PROTEINS: Structure, Function, and Genetics 50:451–463 (2003)

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

Hao Hu,[1] Marcus Elstner,[2] and Jan Hermans[1]*

[1]*Department of Biochemistry and Biophysics, School of Medicine, University of North Carolina, Chapel Hill, North Carolina*
[2]*Department of Theoretical Physics, University of Paderborn, Paderborn, Germany*

ABSTRACT **We compare the conformational distributions of Ace-Ala-Nme and Ace-Gly-Nme sampled in long simulations with several molecular mechanics (MM) force fields and with a fast combined quantum mechanics/molecular mechanics (QM/MM) force field, in which the solute's intramolecular energy and forces are calculated with the self-consistent charge density functional tight binding method (SCCDFTB), and the solvent is represented by either one of the well-known SPC and TIP3P models. All MM force fields give two main states for Ace-Ala-Nme, β and α separated by free energy barriers, but the ratio in which these are sampled varies by a factor of 30, from a high in favor of β of 6 to a low of 1/5. The frequency of transitions between states is particularly low with the amber and charmm force fields, for which the distributions are noticeably narrower, and the energy barriers between states higher. The lower of the two barriers lies between α and β at values of ψ near 0 for all MM simulations except for charmm22. The results of the QM/MM simulations vary less with the choice of MM force field; the ratio β/α varies between 1.5 and 2.2, the easy pass lies at ψ near 0, and transitions between states are more frequent than for amber and charmm, but less frequent than for cedar. For Ace-Gly-Nme, all force fields locate a diffuse stable region around $\phi = \pi$ and $\psi = \pi$, whereas the amber force field gives two additional densely sampled states near $\phi = \pm 100°$ and $\psi = 0$, which are also found with the QM/MM force field. For both solutes, the distribution from the QM/MM simulation shows greater similarity with the distribution in high-resolution protein structures than is the case for any of the MM simulations. Proteins 2003;50:451–463.**

Key words: dynamics simulation; alanine dipeptide; glycine dipeptide; QM/MM model; peptide backbone; solution conformation; Ramachandran plot

INTRODUCTION

At present, simulations with molecular mechanics (MM) force fields, such as amber,[1] charmm,[2] gromos,[3] and opls,[4] offer a comprehensive approach to modeling biological macromolecules in atomic detail over time spans that are commensurate with the relaxation times of these molecules in their native state. Even then, because of limits on available computing power, simulations cannot be performed for sufficiently long times to be able to follow many important processes, such as protein folding and conformation changes of allosteric molecules.

The energetics computed according to a molecular mechanics force field are meant to replace the underlying quantum mechanical (QM) energetics; accordingly, the design (which includes both the form and the values of the associated parameters) of such a force field is often wholly or partly based on a comparison with accurate quantum mechanical calculations, which, per force, can only be conducted for systems with relatively small numbers of atoms.[2,5,6] An alternative (and oldest) route to determine the best values of force field parameters is to impose agreement between measured physical properties and the results of simulations (e.g., Refs. 7–9). In fact, the theoretical and empirical approaches to force-field development can be effectively combined, with the former able to give more accurate parameters for geometric deformation (bond stretching, angle bending), atomic partial charges and coefficients for repulsive forces due to atomic overlap, and the latter giving the more accurate estimates of weak long-range attractions (dispersion terms).

Force fields should be used with an understanding of their accuracy, which can be assessed by comparison of

This article contains Supplementary Materials that can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/2003/50/v50.451.html.

Grant sponsor: National Center for Research Resources, U.S. National Institutes of Health; Grant number: RR08012.

*Correspondence to: Jan Hermans, Department of Biochemistry, University of North Carolina, Chapel Hill, NC 27599-7260. E-mail: hermans@med.unc.edu

Received 3 June 2002; Accepted 19 August 2002

simulated and experimental properties of a few well-characterized model systems. Accuracy is then found to vary, depending on the composition of the simulated system as well as on the physical properties of interest.

In the present study, we set out to enquire into the quality of commonly used force fields with respect to the properties of unfolded polypeptide chains in solution, in terms of preferred conformation and conformational statistics. These are of particular interest in connection with the physical properties of unfolded proteins and with the kinetics and equilibria of protein folding. Debated questions about the "structure" remaining in denatured proteins, even in the presence of strong denaturants, can presumably be illuminated by accurate long simulations of peptides in solution (e.g., Ref. 10), whereas a simulation of protein folding (e.g., Ref. 11) can only be successful given an accurate free energy balance between the ensemble of denatured conformations and the folded state.

As a model system to investigate the performance of molecular mechanics force fields in simulating the unstructured polypeptide chain, we have chosen the terminally blocked amino acid, or "dipeptide" model. Advantages of using this model include the ease with which a reasonably precise conformational distribution (in the two backbone dihedral angles ϕ and ψ) can be obtained simply by extended simulations, the ability to compare the resulting distribution with the distribution of backbone conformation in proteins whose structure is accurately known, and (not in the last place) its familiarity. A disadvantage of using the dipeptide model as a basis is a shortage of experimental data that can report on conformational distributions of flexible molecules in aqueous solution. Although there has long been experimental evidence that the most prevalent conformation of the alanine residue in unfolded polypeptide chains is a "polyproline II" conformation (P_{II}), with ϕ near $-70°$ and ψ near $140°$,[12] this has only recently been confirmed with different approaches.[13–16] This finding agrees with results of an early simulation study of Ace-Ala-Nme in water[17] but less well with a more recent simulation based on a different force field.[18] Disagreement between (in the first instance equivalent) models tends to render these irrelevant to the experimentalist.[15]

Use of accurate quantum-mechanical energies to obtain MM force-field parameters that adequately describe the polypeptide backbone is complicated by the problem that these calculations can be performed only for a limited number of configurations of systems consisting of small numbers of atoms. This would suggest as the model system a molecule such as Ace-Ala-Nme in vacuo. However, as was shown some time ago, the free energy surface of this molecule in vacuo is very different from that in aqueous solution.[19] This finding suggests why it is difficult to design an MM energy function that fits the QM energy surface of Ace-Ala-Nme in vacuo and that is at the same time accurate in the highly polar environment of an aqueous solution or a folded protein,[1,2,5,6] and especially so when the atomic charges in this MM force field are held fixed.

It appears much preferable to use as a QM model system one in which the highly polar solvent also is represented. Obstacles to working with such a model system are (i) the much larger number of atoms in the model, (ii) the fact that for a given conformation of the solute, an ensemble of conformations of the solvent must be considered, and (iii) the need to obtain adequate statistics to assess the relative importance of different conformation states separated by free energy barriers. For the alanine model, Ace-Ala-Nme, this requires of the order of 6 ns of simulation with molecular dynamics, which perhaps can be reduced to 1 ns with optimal use of computation of potentials of mean force along paths connecting free energy minima. Because this appears not feasible with current QM methods, we have resorted to simulation of two of these systems (Ace-Ala-Nme and Ace-Gly-Nme, both in water) with a QM/MM method, in which the solute is treated with quantum mechanics, but the solvent and the solute-solvent interactions are treated with molecular mechanics. For the solvent, we have chosen the (very similar) SPC and TIP3P models,[20,21] which are known to adequately represent many properties of liquid water and which, in conjunction with molecular mechanical descriptions of small molecules, have proven able to represent the solvation of these.

The solute is represented with the self-consistent tight binding (SCCDFTB) method, a fast approximate quantum-mechanical method.[22,23] This method was recently used with good result in a 300-ps simulation of crambin in solution, in which the crambin molecule was treated with the SCCDFTB method, the solute-solvent interactions with the amber force field, and the solvent with the TIP3P water model.[24] Another recent study has used the same approach, except that the solute-solvent interactions were computed with the charmm22 force field.[25] That article includes simulations of peptide helices in solution that show a realistic tendency of formation of α-, rather than 3_{10} helical structure as the molecules become longer.

