal/mol) over the Four Conformations (α_R, β, β_a, and PPII) of Each Amino Acid for Each Method

	Pro	Gly	Ala	Val	Leu	Ile	Met	Phe	Tyr	His	Trp	Cys	Ser	Thr	Asn	Gln	Asp	Glu	Lys	Arg	Mean	SD
M05-2X/cc-pVTZ	0.83	0.98	0.84	0.83	0.77	0.77	0.68	0.99	1.05	0.88	1.15	1.05	0.77	0.67	0.74	0.77	0.30	0.85	0.84	0.92	0.83	0.17
M05-2X-D[a]/cc-pVTZ	2.34	0.58	0.62	0.50	0.48	0.49	0.52	0.59	0.87	0.63	0.82	1.27	0.64	0.55	0.35	0.68	0.91	1.08	0.42	0.54	0.74	0.43
M05-2X/6-31G**	0.34	0.50	0.14	0.38	0.44	0.33	0.25	0.30	0.58	0.22	0.26	1.11	0.34	0.46	0.27	0.48	0.79	0.40	0.41	0.40	0.42	0.21
M05-2X-D[a]/6-31G**	2.72	1.46	1.43	1.31	1.42	1.26	1.37	1.14	1.35	1.21	1.58	1.99	1.43	1.49	1.14	1.50	1.43	1.38	1.11	0.88	1.43	0.37
PBE/cc-pVTZ	2.51	2.80	2.74	2.56	2.63	2.50	2.53	2.40	2.31	2.56	2.84	2.44	2.60	2.68	2.37	2.59	1.35	2.32	2.70	2.08	2.48	0.31
PBE-D[a]/cc-pVTZ	1.12	1.62	1.54	1.59	1.61	1.45	1.52	1.86	1.82	2.12	1.68	1.31	1.36	1.59	1.50	1.34	0.28	1.03	2.15	1.94	1.52	0.40
PBE/6-31G**	1.86	1.94	1.84	1.60	1.66	1.56	1.65	1.51	1.47	1.62	1.84	1.59	1.70	1.57	1.55	1.89	1.18	1.91	1.70	1.64	1.66	0.18
PBE-D[a]/6-31G**	1.33	1.18	0.99	0.90	0.99	0.61	0.98	1.30	1.38	1.50	1.12	0.88	0.72	0.68	0.96	1.15	0.75	0.98	1.43	1.76	1.08	0.29
B3LYP/cc-pVTZ	2.92	3.16	3.20	3.00	3.11	2.94	2.94	2.88	2.81	3.02	3.41	2.92	3.15	3.18	2.84	2.96	1.54	2.61	3.19	2.46	2.91	0.38
B3LYP-D[a]/cc-pVTZ	1.47	1.82	1.87	1.90	1.97	1.86	1.85	2.13	2.08	2.37	2.08	1.64	1.85	2.02	1.86	1.61	0.53	1.32	2.48	2.09	1.84	0.40
B3LYP/6-31G**	2.06	2.15	2.25	2.01	2.05	1.95	2.02	1.89	1.81	1.98	2.28	1.98	2.21	2.02	1.96	2.16	1.20	1.86	2.09	1.72	1.98	0.23
B3LYP-D[a]/6-31G**	1.14	1.00	1.06	1.01	1.06	0.87	1.00	1.32	1.27	1.52	1.18	0.84	0.98	0.90	1.09	1.08	0.56	0.66	1.52	1.53	1.08	0.25
AM1	2.20	2.39	2.14	1.91	2.31	1.95	1.50	2.62	2.56	2.52	2.80	1.62	2.17	1.62	1.37	1.67	1.61	2.14	3.90	2.18	0.57
AM1-D[a]	1.70	1.22	1.07	1.14	1.24	1.49	1.02	1.70	1.63	1.81	1.16	0.57	1.95	0.94	0.44	1.16	1.39	1.58	1.76	4.18	1.46	0.74
PM3	4.85	4.63	4.84	4.85	5.48	4.79	3.90	5.24	5.22	5.19	5.45	4.42	3.82	5.38	4.17	3.56	4.88	5.13	4.83	4.69	4.77	0.54
