Computer Simulation of Met-Enkephalin Using Explicit Atom and United Atom Potentials: Similarities, Differences, and Suggestions for Improvement

Muhammad H. Zaman,[†,‡] Min-Yi Shen,[†,§] R. Stephen Berry,[†,§] and Karl F. Freed*[,†,§]

Department of Chemistry, Institute of Biophysical Dynamics, and The James Franck Institute, The University of Chicago, Chicago, Illinois 60637

Received: September 15, 2002; In Final Form: November 2, 2002

The conformational dynamics of Met-enkephalin are studied using a variety of united atom and explicit atom force fields to compare and contrast the dynamical behavior predicted by these two types of force fields in biological systems. Equilibrium and dynamical properties are computed from 130-ns implicit water, Langevin dynamics (LD) simulations for four explicit atom and two united atom force fields. The dynamical properties discussed provide a reasonable probe of conformational regions not covered by equilibrium folded structures and are very useful in testing and improving force fields. A comparison of these dynamical properties indicates that the explicit atom and united atom force fields produce significantly different peptide dynamical properties.

Introduction

The accuracy, speed, and reliability of computer simulations depend crucially upon the force fields employed. The commonly used "atomistic" force fields can be divided into two broad categories, namely, all-atom (also called explicit atom) and united atom force fields. The explicit atom force fields treat each atom in the molecule as an interaction site,[1] whereas the united atom force fields unite the carbon atoms and their directly bonded hydrogen atoms into single, often spherically symmetric interaction sites.[2] In other words, the explicit atom force fields represent CH_3, CH_2, and CH in terms of four, three, and two interaction sites, respectively, whereas the united atom force fields employ single pseudoatom representations for each CH_3, CH_2, and CH group. The explicit atom force fields appear to be more realistic and are believed to be more appropriate at higher densities. However, the use of united atom representations is quite desirable because the reduction of aliphatic groups to a single pseudoatom can increase the simulation speed by as much as an order of magnitude[3] and can thereby render computationally expensive simulations quite inexpensive and tractable on a personal computer. For example, the ambitious "folding at home" program currently employs the united atom OPLS force field as part of its efforts at simulating the folding rates of small proteins.[4,5]

Equilibrium and dynamic properties of alkanes[3,6−9] and silicone polymers[10] as well as trialanine in water[11,12] have been simulated extensively using both explicit atom and united atom models, and the corresponding structural, equilibrium, and dynamic properties have been compared to experimental data. These studies delineate the regimes where the united atom models agree well with experiments. In contrast, however, very little analysis exists that compares the predictions of these two types of force fields for the conformational dynamics of flexible peptides (i.e., for systems that access a wide range of conformations that depart considerably from those in the neighborhood of the native structures of folded proteins).

As in their application to problems in polymer physics, computer simulations for biopolymers have contributed to the understanding of the equilibrium and dynamical properties of these large and complex systems on short and long time scales. This understanding of the dynamics of these systems is important in identifying potential candidates for better and more effective drug design, for the influence of mutations on different proteins, and for the interaction of proteins with ligands, membranes, and solvents. To perform these simulations with reasonable computational times, united atom models are often used, for example, in studies of protein folding,[13] but no direct comparison has been made of the dynamical properties emerging from simulations using explicit atom and united atom models. Moreover, force fields for proteins have generally been devised by comparison with experimental structural data (perhaps with some ab initio information), but the peptide dynamics samples a wide range of non-native conformations and thereby provides a far more stringent test of the quality of the potential functions. Thus, force fields that predict similar native structures may deviate considerably in their description of the dynamics of very flexible peptides such as the one studied herein.

We examine the long-time dynamical behavior of a pentapeptide, Met-enkephalin, using an implicit solvent method that has been tested by comparison with explicit solvent MD simulations for this same peptide[14] and for the initial stages of folding of the villin headpiece.[15] Shen and Freed have shown that implicit water LD simulations are 200 times faster than explicit solvent MD simulations,[14] therefore making the study of peptide dynamics more tractable on personal computers. The additional speed enhancement from using united atom models scales roughly as the square of the ratio of numbers of united atom groups to explicit atoms, which for Met-enkephalin is a factor of $(75/57)^2 = 1.73$.