To better evaluate the contents of this article, it is useful to briefly review different methods of calculating the energy of macromolecular systems. MM methods use an artificial decomposition of the energy into local terms (e.g., energy terms for bond stretching and bond angle bending, local torsional energy terms, and Lennard–Jones and Coulomb pair energies). Use of a QM representation for the solute unifies the energy expression in a single Hamiltonian, requires no a priori information about molecular geometry, and inherently represents complex effects, such as nonadditivity of terms, changes of polarization coupled to changes of geometry and polarization caused by the local electrostatic field due to intra- and intermolecular interactions. The design and development of a fast approximate QM method, based on judicious approximations that retain these advantages as well as high accuracy is complex. Approximations introduced in SCCDFTB include explicit treatment of only the valence electrons and use of a minimal basis set of pseudoatomic orbitals. Use of precomputed Hamiltonian matrix elements for pairs of atom types leads to a significant speedup. The charge

density is written as a sum of the charge density of neutral atoms (the number of valence electrons) and atomic charge fluctuations (atomic polarization). The Hamiltonian contains a sum of pairwise terms representing the interaction between charge fluctuations, according to an expression that produces the Coulomb energy at large interatomic distance but takes account of exchange-correlation contributions at shorter separation. A double sum of pairwise atom–atom potentials is included in the Hamiltonian, fit to represent the difference between the energy from a high-level DFT calculation and the SCCDFTB electronic energy. The two cited application articles contain more detailed summaries of the method followed in the SCCDFTB calculation.[24,25] For a detailed description of the SCCDFTB methodology, see Refs 22 and 23.

It is not easy to predict the effect of such approximations, and there are few examples of simulations of peptides or proteins in water with QM/MM methods. Therefore, the SCCDFTB method has been tested for various biological model systems, including H-bonded complexes, peptides and DNA bases.[23] Special emphasis was put on the investigation of structures and relative energies of peptides with up to eight amino acid residues in the gas phase. A comparison with results from DFT and MP2 calculations has shown that the SCCDFTB method can reproduce structures and energetics of polypeptides reliably, with an accuracy comparable to that of the higher-level methods.[26,27]

In addition, vibrational frequencies of Ace-Ala-Nme have been investigated for SCCDFTB in comparison with DFT and MP2, and only slightly larger deviations of 6.7% from experimental values have been reported than for DFT (4.4%) and MP2 (3.0%).[28] These studies indicate that the SCCDFTB method describes the potential energy surface around the local minima in the gas phase with good accuracy. However, one lacks the insight and experience needed to know whether system properties are well represented far from the local minima or in solution, especially when the solvent is represented by the SPC or TIP3P models with their well-known shortcomings. Therefore, the present study is in the nature of an exploration, in which we ask how this particular QM/MM method describes the system, and then compare the results with ones of MM simulations and with available experimental data.

In what follows, we present and compare the distributions for alanine and glycine models obtained with simulations with the SCCDFTB/MM method and with simulations with several MM force fields. Because direct experimental information on the conformation of these molecules in solution is so sparse, we compare the results of the simulations also with distributions of conformations of alanine and glycine residues in the database of high-resolution protein structures of known structure as a second, if less direct source of experimental information.

MATERIALS AND METHODS

Dynamics Simulations

With the exceptions noted below, molecular dynamics simulations were performed with the Sigma program.[29] One molecule of Ace-Ala-Nme or Ace-Gly-Nme and 362 water molecules were simulated in a cubic periodic box (with edge slightly >22 Å). Simulations were performed with a multiple timestep scheme,[30] with a basic timestep of 2 fs, doubled to 4 fs for nonbonded interactions at separation between 6 and 11 Å, and again doubled to 8 fs for long-range electrostatic forces calculated via particle-mesh Ewald summation.[31,32] No significant difference was noticed when the Ewald summation was omitted, and thus Ewald summation was not used in all calculations. In addition, no significant difference was noted when the number of water molecules was increased. The nonbonded pairlist was updated every 32 fs. Pressure was maintained at 1 bar and temperature at 300 K with Berendsen manostat and thermostat (with separate thermostats for solute and solvent) using relaxation times of 0.1 ps.[33] Bond lengths were held fixed with the Shake algorithm.[34] The dihedral angles ϕ and ψ were monitored at regular intervals. In each simulation, the choice of water model (TIP3P or SPC) is dictated by the choice of force field.

Simulations were started from extensively equilibrated coordinate sets.

Force Fields

Amber

Simulations with the parm98 amber force field[1] were performed starting with amber topology files prepared for us by Drs. Yong Duan (Ace-Ala-Nme) and Lee Bartolotti (Ace-Gly-Nme). The 1–4 interactions were scaled by division by factors of 2.0 (Lennard–Jones energy) and 1.2 (Coulomb energy). The water model was TIP3P.[21] The accuracy of Sigma's implementation of the amber force field was checked by us by comparison with results of a simulation of Ace-Ala-Nme done with the amber program.[35]

Charmm22

Simulations with the charmm22 force field[2] were performed with use of a protein structure file prepared with the X-plor program[36] using the charmm22 protein dictionary and the charmm22 all-atom parameter files distributed with that program. The water model was TIP3P. The distribution obtained for Ace-Ala-Nme with the Sigma program agrees with that obtained by Smith.[18]

Cedar

Simulations with the cedar all-atom force field[37,38] were performed by using a protein structure file prepared with the X-plor program and the cedar all-atom topology and parameter files. The water model was SPC.[20]

Gromos

Simulations with the gromos96 force field were performed with the gromos96 program.[3] In this one case, the CH and CH_3 groups of the solute were represented as single centers for nonbonded force calculations. The simulations with the gromos force field were performed also in a cubic periodic system but with 2795 (Ace-Ala-Nme) or 1497 water molecules (Ace-Gly-Nme). The water model

was SPC. (Larger numbers of water molecules were used in the simulations with gromos and with QM/MM as a result of inadequate communication between the authors.)

OPLS

A simulation of Ace-Ala-Nme with the all-atom opls-aa force field[4,6] was performed by starting with a topology file in amber format prepared for us by Dr. Julian Tirado-Rives. The 1–4 interactions were scaled by division by factors of 2.0 (Lennard–Jones energy) and 1.2 (Coulomb energy). The water model was TIP3P.

QM/MM

The QM/MM simulations were performed with the SCCDFTB method,[23] with code and data files incorporated into the Sigma program. As was done in an earlier study of crambin from this laboratory,[24] the intramolecular forces of the solute were computed with the SCCDFTB method, the solvent-solvent interactions were calculated as appropriate for the (SPC or TIP3P) water model, electrostatic interactions between solvent and solute were evaluated as part of the SCCDFTB calculation, and Lennard–Jones interactions between solvent and solute were computed with molecular mechanics by using nonbonded parameters for water-solute interactions from one of these molecular mechanics force fields: charmm22, amber or cedar, with the water model (SPC or TIP3P) belonging to that force field. The QM/MM simulations were performed in a cubic periodic system, with 2795 (Ace-Ala-Nme) or 1497 water molecules (Ace-Gly-Nme). The near and far cutoffs were 8 and 12 Å, with timesteps of 1 and 3 fs. For technical reasons, no Ewald summation was applied, and the simulations were run at constant volume of 85,184 $Å^3$ (Ace-Ala-Nme) or 46,656 $Å^3$ (Ace-Gly-Nme), rather than at constant pressure.

The study of crambin showed that the application of explicit dispersion forces ($1/r^6$ energy terms) within the crambin molecule (which was in its entirety represented with the QM force field) was necessary to retain a native-like crambin structure during the simulation.[24] In that study, the dispersion terms were damped at short distance with a switching function.[39] Results of simulations of Ace-Ala-Nme in which such damped dispersion terms were applied by using the attractive Lennard–Jones parameters of the same force field that was used for the water-solute interactions (data not shown) depended much more strongly on the choice of MM force field than was the case when these terms were omitted. (We believe that this is due to the use of a switching function, and this is something we are investigating.) Because we find that in the MM simulations the mean dispersion energy of Ace-Ala-Nme is not significantly different for conformations with $\psi < 0$ and for conformations with $\psi > 0$, we report here only the results of simulations in which intramolecular dispersion forces have not been applied.

Obtaining Adequate Sampling

Ace-Ala-Nme

With several of the force fields the simulations explore conformation space (as represented by the dihedral angles ϕ and ψ) in a simulation time of 6 ns. With all force fields, the exploration of conformations with $-\pi < \phi < 0$ is effective; in that region, the distribution consists of two main clusters, one having $-\pi < \psi < 0$, and the other having $0 < \psi < \pi$, and the extent to which the two regions are fairly represented in the sample depends on how many transitions occur between these two regions of conformation space. The frequency of such transitions was analyzed for each of the simulations.