PM3-D[a]	2.85	3.37	3.58	3.64	4.35	3.99	2.85	4.10	4.06	4.20	3.96	3.26	2.91	4.27	3.23	2.37	4.24	3.93	4.04	4.35	3.68	0.58
PM3MM	4.92	4.92	4.84	4.76	5.40	4.70	3.96	5.22	5.15	5.11	5.40	4.55	4.64	5.78	4.37	3.85	4.82	4.64	5.28	4.96	4.86	0.46
PM3MM-D[a]	2.77	3.55	3.50	3.54	4.23	3.83	2.84	4.14	4.04	4.16	3.84	3.25	3.68	4.59	3.33	2.54	4.22	3.43	4.48	4.58	3.73	0.58
AMOEBA	3.92	0.70	0.70	1.62	0.83	1.07	0.53	0.46	0.27	1.56	1.01	0.90	0.51	1.94	1.58	0.23	2.50	2.10	1.38	0.91	1.24	0.87
AMBEREP	2.91	1.35	0.72	0.42	0.93	0.54	1.30	0.73	0.78	0.50	1.42	1.96	0.95	1.07	1.66	0.68	1.66	2.94	0.31	1.73	1.23	0.73
AMBERPOL	3.92	0.58	1.29	0.99	1.23	1.40	1.38	1.12	1.00	0.70	1.91	1.89	1.00	1.56	1.21	1.40	1.44	2.96	0.76	2.40	1.51	0.78
AMBER94	4.03	1.37	2.16	1.75	2.11	2.09	2.18	2.05	1.87	2.19	2.40	2.94	2.41	1.87	1.35	2.42	2.87	3.99	2.27	3.98	2.42	0.77
AMBER96	1.18	2.49	1.79	2.15	1.75	2.37	1.71	1.81	1.98	1.72	1.67	1.89	1.88	2.02	2.59	1.60	1.77	2.22	1.99	2.09	1.93	0.32
AMBER99	2.47	1.38	1.47	1.29	1.24	1.30	1.81	1.44	1.30	1.25	1.90	2.47	1.72	1.62	0.60	1.74	2.87	3.92	1.23	2.22	1.76	0.72
AMBER99SB	2.29	2.32	2.35	2.59	2.02	3.16	1.94	2.24	2.36	2.25	2.28	2.70	2.65	2.42	2.65	1.91	2.00	2.57	2.72	3.02	2.42	0.33
AMBER03	2.43	1.48	1.57	1.87	1.39	2.81	1.16	1.23	1.30	1.76	1.12	2.27	1.73	2.72	1.18	2.34	2.01	3.03	3.72	4.13	2.06	0.84
CHARMM27	2.38	1.86	1.35	1.49	1.71	1.86	0.91	1.97	1.99	2.30	1.53	1.82	0.64	1.26	1.69	0.99	4.22	2.64	2.03	3.53	1.91	0.82
OPLS-AA	4.13	2.00	2.00	2.13	2.33	2.65	2.29	2.71	2.75	2.42	1.94	2.21	2.24	2.57	3.59	2.11	1.86	2.16	3.23	3.36	2.53	0.59
OPLS-AA/L	4.78	4.02	1.97	2.04	2.36	1.80	1.79	2.69	2.71	2.42	1.75	2.59	3.11	2.49	2.74	2.13	1.95	1.70	2.77	3.37	2.56	0.77
AMBERUA	2.56	1.53	1.22	1.94	1.01	1.54	0.92	1.28	1.91	1.37	1.59	2.16	0.96	0.83	1.30	1.83	1.44	2.62	1.36	3.47	1.64	0.64
GROMOS(G43b1)	2.33	3.31	3.18	3.14	3.34	3.07	2.92	3.14	3.22	2.95	2.06	3.00	2.65	3.15	3.24	2.71	7.60	5.63	3.75	6.22	3.53	1.33
GROMOS(G45a3)	2.47	3.48	3.38	3.25	3.54	3.11	3.09	3.40	3.43	3.10	2.25	3.15	2.83	3.30	3.40	2.83	5.24	5.66	3.62	5.32	3.49	0.88
GROMOS(G53a6)	7.15	4.68	4.56	4.34	4.80	4.22	3.72	4.60	4.62	4.25	3.52	4.17	3.87	4.24	4.24	3.85	4.33	4.71	4.47	5.13	4.47	0.72