Met-enkephalin (Tyr-Gly-Gly-Phe-Met) is one of the smallest neurotransmitter peptides and was first isolated from pig brains.[16] This peptide has been studied extensively using X-ray crystallography,[17] NMR,[18] and computer simulations.[14,19,20] This large body of research has established that similar to other short peptides Met-enkephalin does not exhibit a single native conformation. Rather, Met-enkephalin rapidly traverses a wide

* To whom correspondence should be addressed. E-mail: k-freed@uchicago.edu.
† Department of Chemistry, The University of Chicago.
‡ Institute of Biophysical Dynamics, The University of Chicago.
§ The James Franck Institute, The University of Chicago.

10.1021/jp026994s CCC: $25.00
Published on Web 01/29/2003

range of different conformations in aqueous solution.[14,21] To compare and contrast the predicted dynamical behavior of Met-enkephalin by united atom and explicit atom models, we study the long-time behavior (~130 ns) using six different commonly used force fields. Four of these force fields (AMBER 94,[21] AMBER 96,[22] CHARMM-27,[24,25] and OPLS all-atom[26]) are explicit atom force fields, whereas two of them (OPLS and CHARMM-19[27]) employ the united atom method. Three pairs are matched sets that have been developed by the same group and therefore provide checks on their internal consistency. The main aim of our study is not to show the superiority or inferiority of any of the force fields; rather, it is aimed at suggesting criteria that can be used to test and improve existing force fields.

Computational Details

The implicit water Langevin dynamics (LD) simulations follow the procedures discussed in detail by Shen and Freed.[14] Thus, the method is reviewed only briefly in this section. Within the implicit solvent model, the total system energy is given by

$$U_{\text{tot}} = U_{\text{b}} + U_{\text{bend}} + U_{\text{imp-tors}} + U_{\text{tors}} + U_{\text{ch}}(\epsilon) + U_{\text{vdw}} + U_{\text{solv}} \quad (1)$$

where the subscripts b, bend, imp-tors, tors, and vdw denote the bonding, bond bending, improper torsions, torsions, and van der Waals interaction terms in the overall system energy. The subscript ch designates the contribution involving the dielectric screening of electrostatic interactions, and the subscript solv denotes the solvation potential portion of the overall system energy. The implicit water expression for the overall system energies differs from its explicit water counterpart in the presence of the dielectric screening in $U_{\text{ch}}(\epsilon)$ and in the solvation term (U_{solv}) that replace the Coulomb interactions and the protein−solvent and solvent−solvent interactions, respectively, in the explicit solvent treatments.

We utilize the macroscopic solvation potentials U_{sol} given by the Ooi−Scheraga solvent-accessible surface area (SASA)[28] method because comparisons between implicit and explicit solvent simulations demonstrate the superiority of this potential in more faithfully approximating the results of explicit solvent simulations.[14] The potential contains a contact free energy term that is evaluated in terms of the accessible surface area (σ_i) of all atoms i in the peptide. The SASA accessible surface area is computed from a hypersurface bisecting the first solvent shell using a water (probe) radius of 1.4 Å. Therefore, the overall solvation free energy can be written as a sum of free energy contribution from all atoms,

$$U_{\text{solv}} = \sum_{i=1}^{i=N} g_i \sigma_i \quad (2)$$

where g_i denotes the empirical atom solvation energy parameters[28] determined by fitting experimental aqueous solvation free energies of amino acids and selected organic compounds to eq 2.