With the charmm22 and amber force fields, no sampling occurred in the region of positive ϕ (not counting sampling at values of ϕ close to π, that are part of the two main clusters). Separate simulations were run with these force fields, with the value of ϕ restrained to lie between 0 and π, and this served to locate a third cluster for both. The relative importance of this cluster was in both cases determined by a potential of mean force calculation in which the value of ϕ was forced to change by 2π.[17]

Ace-Gly-Nme

Because this solute is achiral, the probability density is the same for a given (ϕ, ψ) as it is for the inverted geometry, that is, ($-\phi$, $-\psi$). The results have been reported without consideration of the symmetry of the distribution, and the symmetry of the figures thus gives a qualitative sense of the extent of convergence of the distribution or lack thereof. As indicated by the approximate symmetry of the sample distribution, all simulations of Ace-Gly-Nme achieved reasonable sampling of the accessible conformation space.

Reproducibility

At the suggestion of a reviewer, we make available the bonded and nonbonded parameters as supplementary material. All elements needed to compute energy and forces, with exception of atomic coordinates, for the simulations of Ace-Ala-Nme and Ace-Gly-Nme with the amber, cedar, charmm, and opls force fields have been collected in the identical, self-documenting format, preceded by an annotated example and have in this form been made available as part of the supplementary material of this article. (These data can be used as input for the Sigma program.) For the gromos force field we have deposited the gromos topology file for Ace-Ala-Nme, plus a list of differences between the gromos topology file for Ace-Gly-Nme and the former file, with which the latter file can be regenerated, and refer the reader to the extensive documentation of the gromos force field.[3]

The SCCDFTB tables occupy >2 megabytes of storage on a unix system and contain no indications as to the meaning of the contents. The SCCDFTB code and the integral and spline tables for O, N, C, and H are available on request to the second author, Marcus Elstner (E-mail: m.elstner@dkfz-heidelberg.de). The tables used in this study are identified as "integral tables for O N C H, version 1, created 5/1999." The Sigma code with built-in SCCDFTB code but without these tables is available from either one of the other two authors or via the web site //femto.med.unc.edu/SIGMA.

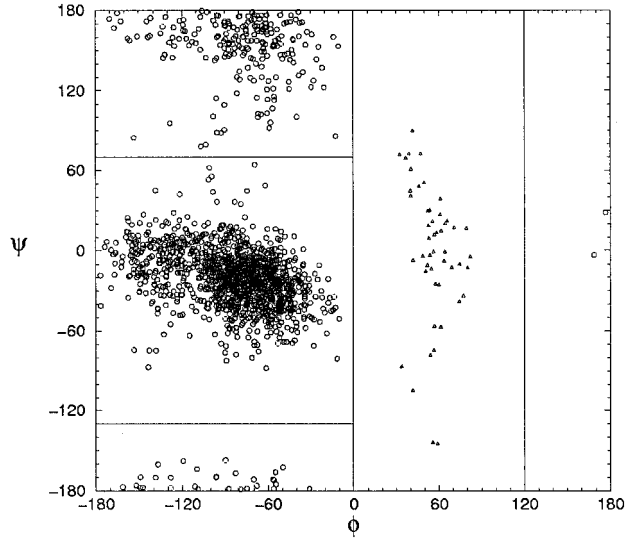

Fig. 1. Sampled conformational distribution of Ace-Ala-Nme with the amber force field. The circles and triangles represent results from two independent simulations that have been scaled together with a third simulation in which a potential of mean force was calculated.

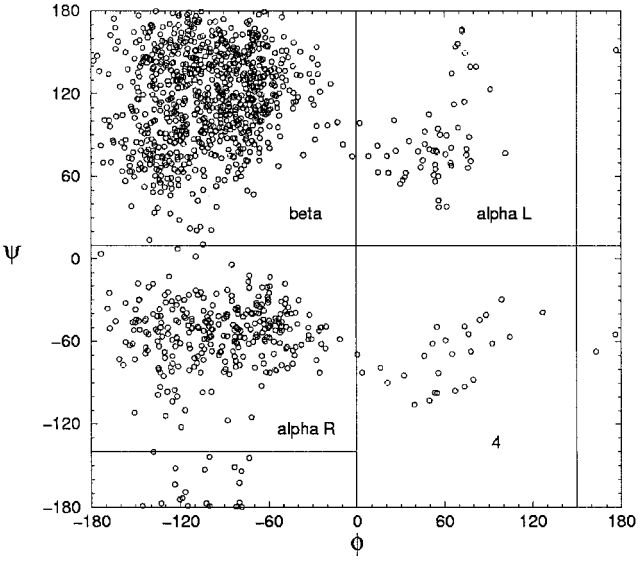

Fig. 2. Sampled conformational distribution of Ace-Ala-Nme with the cedar force field.

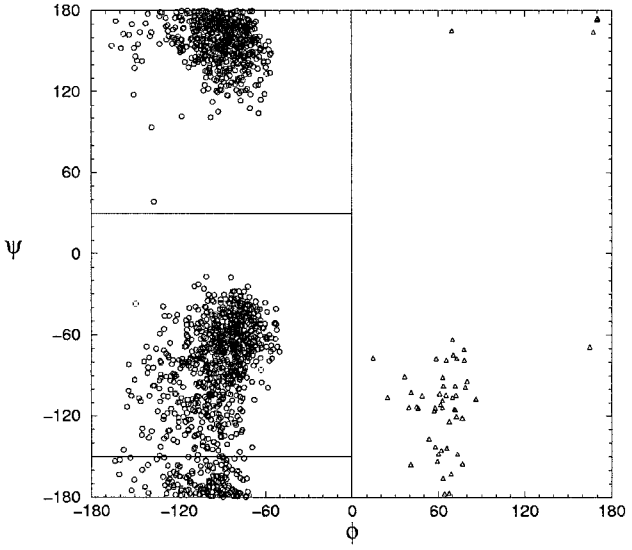

Fig. 3. Sampled conformational distribution of Ace-Ala-Nme with the charmm22 force field. See legend for Figure 1 for explanation of symbols.

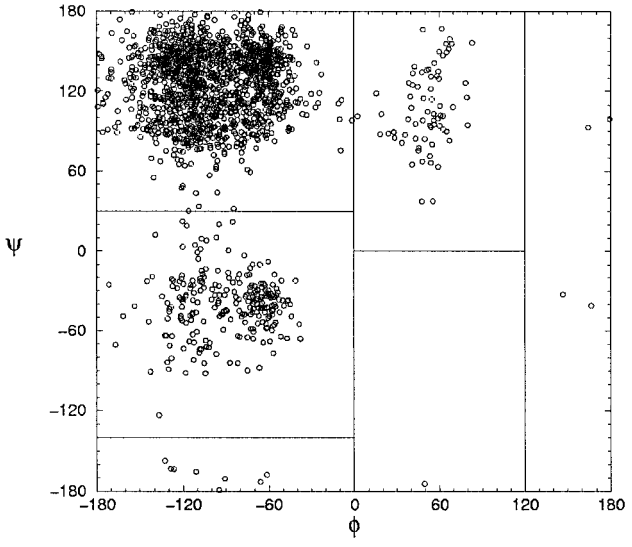

Fig. 4. Sampled conformational distribution of Ace-Ala-Nme with the gromos force field.