[a] AMBER99 dispersion correction.

nitrogen in PM3MM ($\mu = 5.92$ kcal/mol). The additions of MM-dispersion interactions slightly improve their performance by 0.43 kcal/mol for AM1 and 0.78 kcal/mol for PM3 and PM3MM. Note that the mean RMS-C of AM1, 2.91 kcal/mol, is comparable with that (3.24 kcal/mol) of the least accurate B3LYP/cc-pVTZ method. The standard RMS-C deviations (σ(AM1) = 1.02 kcal/mol and σ(PM3) = 1.90 kcal/mol) also indicate that the examined semiempirical methods have less consistent descriptions of the five conformations than DFT methods which have σ values ranging from 0.2−0.8 kcal/mol.

Remarkably, the polarizable force field, AMOEBA ($\mu = 1.56$ kcal/mol), performs better than the B3LYP/cc-pVTZ ($\mu = 3.24$ kcal/mol), B3LYP/6-31G** ($\mu = 1.90$ kcal/mol), and semiempirical AM1 ($\mu = 2.91$ kcal/mol) and PM3 ($\mu = 5.55$ kcal/mol) but less accurate than M05-2X with both cc-pVTZ or 6-31G** ($\mu \approx 0.8$ kcal/mol). The standard RMS-C deviation ($\sigma = 0.19$ kcal/mol) of AMOEBA reaches the value of M05-2X, indicating that the force field has a consistent description over all five conformations. These are encouraging signs for developing such a polarizable force field to simulate biological molecules. However, the performance of the two versions

Figure 3. Mean (μ) and standard deviation (σ) of RMS-C of each conformation calculated over 20 amino acids for each method.
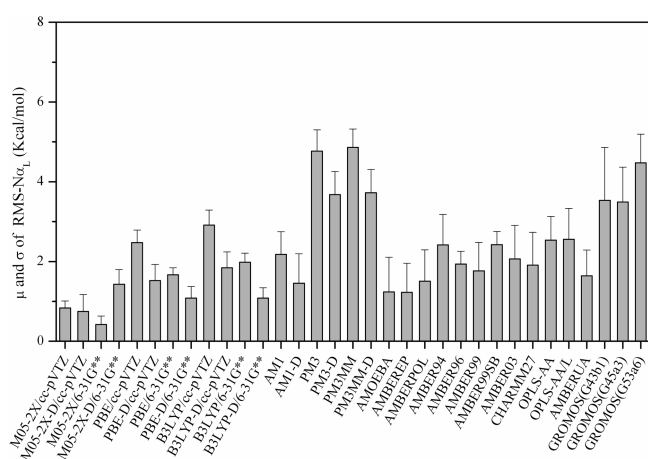
Figure 4. Mean (μ) and standard deviation (σ) of RMS of each amino acid calculated without α_L conformations for each method.

of AMBER polarizable force fields (with/without extra points) is not improved in comparison with their nonpolarizable AMBER versions (see Table 1). A systematic parametrization is necessary to improve the accuracy.

Unexpectedly, some of the additive all-atom force fields outperform semiempirical methods (see Table 2 for comparing the mean RMS-C of these force fields with the semiempirical methods). Large standard deviations of RMS-C are observed for AMBER96 (2.86 kcal/mol), AMBER03 (2.69 kcal/mol), and CHARMM27 (4.37 kcal/mol), indicating that these force fields have imbalanced descriptions on some conformations. Indeed, these three force fields have larger RMS-C's for the α_L conformation, being 9.55, 9.11, and 14.55 kcal/mol, respectively, than the other four conformations. This defect is probably caused by the less attention paid to the α_L conformation in the force field parametrization. The OPLS-AA/OPLS-AA/L and AMBER99/AMBER99SB force fields suffer such a defect less severely; their σ values are 0.74/1.23 and 0.93/0.78 kcal/mol, respectively. If excluding the α_L conformation, the mean RMS-C-Nα_L of AMBER96/99 and CHARMM27 are respectively 1.82/1.81 and 2.05 kcal/mol (Table S4c of SIA), which are smaller than the best semiempirical AM1 method (2.18 kcal/mol of AM1) but are still larger than the polarizable AMOEBA force field (1.50 kcal/mol). We emphasized that the tetrapeptide models used in this study are no longer the alanine dipeptide which is often used in force field parametrization, and they can mimic the H-bonds in protein helix secondary structures. For the AMBER series, the AMBER99 and AMBER96 perform slightly better than the others in the gas phase after excluding the α_L data. However, caution should be taken that AMBER99SB and AMBER03 were developed to implicitly include the solvent effect. Further evaluation in the condensed phase is necessary. For the OPLS series of force fields, it is unexpected that OPLS-AA/L performs slightly worse than OPLS-AA, because the torsion parameters of the former were refined using QM-based conformational energies of a large amount of different conformations in the gas phase. Again, further evaluation in the condensed phase is necessary to provide a more reasonable assessment.[18]