The LD simulations are based on nonlinear generalized Langevin equations (GLE) and a procedure similar to MD simulations, apart from the need for an additional algorithm for computing the frictional forces and the corresponding random forces that represent the frictional forces due to implicitly treated water. These friction coefficients are computed by the method of Pastor and Karplus[29] and are updated every 100 integration steps. More explicitly, the LD simulations are generated by integrating the atom positions and velocities by using the standard velocity Verlet algorithm,[30]

$$r_i(t + \Delta t) = r_i(t) + c_{1i} v_i(t) \Delta t + \frac{1}{2} c_{2i} a_i(t) \Delta t^2 + r_{gi} \quad (3)$$

$$v_i(t + \Delta t) = c_{0i} v_i(t) + c_{1i} a_i(t) \Delta t + v_{gi} \quad (4)$$

The coefficients c_{0i}, c_{1i}, and c_{2i} are given by

$$c_{0i} = \exp\left(-\zeta_i \frac{\Delta t}{m_i}\right) \quad (5)$$

$$c_{1i} = \left(-\zeta_i \frac{\Delta t}{m_i}\right)^{-1} (1 - c_{0i}) \quad (6)$$

$$c_{2i} = \left(-\zeta_i \frac{\Delta t}{m_i}\right)^{-1} (1 - c_{1i}) \quad (7)$$

where m_i is the mass of ith atom. r_{gi} and v_{gi} are Gaussian random variables with variances depending upon the friction coefficients in the usual manner.[28] The friction coefficients ζ_i are determined from the solvent-accessible surface area (σ_i') with zero probe radius using stick boundary conditions,

$$\zeta_i = 6\pi\eta r_{\text{eff:}i} \quad (8)$$

where η is the solvent viscosity and $r_{\text{eff:}i}$ is the effective hydrodynamic radius of atom i that is computed from the solvent-accessible area by

$$r_{\text{eff:}i} = \sqrt{\frac{\sigma_i'}{4\pi}} \quad (9)$$

As mentioned above, we compare and contrast the results of two united atom and four all-atom force fields that are present in commercially available packages. The LD simulations have been performed using a modified version (by Shen and Freed[14]) of the TINKER software package[31] for protein molecular dynamics simulations. The parameters of the CHARMM-19 force field were incorporated into the TINKER package using the published parameters by Karplus and co-workers.[27] Both the explicit atom and united atom force fields have been used to generate the 130-ns trajectory of Met-enkephalin. The simulations use a 1.5-fs time step and are run on 1.4-GHz Pentium IV and 1.2-GHz AMD machines.

Results and Discussion

Met-enkephalin does not occupy a unique native state in aqueous solution; rather, its dynamics are highly flexible. Previous studies show that the molecule jumps between extended, semipacked, and packed states[14] (Figure 1 a, b, and c, respectively) that may be classified in terms of the distribution for the square of the radius of gyration, which is defined by

$$R_{\text{g}}^2 = \frac{1}{N} \sum_{i=1}^{i=N} (\mathbf{r}_i - \mathbf{r}_{\text{g}})^2 \quad (10)$$

where $\mathbf{r}_{\text{g}}$ denotes the position of the center of gravity of the molecule and $\mathbf{r}_i$ denotes the position of the ith atom. The extended state is classified as the range of conformations where $R_{\text{g}}^2 > 40$ Å^2, the semipacked state for conformations having R_{g}^2 between 20 Å^2 and 40 Å^2 and the packed state for conformations with R_{g}^2 less than 20 Å^2. The normalized distributions for R_{g}^2 are computed by binning the values of R_{g}^2 into 100 discrete regions with values between 0 and 60 Å^2.

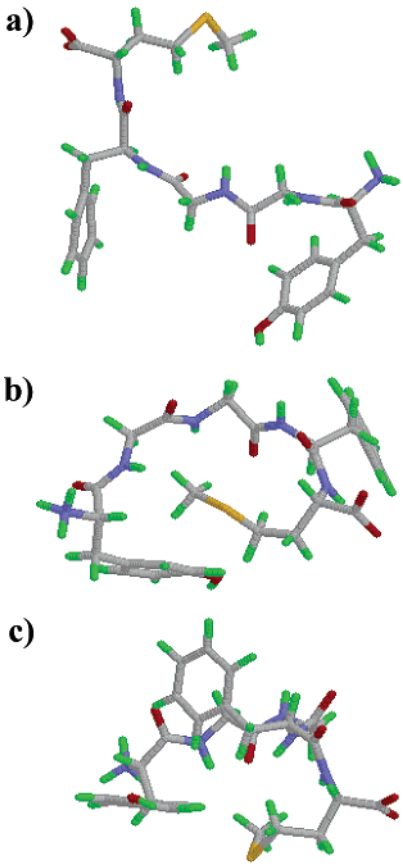

Figure 1. Three classes of conformations accessed during the dynamics of Met-enkephalin. (a) Typical examples of extended ($R_g^2 \approx 42$ Å^2), (b) semicompact ($R_g^2 \approx 33$ Å^2), and (c) compact ($R_g^2 \approx 18$ A^2) conformations. Carbon atoms are shown in gray, nitrogen in blue, oxygen in red, and hydrogen in green.