RESULTS

Ace-Ala-Nme

Figures 1–5 show the distributions from simulations of Ace-Ala-Nme in water with five different MM force fields, each for periods exceeding 6 ns. As can be seen, in all cases, conformation states with $\phi < 0$ predominate (we include the "overflow" at values slightly below π in these states), and within this category two states (one with $\psi < 0$ and the other with $\psi > 0$) are clearly separated. The simulations with the amber and charmm22 force fields did not at any time sample states with $\phi > 0$; these states occur sparsely in the samples obtained with the other three force fields: cedar, gromos, and opls. Additional simulations with the amber and charmm22 force fields, in which the conformation was prevented from assuming either of the two dominant states, showed a third locally stable conformation state for both amber and charmm22 (indicated with triangles in Figs. 1 and 3). The relative importance of this state was assessed by potential of mean force calculations in which the value of ϕ was forced through a range of 2π. These indicated that this state is at least 3 kcal/mol higher in free energy and that a barrier of at least 6 kcal/mol must be overcome in a transition from states centered at $\phi < 0$. For the other force fields, states centered at $\phi > 0$ also have higher free energy (by $\geq$1.5 kcal/mol), but passage to these states from conformation states centered at $\phi < 0$ is easier. The distribution obtained with the charmm22 force field is particularly tight and also is distinct from the others in

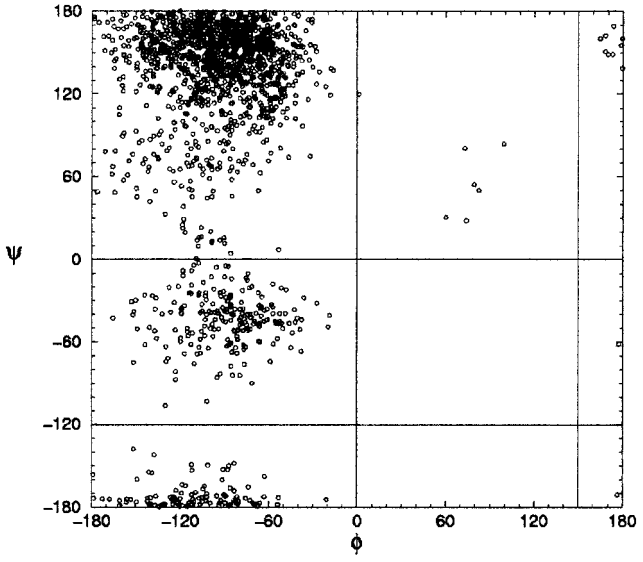

Fig. 5. Sampled conformational distribution of Ace-Ala-Nme with the opls-aa force field.

TABLE I. Sampling of Different Conformation States (Local Minima) of Ace-Ala-Nme in Simulations With MM Force Fields

Force field	Amber	Charmm22	Cedar	Gromos	Opls
beta	0.16	0.50	0.71	0.82	0.86
alpha R	0.84	0.50	0.22	0.13	0.135
alpha L	—	—	0.05	0.04	.004
state 4	—	—	0.02	0.0005	.0006

having the highest population density between the two principal states at values of ψ near $-150°$, whereas this "easy pass" lies near $+30°$ for the other four force fields. Finally, one notices that with the amber force field, the global maximum of the distribution occurs at $\psi < 0$, whereas this lies at a value of $\psi > 0$ for all others. The backbone conformation found in α-helices is close to the location of this maximum, and this leads one to suspect that simulation of oligopeptides with this force field may produce an exaggerated predominance of the α-helix conformation, something that has indeed been shown recently in a number of instances.[40] Table I shows the relative sampling of the conformation states separated by the lines drawn (somewhat subjectively) in each of the figures.

The QM/MM simulations were performed with each of three force fields determining the interaction of the MM water model with the solute. Results of these simulations are very similar; a single distribution is shown in Figure 6. With one exception, these samplings represent four different conformation states, with relatively easy passage between states. Table II shows the relative sampling of conformation states in these simulations; as indicated by the lines drawn in Figure 6, we have defined a fifth state, labeled "pass," close to the internally hydrogen-bonded $C7_{eq}$ conformation (the global minimum in vacuo). The distributions found with the QM/MM simulations more

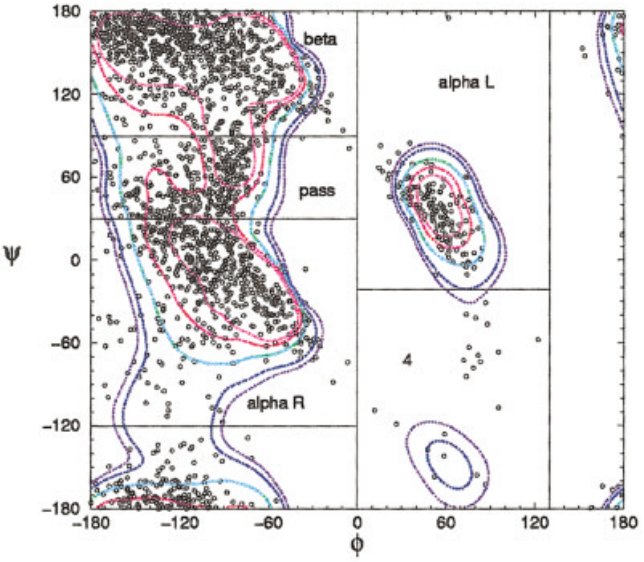

Fig. 6. Sampled conformational distribution of Ace-Ala-Nme with QM/MM (SCCDFTB/amber.) Successive contours enclose 99.8% (purple), 99.5%, 98%, 95%, and 90% (pink) of the data points for alanine residues in the new data base.[41]

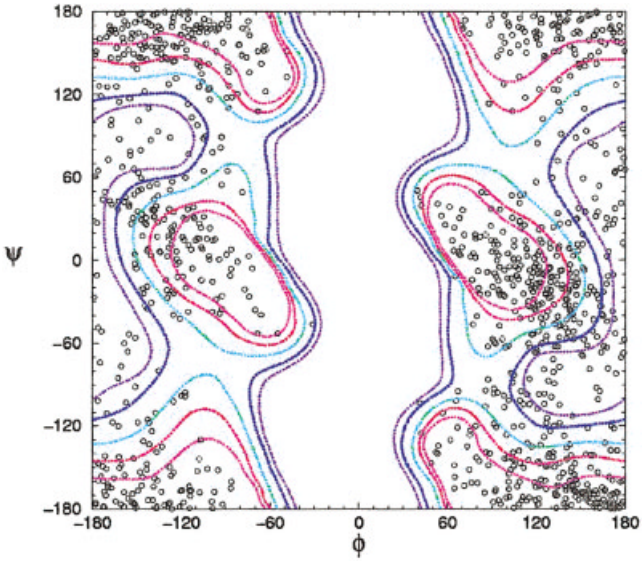

Fig. 12. Sampled conformational distribution of Ace-Gly-Nme with QM/MM (SCCDFTB/cedar; 7-ns simulation.) Successive contours enclose 99.8% (purple), 99.5%, 98%, 95%, and 90% (pink) of the data points for glycine residues in the new data base.[41]

strongly resemble those obtained with the cedar, gromos, and opls force fields. For these and for the QM/MM simulations, the most stable state is centered at $\phi < 0$ and $\psi > 0$, in distinction to what is obtained with the amber force field, and the easy pass lies at positive ψ, in distinction to what is obtained with the charmm22 force field.

More than any of the MM distributions, the QM/MM distribution resembles that of alanine residues in a database of well-ordered residues in high-resolution X-ray structures, as updated in the accompanying paper by Lovell et al.[41] Contours enclosing successively more "fa-

TABLE II. Sampling of Different Regions in Conformation Space of Ace-Ala-Nme in QM/MM Simulations

Force field	SCCDFTB cedar	SCCDFTB amber	SCCDFTB charmm22
beta	0.61	0.48	0.48
pass	0.12	0.16	0.14
alpha R	0.27	0.27	0.33
alpha L	—	0.07	0.03
state 4	—	0.01	0.01

vored" regions, encompassing, respectively, 99.8%, 99.5%, 98%, 95%, and 90% of the data points for alanine residues in the data base are included in Figure 6. (Alanine residues in repetitive secondary structure and preproline alanines had been excluded from the contoured data set. These contours were computed as described by Lovell et al.[41]) One notes in particular, that the distributions near ($\phi = -60$, $\psi = 0$) and ($\phi = 60$, $\psi = 0$) scatter around axes that make angles of about 45° with the coordinate axes, as is the case for the distribution in the database, but not for any of the MM simulations.

To assess the extent of convergence of these distributions, cumulative averages were calculated of the fraction of (recorded) instances in which the conformation fell within the limits of the various conformations designated as "beta," "alpha-R," and so forth. Results for the SCCDFTB simulation of Figure 6 have been plotted in Figure 7. For each conformation, periods of increasing and decreasing fraction alternate. The number of transitions to conformations with positive values of ϕ ($0 < \phi < 130°$) is only 7, and the estimated prevalence of 8% for this set of conformations (Table II) obviously has a large relative uncertainty.