In the category of the united atom force fields, AMBERUA is comparable to the GROMOS96 series when all five conformations are included. But when α_L is excluded, AMBERUA performs much better than the GROMOS series (~1.7 kcal/mol vs ~4.0 kcal/mol of mean RMS-C). Since the versions of G45a3 and G53ab were optimized in the solution phase, further evaluation in the solution phase is necessary. In addition, we argue that other important aspects are needed to be considered for a comprehensive assessment of the force fields (especially for the additive ones): inter-peptide interactions,[66,79,80] peptide–water interactions,[14,71] thermodynamic properties,[15,16] even kinetic properties, etc. In fact, the GROMOS force fields, which perform worse than others in this study, have been successfully applied to many protein simulations.

The means (μ) of RMS (Table S3d, Supporting Information) and RMS-Nα_L (Table S4d/Table 3) give the same information on the overall performance of the examined methods as those of RMS-C (Table S3c/Table 2) and

RMS-C-Nα_L (Table S4c) because they originate from the same data sets. However, the individual RMS/RMS-Nα_L can tell us the performance of the examined methods on the individual amino acid. It is well-known and we also confirmed that the MM methods are not able to describe the α_L conformation properly. We thus use RMS-Nα_L (Table 3 and Figure 4) for the following discussion, since α_L is not important for modeling the native protein structure. Understandably, the QM methods (M05-2X, PBE, B3LYP, AM1, and PM3) are generally more consistent in describing all 20 amino acids, although some of them (e.g., PBE/cc-pVTZ, B3LYP/cc-pVTZ, PM3, and PM3MM) have poor overall performance. The polarizable AMOEBA, which has overall good performance, is unsatisfied with some amino acids such as Pro, Asp, and Glu in particular. The developer needs to pay attention to these problematic amino acids. For the additive force fields, AMBER96 gives more consistent descriptions to all the amino acids than the other additive force fields. Its standard deviation, 0.32 kcal/mol, is comparable with those of the M05-2X methods, although the overall performance of the force field is not as good as those of the M05-2X methods. The readers can refer the Table 3 to identify the problematic amino acids for other force fields.

4. Conclusions

Using 100 tetrapeptide structures which cover all 20 amino acids and five major conformations (α_R, α_L, β, β_a, and PPII), we estimated their conformation energies in the gas phase at the MP2/cc-pVTZ//B3LYP/6-31G** level. The results indicate that the energetic patterns (the order and the energy gap) of the five conformations bear certain resemblances among the amino acids in the same class but is quite different among the amino acids in the different classes (e.g., hydrophobic, aromatic, polar, and charged classes). Using these MP2 energies of 100 tetrapeptide structures as "standard", we further evaluated the performance of various methods in terms of RMS and RMS-C and draw the following conclusions: (1) The M05-2X DFT functional outperforms PBE and B3LYP. (2) The semiempirical methods (AM1, PM3, and PM3MM) are not accurate enough to describe the relative energies of the conformations. (3) The AMOEBA polarizable force field outperforms the semiempirical methods and the B3LYP method. However, the current AMBER polarizable force fields do not improve the accuracy with respect to the related additive versions, which suggest a systematic parametrization is necessary to improve the accuracy. (4) The additive force fields are less accurate than the three DFT methods, but some of them are more accurate than the semiempirical methods. (5) If excluding the α_L conformation, the examined force fields have comparable performance; the RMS-C means are 2.4 kcal/mol for AMBER94, 1.8 kcal/mol for AMBER96/99, 2.3 kcal/mol for AMBER99SB, 2.2 kcal/mol for AMBER03, 2.0 kcal/mol for CHARMM27, and 2.5 kcal/mol for OPLS-AA and OPLS-AA/L. However, it should be pointed out that some of the force fields are parametrized to include the solvent effects implicitly, while our calculations were carried out in