For the MM force fields, the number of transitions between beta and alpha R regions was established in a similar manner. This number was lowest for the simulation with the amber force field (seven transitions); it was twice as large for charmm22 and higher still for the other three force fields. A second, 40-ns-long simulation with the amber force field showed close to 50 transitions, and the fraction of times at which the alpha-R conformation was sampled was 0.87, versus 0.84 in the first simulation (cf. Table II). In conclusion, the precision of the distributions achieved in these simulations appears adequate for purposes of comparison of the kind made in this article.

Ace-Gly-Nme

Figures 8–11 show the distributions from simulations of Ace-Gly-Nme in water with four different MM force fields, each for a period exceeding 1 ns. As can be seen, all simulations produce a broadly distributed set of conformations having both ϕ and ψ near π. In addition, the simulation with the amber force field produces two additional states, with ψ near 0.

The QM/MM sample distribution is given in Figure 12, which also includes contours of the database distribution of Lovell et al.[41] for glycine residues, as described above for alanine residues. Although the QM/MM force field and the MM force fields all sample a state centered at $\phi = \psi = \pi$, the QM distribution also samples a second and third state centered at $\phi = \pm 120°$ and ψ near 0°, including conformations that are only sparsely sampled in simulations with all MM force fields except amber. This is a qualitative difference; remarkably, the distribution of glycine residues in the database of well-ordered residues in high-resolution globular proteins[41] shows dense sampling in *all three* areas. The correspondence between database and simulation results is less marked than for alanine.

DISCUSSION

A Failure of MM Force Fields

The results presented here show that molecular mechanics force fields represent the conformation of unfolded polypeptide chains in aqueous solution with insufficient accuracy. This conclusion follows from the wide spread of results with different force fields, rather than from a direct comparison with experimental data. In fact, for several of the force fields, the most common conformation of the alanine residue in solution is a more or less spread-out region with a relatively high density near the P_{II} conformation (ϕ near $-70°$ and ψ near 140°), the conformation detected with experimental studies,[12–16] but for one force field (amber parm98) the α_R conformation, with ϕ near $-60°$ and ψ near $-50°$, dominates, and for another (charmm22) the α_R state is equally populated. In this context, it is useful to mention that the single-residue model, Ace-Ala-Nme, in first approximation provides an adequate representation of the backbone of an Ala residue in a modeled unfolded chain composed of several alanine residues; this can be seen by comparing the results of simulations with the same force field (gromos96) for Ace-Ala-Nme (Fig. 4 of this article), for Ala_3 (Fig. 2 of Ref. 42), and for Ala_8 (Fig. 1 of Ref. 10). Although that may progressively become less true of longer chains and for residues other than alanine, it seems safe to assume that, quite generally, an unfolded polypeptide cannot be modeled correctly with an MM force field that fails the simplest members of the family.

The MM force fields used in this article have been developed by following very similar principles, as described in Introduction; consequently, unless *all* agree, *all* fail to provide guidance in interpretation of experimental studies.[15] Of course, the models have been found inadequate only under one specific set of physical conditions (the unfolded backbone in solution), under which the properties obviously are sensitive to small changes in the energy function, and we do not for a moment suggest that the entire body of extant simulations of biological macromolecules with MM force fields has somehow become invalidated. A recent study indicates that different MM force fields behave comparably in tests with a different set of systems.[43] Nevertheless, the use of many different force fields having basically the same form, but with different parameters, each by a different set of laboratories, should be considered an embarrassment.

Origin of the Discrepancies

On the energy surface of Ace-Ala-Nme in vacuo, the two conformations of lowest energy are those for which the two

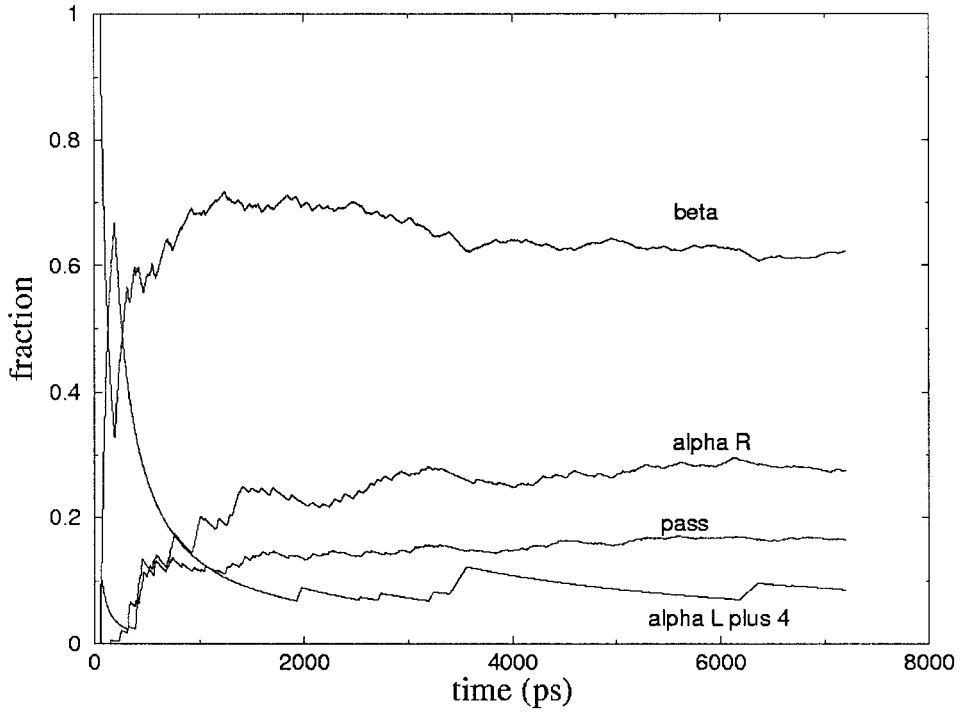

Fig. 7. Cumulative average of the fraction of time each of four conformations occurs in the simulation of Ace-Ala-Nme with QM/MM of Figure 6.

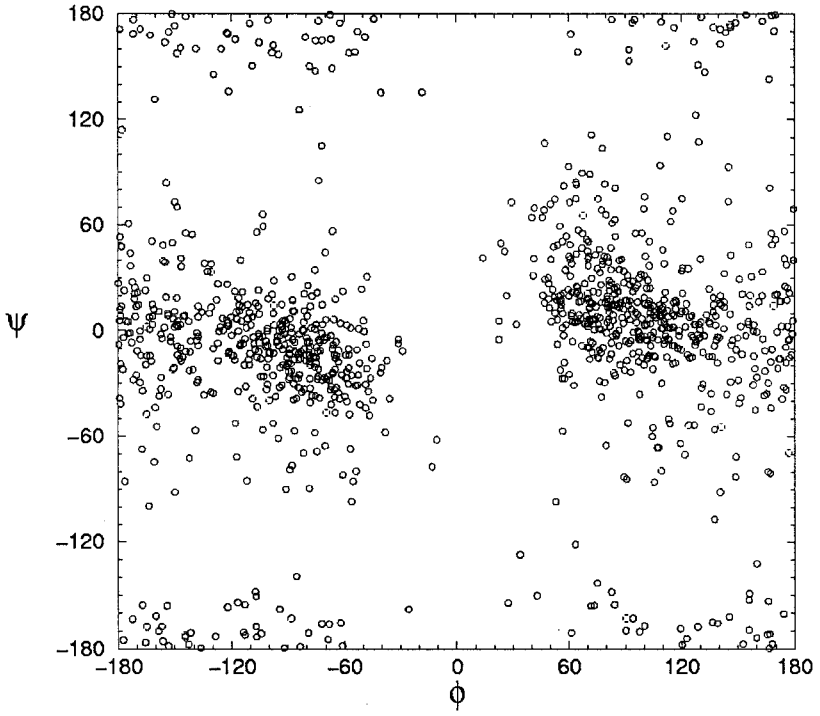

Fig. 8. Sampled conformational distribution of Ace-Gly-Nme with the amber force field (3.1-ns simulation).

highly polar peptide groups form close contacts, with formation of a (distorted) intramolecular hydrogen bond, namely, the $C7_{eq}$ and $C7_{ax}$ conformations, with (ϕ,ψ), respectively, near $(-80,80)$ and $(80,-50)$ (e.g., Ref. 6); these are reasonably well reproduced by a MM model in vacuo.[19] These conformations are unstable relative to other conformations when these are stabilized by hydrogen bonds, either with other polar groups of the polypep-

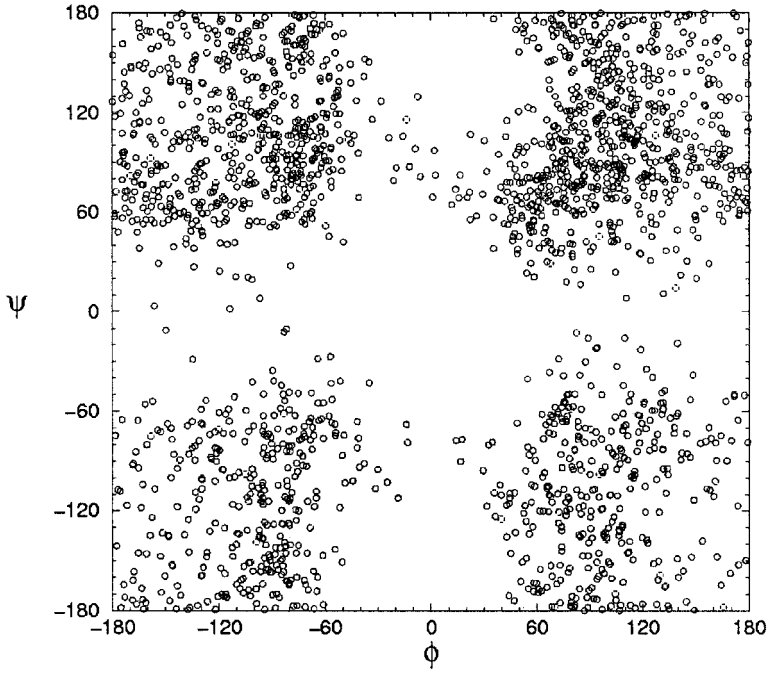

Fig. 9. Sampled conformational distribution of Ace-Gly-Nme with the cedar force field (1.2-ns simulation).

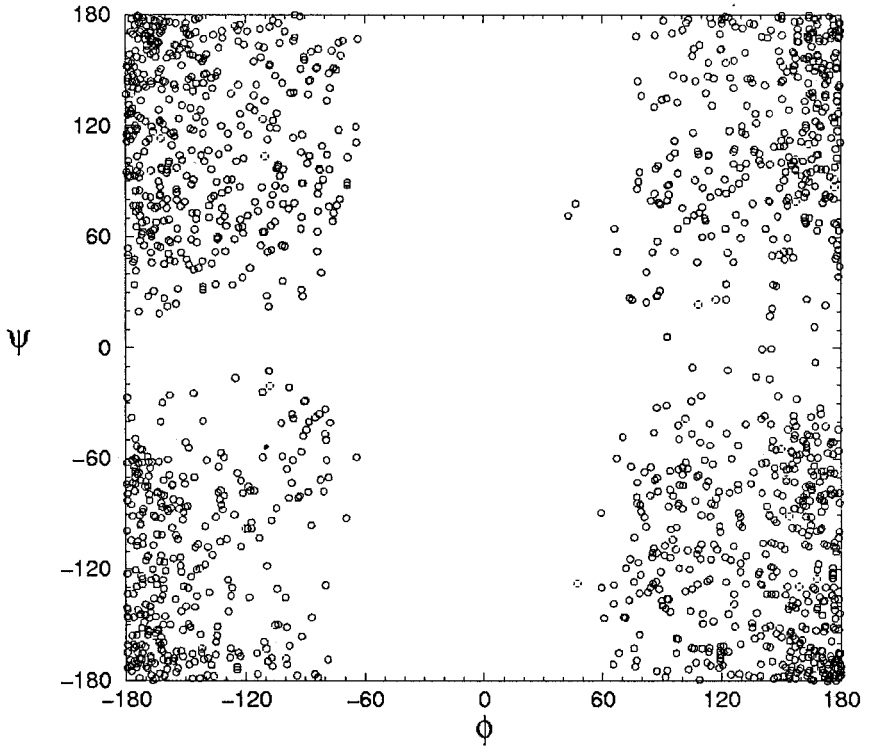

Fig. 10. Sampled conformational distribution of Ace-Gly-Nme with the charmm22 force field (1.1-ns simulation).

tide (often, other peptide groups), or with water, and they are uncommon in folded proteins. Conversely, conformations that are common in proteins have considerably higher energies in the in vacuo model. The most stable conformations of dipeptides in aqueous solution found by simulation with an MM energy function obtained as a fit to a QM energy surface of Ace-Ala-Nme in vacuo, correspond to high-energy conformations of the isolated molecule. It is

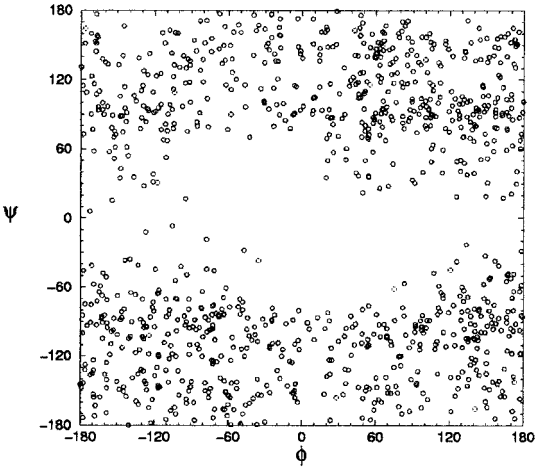

Fig. 11. Sampled conformational distribution of Ace-Gly-Nme with the gromos force field (7-ns simulation).

obviously difficult to design an MM energy function, based on a fit to the QM energy surface of Ace-Ala-Nme in vacuo, that is accurately transferable to the highly polar environment of an aqueous solution or a folded protein. Although a model that is accurate under a wide variety of conditions would appear advantageous, a greater advantage is achieved in practice by having a model that is highly accurate under the conditions of interest.

Importance of Explicit Torsional Energy Terms

The five MM energy functions have the same form but different parameters. Differences that may be at the root of the observed differences between the free energy surfaces in water are: atomic partial charges of the peptide group, repulsive parameters of the Lennard–Jones potential, and torsion parameters. Equilibrium values and force constants of bond lengths and bond angles and attractive parameters of the Lennard–Jones potential do not differ greatly between force fields, and, also, the two water models (SPC and TIP3P) are very similar. With one exception, we have not attempted to correlate the differences in conformational distributions with differences in parameters. The MM force fields include energy terms, U_χ that depend explicitly on the value of the dihedral angle, χ (ϕ or ψ) in the form of one or several Fourier terms,

$$U_\chi = \sum_i \frac{U_{\chi,i}}{2}[1 - \cos\{n_i(\chi - \chi_{0,i})\}] \qquad (1)$$

and the force fields disagree on the details of these terms. Thus, for ψ, cedar has a single threefold term ($n_i = 3$) with a small barrier ($U_{\chi,i} = 0.2$ kcal/mol), whereas opls has a set of three terms with $n_i = 1$, 2, and 3, and amber a set of three terms with $n_i = 1$, 2, and 4. The torsional potential for amber favors values of ψ near $-60°$ over values near 180° by several kcal/mol and appears responsible for the high probability of conformations with ψ near $-60°$ for both Ace-Ala-Nme and Ace-Gly-Nme (Figs. 1 and 8. It would be easy to adjust the constants in U_χ for amber and thereby achieve a different balance between population of conformations. Backbone torsional parameters in amber94[44] and amber99[45] are different from those in amber98.) The gromos potential contains a sixfold energy term for ψ, with $U_{\chi,i} = 0.48$ kcal/mol, the effect of which is discernable in the distribution for Ace-Gly-Nme (Fig. 11).