Popular QM and MM Methods

J. Chem. Theory Comput., Vol. 6, No. 4, 2010 **1207**

the gas phase. (6) If excluding the α_L conformation, the united atom force field (AMBERUA) with a 1.7 kcal/mol of mean RMS-C, has an accuracy comparable with that of the all-atom force field. (7) With respect to the MP2 energies, overestimating the energies of the compact helical conformations (α_R and α_L), but underestimating those of the extended conformations (β, β_a, and PPII), is a general error trend for methods M05-2X/cc-pVTZ, PBE and B3LYP, AM1, AMBER99SB, OPLS-AA, and OPLS-AA/L and GROMOS. (8) Semiempirical and empirical force field methods perform poorly on Pro and the charged amino acids.

The structures and energies of the 100 tetrapeptide structures can serve as a database to systematically develop/calibrate force fields for modeling proteins. In addition to the data provided in the Supporting Information, other preliminary data such as the Cartesian coordinates of the 100 tetrapeptides are available upon request.

Supporting Information Available: The main-chain and side-chain dihedral angles adopted in the calculations (Table S1 of SIA). The relative energies of α_L, β, β_a, and PPII to α_R at all considered levels (Table S2 of SIA and Table S1 of SIB). The reference energy offsets ($E_C/E_C - N\alpha_L$) to minimize RMS/RMS-Nα_L (Table S3a/Table S4a of SIA and Table S2a/Table S3a of SIB). The signed energy errors (error = $E_{ai} - E_{bi} + E_c$) and errors (Table S3b of SIA and Table S2b of SIB)/error-Nα_L (Table S4b of SIA and Table S3b of SIB). The signed RMS-C/RMS-C-Nα_L values (Table S3c/Table S4c of SIA and Table S2c and Table S3c of SIB). The unsigned RMS/RMS-Nα_L values (Table S3d/Table S4d of SIA and Table S2d/Table S3d of SIB). Representative geometries for five conformations of each tetrapeptide (Figure S1 of SIA). Relative energies of the MP2/cc-pVTZ//M05-2X/6-31G** method (Figure S1 of SIB). The signed energy errors, signed RMS-C/RMS-C-Nα_L values, and RMS values of each method are plotted in Figure S2-S34 of SIA and Figure S2-S34 of SIB. Mean and stand deviation of RMS-C of each conformation calculated over 20 amino acids relative to MP2/cc-pVTZ//M05-2X/6-31G** method (Figure S35 of SIB). Comparison of the RMS-Nα_L using two set of geometries (B3LYP/6-31G** and M05-2X/6-31G**; Figure S36 of SIB). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

(2) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.

(3) Axel, D. B. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(4) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

(5) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(6) Yan, Z.; Donald, G. T. *J. Chem. Phys.* **2006**, *125*, 194101–194118.

(7) Hohenstein, E. G.; Chill, S. T.; Sherrill, C. D. *J. Chem. Theory Comput.* **2008**, *4*, 1996–2000.

(8) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289–300.

(9) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(10) James, J. P. S. *J. Comput. Chem.* **1989**, *10*, 209–220.

(11) James, J. P. S. *J. Comput. Chem.* **1989**, *10*, 221–264.

(12) Jenn-Huei, L.; Norman, L. A. *J. Comput. Chem.* **1991**, *12*, 186–199.

(13) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(14) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(15) Schuler, L. D.; Daura, X.; Gunsteren, W. F. v. *J. Comput. Chem.* **2001**, *22*, 1205–1218.

(16) Oostenbrink, C.; Villa, A.; Mark, A. E.; Gunsteren, W. F. V. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(17) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(18) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(19) Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. *J. Comput. Chem.* **2006**, *27*, 781–790.

(20) Dixon, R. W.; Kollman, P. A. *J. Comput. Chem.* **1997**, *18*, 1632–1646.