Quality of the QM/MM Method

The choice of a QM/MM method is a compromise between accuracy and speed. The SCCDFTB method is among the fastest approximate QM methods; nevertheless, SCCDFTB closely reproduces minimum-energy molecular geometries found with high-level methods. The energies of several conformations of Ace-Ala-Nme in vacuo have been found to be properly ranked, which is not the case for several other fast QM methods.[23] We have found the energy surface to be qualitatively very similar to that reported for a high-level QM method, LMP2[6]; the largest deviation is found for the $C7_{ax}$ conformation; the energy of the $C7_{ax}$ conformation lies above that of the $C7_{eq}$ conformation by 1.1 kcal/mol, compared with 2.71 kcal/mol for LMP2. Application of SCCDFTB to helices of short alanine oligomers in vacuum, to crambin and to alanine oligomers in aqueous solution has amply confirmed these findings.[24,25] On the negative side, we mention that SCCDFTB provides a poor representation of liquid water in a dynamics simulation of several hundred water molecules with periodic boundary conditions (results not shown). Furthermore, the conformational distribution of Ace-Ala-Nme obtained here with this QM/MM force field does not show the preference for the P_{II} conformation that has been indicated by experimental studies.[12–16] A previous study investigating structures and relative energies of solvated Ace-Ala-Nme finds that the P_{II} conformation is the global minimum on the potential energy surface when the whole system (solute and solvent) is treated with QM, whereas in a QM/MM description, with solute treated with QM and solvent with MM, the α_R conformation is favored over the P_{II} conformation.[46]

We have seen here that the results obtained with QM/MM are sensitive to the choice of MM parameters, mainly the repulsive Lennard–Jones parameters, the water models, and the attractive Lennard–Jones parameters being similar. It will be appropriate to optimize a set of Lennard–Jones parameters specifically for describing interactions between the SPC or TIP3P water model and solutes represented with SCCDFTB. (These may then be used in combination with a recently developed set of intramolecular long-range potentials.[47]) Development of such an intermolecular parameter set on the basis of free energies of transfer of selected solutes is in progress. Preliminary results indicate that the properties of a single QM H_2O molecule represented with SCCDFTB are very sensitive to the strength of the Lennard–Jones repulsive interaction between it and surrounding SPC water molecules. This finding suggests that optimization of these parameters should be a first step, only after which one may want to consider use of more sophisticated MM water

models, such as models with additional charge centers and polarizable models.

Relevance of Database Statistics

We have used the extent of agreement between distributions of backbone conformations from simulations of Ace-Ala-Nme and Ace-Gly-Nme in water and these same distributions of alanine and glycine residues in proteins as evidence indicating the accuracy of the QM/MM model with SCCDFTB. A direct relation between the two had been proposed by two groups, who suggested also that these distributions (corrected for the prevalence of regular secondary structures) could be used in a statistical description of the conformation of unfolded peptides in solution.[48,49] Later, discrepancies between the database distributions and distributions from simulations with a MM force field, in particular for the glycine residue, were adduced as an argument that the different nature of the two environments (aqueous solution and folded protein) would tend to limit the proposed agreement.[50] The present good agreement suggests that because virtually all polar groups participate in hydrogen bonds in the protein interior, this may be considered a polar environment not very dissimilar to water for polar groups, while at the same time presumably providing a much less polar environment for apolar groups. Although it is not surprising that conformations with high energy due to atomic overlap are absent in folded proteins as well as in the simulated sample, it remains debatable if, and if so, why, the statistics of, for example, the alanine backbone conformation in proteins follow the equilibrium distribution of the alanine residue of peptides such as Ace-Ala-Nme in water.[51,52] Ultimately, the accuracy of simulated distributions of backbone geometry of unfolded peptides in water will presumably be superior to that provided by the database, but at present such is not the case.

Protein Stability

An accurate representation of the unfolded state is required for simulations of conformation change from the unfolded to an ordered folded state, such as helix formation of polypeptides and folding of globular proteins. Accurate free energy differences between folded and unfolded states are known from experiment for many such systems, and it is important to establish the accuracy with which these are reproduced in the simulated system. A problem is that, given the actual rates, transitions from unfolded to folded states are unlikely within the times over which such systems can be simulated in practice. Formation of helices by β-peptides (chains formed by β-amino acids linked by amide bonds) in aqueous solution is exceptionally fast, and this is one system for which simulations (with the gromos force field) have proven to reproduce the observed equilibria quite well.[53–55]

Formation of α-helices by polypeptides is slower because of the cooperative nature of the helix initiation step, although the accretion or loss of helical structure at the ends of α-helical regions occurs on a timescale where the process can presumably be studied with free simulations. Whether the force field accurately represents the equilibrium, depends, in the first instance, on the free energy difference between the unfolded and folded states. Special simulation techniques exist, which allow one to calculate free energy differences; in one example, the free energy of formation of α-helix by short alanine oligomers in solution has been computed for the cedar force field and shown to be in reasonable agreement with experiment.[56] Other simulations directed at assessing stability of elements of secondary structure of proteins include studies of reverse turns, β-sheet and α-helix[57–60] and of designed β-sheet proteins.[61–63] These simulations have been able to describe believable free-energy landscapes for the folding of these molecules, including the marginal stability of the folded state. It has been theorized[64,65] that marginal stability is required for rapid folding of proteins, in which case MM force fields used in simulation of the folding process would of necessity have to represent only *marginal* stability. Inaccuracies in the force field could then easily render the folded conformation unstable relative to the ensemble of all other solution conformations. In this light, the expectation of successful, accurate protein folding in long simulations of an unfolded protein molecule[11] appears naive unless it can first be established that the used MM force field produces a relatively small balance of the free energy in favor of the folded state.

ACKNOWLEDGMENT

Supercomputing time was provided by the North Carolina Supercomputing Center. We thank Drs. Jane and David Richardson and Ian Davis for helpful discussions and providing results of recent database analysis studies in their laboratory. We thank Drs. Julian Tirado-Rives, Yong Duan, and Lee Bartolotti for providing us topology files in amber format and Drs. Weitao Yang and Haiyan Liu for advice on use of SCCDFTB.

REFERENCES

1. Cornell WD, Cieplak P, Bayly C, Gould IR, Merz KMJ, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins and nucleic acids. J Am Chem Soc 1995;117:5179–5197.
2. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem 1998;B 102:3586–3616.
3. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG. Biomolecular simulation: the GROMOS96 manual and user guide. Zürich: Vdf Hochschulverlag AG an der ETH Zürich; 1996.
4. Jorgensen WL. OPLS force fields. In: Schleyer PvR, editor. Encyclopedia of computational chemistry. Vol. 3. New York: Wiley; 1998.
5. Maple JR, Hwang M-J, Jalkanen KJ, Stockfisch TP, Hagler AT. Derivation of class II force fields. V. Quantum force field for amides, peptides, and related compounds. J Comput Chem 1998; 19:430–458.
6. Kaminski G, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA forcefield for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem 2001;B 105:6474–6487.