(21) Guillaume, L.; Benoit, R. *J. Chem. Phys.* **2003**, *119*, 3025–3039.

(22) Lamoureux, G.; MacKerell, A. D.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185–5197.

(23) Geerke, D. P.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2007**, *3*, 2128–2137.

(24) Jorgensen, W. L.; Jensen, K. P.; Alexandrova, A. N. *J. Chem. Theory Comput.* **2007**, *3*, 1987–1992.

(25) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. A* **2003**, *108*, 621–627.

(26) Friesner, R. A.; Robert, L. B.; David, B. *Adv. Protein Chem.* **2005**, *72*, 79–104.

(27) Yang, Z.-Z.; Wang, C.-S. *J. Phys. Chem. A* **1997**, *101*, 6315–6321.

(28) Ponder, J. W.; Case, D. A.; Valerie, D. *Adv. Protein Chem.* **2003**, *66*, 27–85.

(29) Nohad, G.; Sherif, A. K.; Jean-Francois, T.; Dennis, R. S. *J. Comput. Chem.* **2004**, *25*, 823–834.

(30) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.

1208 *J. Chem. Theory Comput., Vol. 6, No. 4, 2010*

Jiang et al.

(31) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(32) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L.; Pople, J. A. *J. Am. Chem. Soc.* **1991**, *113*, 5989–5997.

(33) Viviani, W.; Rivail, J. L.; Perczel, A.; Csizmadia, I. G. *J. Am. Chem. Soc.* **1993**, *115*, 8321–8329.

(34) Alemán, C.; Casanovas, J. *J. Chem. Soc., Perkin Trans. 2* **1994**, 563–568.

(35) Gould, I. R.; Cornell, W. D.; Hillier, I. H. *J. Am. Chem. Soc.* **1994**, *116*, 9250–9256.

(36) Gronert, S.; O'Hair, R. A. J. *J. Am. Chem. Soc.* **1995**, *117*, 2071–2081.

(37) Cornell, W. D.; Gould, I. R.; Kollman, P. A. *THEOCHEM* **1997**, *392*, 101–109.

(38) Mohle, K.; Gussmann, M.; Rost, A.; Cimiraglia, R.; Hofmann, H. J. *J. Phys. Chem. A* **1997**, *101*, 8571–8574.

(39) Kaschner, R.; Hohl, D. *J. Phys. Chem. A* **1998**, *102*, 5111–5116.

(40) Aleman, C. *J. Phys. Chem. A* **2000**, *104*, 7612–7616.

(41) Barroso, M. N.; Cerutti, E. S.; Rodríguez, A. M.; Jáuregui, E. A.; Farkas, O.; Perczel, A.; Enriz, R. D. *THEOCHEM* **2001**, *548*, 21–37.

(42) Masman, M. F.; Amaya, M. G.; Rodríguez, A. M.; Suvire, F. D.; Chasse, G. A.; Farkas, O.; Perczel, A.; Enriz, R. D. *THEOCHEM* **2001**, *543*, 203–222.

(43) Yu, C.-H.; Norman, M. A.; Schafer, L.; Ramek, M.; Peeters, A.; van Alsenoy, C. *J. Mol. Struct.* **2001**, *567−568*, 361–374.

(44) Aleman, C.; Jimenez, A. I.; Cativiela, C.; Perez, J. J.; Casanovas, J. *J. Phys. Chem. B* **2002**, *106*, 11849–11858.

(45) Vargas, R.; Garza, J.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* **2002**, *106*, 3213–3218.

(46) Andras, P.; Odon, F.; Imre, J.; Igor, A. T.; Imre, G. C. *J. Comput. Chem.* **2003**, *24*, 1026–1042.

(47) Klipfel, M. W.; Zamora, M. A.; Rodriguez, A. M.; Fidanza, N. G.; Enriz, R. D.; Csizmadia, I. G. *J. Phys. Chem. A* **2003**, *107*, 5079–5091.

(48) Chin, W.; Mons, M.; Dognon, J. P.; Mirasol, R.; Chass, G.; Dimicoli, I.; Piuzzi, F.; Butz, P.; Tardivel, B.; Compagnon, I.; vonHelden, G.; Meijer, G. *J. Phys. Chem. A* **2005**, *109*, 5281–5288.