7. Williams DE. Nonbonded potential parameters derived from crystalline hydrocarbons. J Chem Phys 1967;47:4680–4684.
8. Scheraga HA. Calculations of conformations of polypeptides. Adv Phys Org Chem 1968;6:103–184.
9. Ferro D, Hermans J. Semiempirical energy calculations on model compounds of polypeptides. Crystal structures of DL-acetylleucine N-methylamide and DL-acetyl-n-butyric acid N-methylamide. Biopolymers 1972;11:105–117.
10. Sreerama N, Woody RW. Molecular dynamics simulations of polypeptide conformations in water: a comparison of α, β, and poly(Pro)II conformations. Proteins 1999;36:400–406.
11. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–744.
12. Sreerama N, Woody RW. Poly(Pro) II type structure in globular proteins—identification and CD analysis. Biochemistry 1994;33: 10022–10025.
13. Poon CD, Samulski ET, Weise CF, Weisshaar JC. Do bridging water molecules dictate the structure of a model dipeptide in aqueous solution? J Am Chem Soc 2000;122:5642–5643.
14. Schweitzer-Stenner R, Eker F, Huang Q, Griebenow K. Dihedral angles of trialanine in D_2O determined by combining FTIR and polarized visible Raman spectroscopy. J Am Chem Soc 2001;123: 8628–9633.
15. Woutersen S, Hamm P. Structure determination of trialanine in water using polarization sensitive two-dimensional vibrational spectroscopy. J Phys Chem 2001;B 104:11316–11320.
16. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II structure in a 7-residue alanine peptide. Proc Natl Acad Sci USA 2002;accepted for publication.
17. Anderson AG, Hermans J. Microfolding: conformational probability map for the alanine dipeptide in water from molecular dynamic simulation. Proteins 1988;3:262–265.
18. Smith PE. The alanine dipeptide free energy surface in solution. J Chem Phys 1999;111:5568–5579.
19. Tobias DJ, Brooks CL. Conformational equilibrium in the alanine dipeptide in the gas phase and aqueous solution—a comparison of theoretical results. J Phys Chem 1992;96:3864–3870.
20. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. Intermolecular forces. Dordrecht, Holland: Reidel; 1981. p 331–342.
21. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
22. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G. Self-consistent charge density functional tight-binding method for simulation of complex material properties. Phys Rev B 1998;58:7260–7268.
23. Elstner M, Frauenheim T, Kaxiras E, Seifert G, Suhai S. A self-consistent charge density-functional based tight-binding scheme for large biomolecules. Phys Status Solidi B 2000;217:357–376.
24. Liu H, Elstner M, Kaxiras E, Frauenheim T, Hermans J, Yang W. Quantum mechanics simulation of protein dynamics on long time scale. Proteins 2001;44:484–489.
25. Cui Q, Elstner M, Kaxiras E, Frauenheim T, Karplus M. A QM/MM implementation of the self-consistent-charge density functional tight binding (SCC-DFTB) method. J Phys Chem 2001;B 105:569–585.
26. Elstner M, Jalkanen K, Knapp-Mohammady M, Frauenheim T, Suhai S. DFT studies on helix formation in N-acetyl-$(L\text{-alanyl})_n$-N′-methylamide for n = 1–20. Chem Phys 2000;256:15–27.
27. Elstner M, Jalkanen KJ, Knapp-Mohammady M, Frauenheim T, Suhai S. Energetics and structure of glycine and alanine based model peptides: approximate SCC-DFTB, AM1 and PM3 methods in comparison with DFT, HF and MP2 calculations. Chem Phys 2001;263:203–219.
28. Bohr HG, Jalkanen KJ, Elstner M, Frimand K, Suhai S. A comparative study of MP2, B3LYP, RHF and SCC-DFTB force fields in predicting the vibrational spectra of N-acetyl-(L-alanine)-N′-methyl amide: VA and VCD spectra. Chem Phys 1999;246:13–36.
29. Mann G, Yun RH, Nyland L, Prins J, Board J, Hermans J. The Sigma MD program and a *generic* interface applicable to multifunctional programs with complex, hierarchical command structure. In: Schlick T, Gan HH, editors. Computational methods for macromolecules: challenges and applications. Proceedings of the 3rd International Workshop on Algorithms for Macromolecular Modelling, New York, October 12–14, 2000. Berlin and New York: Springer-Verlag; 2002. Forthcoming.
30. Tuckerman ME, Berne BJ, Martyna GJ. Reversible multiple time scale molecular dynamics. J Chem Phys 1992;97:1990–2001.
31. Darden TA, York DM, Pedersen LG. Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. J Chem Phys 1993;98:10089–10092.
32. Schlick T, Skeel RD, Brünger AT, Kalé LV, Board JA, Hermans J, Schulten K. Algorithmic challenges in computational molecular biophysics. J Comput Phys 1999;151:9–48.
33. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–3690.
34. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–341.
35. Weiner PK, Kollman PA. AMBER: assisted model building with energy refinement. A general program for modeling molecules and their interactions. J Comput Chem 1981;2:287–303.
36. Brunger AT. X-PLOR, a system for X-ray crystallography and NMR. New Haven, CT: Yale University Press; 1992.
37. Ferro DR, McQueen JE, McCown JT, Hermans J. Energy minimization of rubredoxin. J Mol Biol 1980;136:1–18.
38. Hermans J, Berendsen HJC, van Gunsteren WF, Postma JPM. A consistent empirical potential for water-protein interactions. Biopolymers 1984;23:1513–1518.
39. Mooij WTM, van Duijneveldt FB, van Duijneveldt-van de Rijdt JGCM, van Eijck BP. Transferable ab initio molecular potentials.1. Derivation from methanol dimer and trimer calculations. J Phys Chem A 1999;103:9872–9882.
40. Okur A, Strockbine B, Hornak V, Simmerling C. Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. J Comput Chem 2002; in press.
41. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. Proteins 2003;50:437–450.
42. Mu Y, Stock G. Conformational dynamics of trialanine in water: a molecular dynamics study. J Phys Chem 2002;B 106:5294–5301.
43. Price DJ, Brooks CL. Modern protein force fields behave comparably in molecular dynamics simulations. J Comput Chem 2002;23: 1045–1056.
44. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr, Weiner P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984;106: 765–784.
45. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J Comput Chem 2000;21:1049–1074.
46. Han W, Elstner M, Jalkanen KJ, Frauenheim T, Suhai S. Hybrid SCC-DFTB/molecular mechanical studies of H-bonded systems and of N-acetyl-$(L\text{-Ala})_n$-N′-methylamide helices in water solution. Int J Quant Chem 2000;78:459–479.
47. Elstner M, Hobza P, Frauenheim T, Suhai S, Kaxiras E. Hydrogen bonding and stacking interactions of nucleic acid base pairs: a density-functional-theory based treatment. J Chem Phys 2001;114: 5149–5155.
48. Muñoz V, Serrano L. Intrinsic secondary structure propensities of the amino acids, using statistical ϕ-ψ matrices: comparison with experimental scales. Proteins 1994;20:301–311.
49. Swindells MB, MacArthur MW, Thornton JM. Intrinsic ϕμψ propensities of amino acids, derived from the coil regions of known structures. Nat Struct Biol 1995;2:596–603.
50. O'Connell T, Wang L, Tropsha A, Hermans J. Comparison of models from database statistics and molecular simulations. Proteins 1999;36:407–418.
51. Sippl MJ. Boltzmann's principle, knowledge-based mean fields, and protein folding. An approach to the computational determination of protein structures. J Comput Aided Mol Design 1993;7:473–501.
52. Thomas PD, Dill KA. Statistical potentials extracted from protein structures: how accurate are they? J Mol Biol 1996;257:457–469.
53. Daura X, van Gunsteren WF, Rigo D, Jaun B, Seebach D.

Studying the stability of a helical β-heptapeptide by molecular dynamics simulation. Chem Eur J 1997;3:1410–1417.
54. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE. Reversible peptide folding in solution by molecular dynamics simulation. J Mol Biol 1998;280:925–932.
55. Daura X, van Gunsteren WF, Mark AE. Folding-unfolding thermodynamics of a β-heptapeptide from equilibrium simulations. Proteins 1999;34:269–280.
56. Wang L, O'Connell T, Tropsha A, Hermans J. Thermodynamic parameters for the helix-coil transition of oligopeptides: molecular dynamics simulation with the peptide growth method. Proc Natl Acad Sci USA 1995;92:10924–10928.
57. Tobias DJ, Sneddon SF, Brooks CL. Reverse turns in blocked dipeptides are intrinsically unstable in water. J Mol Biol 1990;216: 783–796.
58. Tobias DJ, Mertz JE, Brooks CL. Nanosecond timescale folding dynamics of a pentapeptide in water. Biochemistry 1991;30:6054–6058.
59. Tobias DJ, Brooks CL. Thermodynamics and mechanism of α-helix initiation in alanine and valine peptides. Biochemistry 1991;30: 6059–6070.
60. Tobias DJ, Sneddon SF, Brooks CL. Stability of a model beta-sheet in water. J Mol Biol 1992;227:1244–1252.
61. Bursulaya BD, Brooks CL. Folding free energy surface of a three-stranded β-sheet protein. J Am Chem Soc 1999;121:9947–9951.
62. Ferrara P, Caflisch A. Folding simulations of a three-stranded antiparallel β-sheet peptide. Proc Natl Acad Sci USA 2000;97: 10780–10785.
63. Cavalli A, Ferrara P, Caflisch A. Weak temperature dependence of the free energy surface and folding pathways of structured peptides. Proteins 2002;47:305–314.
64. Klimov DK, Thirumalai D. Factors governing the foldability of proteins. Proteins 1996;26:411–441.
65. Dinner AR, Abkevich V, Shakhnovich EI, Karplus M. Factors that affect folding ability of proteins. Proteins 1999;35:34–40.