(49) Masman, M. F.; Lovas, S.; Murphy, R. F.; Enriz, R. D.; Rodriguez, A. M. *J. Phys. Chem. A* **2007**, *111*, 10682–10691.

(50) Young Kee, K.; Nam Sook, K. *J. Comput. Chem.* **2009**, *30*, 1116–1127.

(51) Möhle, K.; Höfmann, H.-J. *J. Mol. Model.* **1998**, *4*, 53–60.

(52) Langella, E.; Rega, N.; Improta, R.; Crescenzi, O.; Barone, V. *J. Comput. Chem.* **2002**, *23*, 650–661.

(53) Schlund, S.; Müller, R.; Graβmann, C.; Engels, B. *J. Comput. Chem.* **2008**, *29*, 407–415.

(54) Böehm, H. J.; Brode, S. *J. Am. Chem. Soc.* **1991**, *113*, 7129–7135.

(55) Gould, I. R.; Kollman, P. A. *J. Phys. Chem.* **1992**, *96*, 9255–9258.

(56) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 5908–5920.

(57) Kaminsky, J.; Jensen, F. *J. Chem. Theory Comput.* **2007**, *3*, 1774–1788.

(58) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

(59) Improta, R.; Barone, V. *J. Chem. Phys.* **2001**, *114*, 2541–2549.

(60) Improta, R.; Barone, V.; Kudin, K. N.; Scuseria, G. E. *J. Am. Chem. Soc.* **2001**, *123*, 3311–3322.

(61) Ireta, J.; Neugebauer, J.; Scheffler, M.; Rojo, A.; Galvan, M. *J. Phys. Chem. B* **2003**, *107*, 1432–1437.

(62) Bogár, F.; S., Z.; Bartha, F.; Penke, B.; Ladik, J. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2965–2969.

(63) Ireta, J.; Neugebauer, J.; Scheffler, M.; Rojo, A.; Galvan, M. *J. Am. Chem. Soc.* **2005**, *127*, 17241–17244.

(64) Ireta, J.; Scheffler, M. *J. Chem. Phys.* **2009**, *131*, 085104−085108.

(65) Penev, E.; Ireta, J.; Shea, J.-E. *J. Phys. Chem. B* **2008**, *112*, 6872–6877.

(66) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(67) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.

(68) Yoda, T.; Sugita, Y.; Okamoto, Y. *Chem. Phys.* **2004**, *307*, 269–283.

(69) Shell, M. S.; Ritterson, R.; Dill, K. A. *J. Phys. Chem. B* **2008**, *112*, 6878–6886.

(70) Dunbrack, R. L. *Curr. Opin. Struct. Biol.* **2002**, *12*, 431–440.

(71) Wang, Z.-X.; Duan, Y. *J. Comput. Chem.* **2004**, *25*, 1699–1716.

(72) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; N. Rega; Petersson, G. A.; 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, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision E.01; Gaussian, Inc.: Wallingford, CT, 2004.

(73) Ponder, J. W. *Tinker*, version 4.2; Biochemistry & Molecular Biophysics, Washington University School of Medicine: St. Louis, MO, 2004.

(74) Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. In *Computer Simulation of Biomolecular Systems*; Wilkinson, A.,Weiner, P., van Gunsteren, W. F., Eds.; Elsevier: New York, 1997; Vol. 3, pp 83−96.

Popular QM and MM Methods

J. Chem. Theory Comput., Vol. 6, No. 4, 2010 **1209**

(75) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.

(76) Yang, L.; Tan, C.-h.; Hsieh, M.-J.; Wang, J.; Duan, Y.; Cieplak, P.; Caldwell, J.; Kollman, P. A.; Luo, R. *J. Phys. Chem. B* **2006**, *110*, 13166–13176.

(77) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross; W. S. Kollman, P. A. *Amber 9*; University of California: San Francisco, 2006.

(78) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(79) Alexander, D.; Mackerell, M., Jr. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(80) Wang, Z.-X.; Wu, C.; Lei, H.; Duan, Y. *J. Chem. Theory Comput.* **2007**, *3*, 1527–1537.

CT100008Q