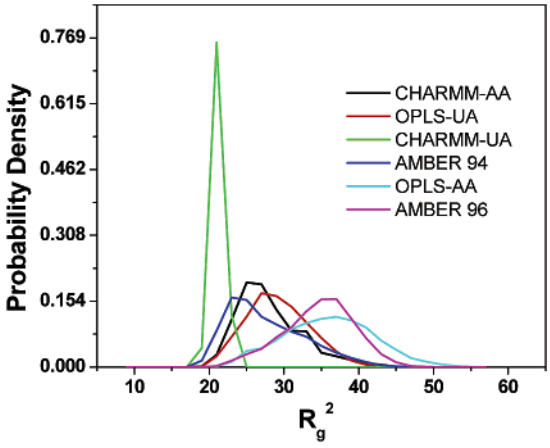

Figure 2. Probability distribution for the radius of gyration R_g^2 for Met-enkephalin as computed with the six different force fields. The radius of gyration is computed using trajectories of 130 ns by employing the implicit solvent LD simulations using the methods of Shen and Freed.[14] The plot displays the R_g^2 distribution for the six force fields discussed in the text.

Figures 2 exhibit the probability distributions of R_g^2 for the six different force fields. The first interesting feature of Figure 2 is the contrast between the explicit atom force fields. Whereas the all-atom AMBER 94 and CHARMM-27 have fairly similar behaviors, the OPLS all-atom force field displays different dynamics in which the Met-enkephalin molecule spends most of its time on the border between semicompact and extended conformations and almost never samples the compact state. AMBER 96 shows a behavior similar to that of the OPLS all-atom force field. In contrast, the CHARMM-27 and AMBER 94 all-atom simulations describe Met-enkephalin as remaining mostly on the border between compact and semicompact states and spending very little time in the extended state. These differences are, however, rather small compared to the departures between the explicit atom and united atom distributions. The LD simulations with the CHARMM-19 united atom force field suggest that the molecule remains in the semicompact state (very close to the compact state) essentially all of the time, in sharp contrast with the other force-field predictions for which Met-enkephalin prefers one state but still significantly samples all of the other conformations. The OPLS united atom distribution departs considerably from the OPLS all-atom distribution but is more similar to the other all-atom distributions for the radius of gyration. The OPLS united atom simulations thus do not overly constrict the peptide to one conformation as is found for the CHARMM-19 united atom simulations, but the OPLS united atom distribution undersamples the compact state.

The comparison of the united atom and explicit atom distributions in Figure 2 raises two important questions, namely, the origin of the differences between the two kinds of force fields and the observation that the CHARMM-19 force field produces a distribution significantly different from those of the other force fields. The answers to both of these questions require a careful analysis of the parameters of the force fields. The primary difference between united atom and explicit atom force fields (apart from the numbers of interaction sites) is the presence of partial charges on aliphatic groups. Both CHARMM-19 and OPLS united atom force fields assign a vanishing partial charge to all CH_3, CH_2, and CH groups that are connected to other aliphatic groups. This difference can then lead to energetics of molecular conformations that significantly differ from the ones that are observed for the explicit atom force fields where these groups have partial charges on the carbon and hydrogen atoms. Consequently, both of the united atom force fields lead to dynamics that sample only parts of the overall conformational space available and, more explicitly, constrict the molecule to only one region of conformational space. The resolution of the second question regarding the differences between the simulations with CHARMM-19 and with the other force fields also requires a careful analysis of the parameters. The CHARMM-19 force field uses a wide variety of "wild-card" parameters; in other words, there are several torsional and improper torsional parameters that have been assigned the same ad hoc value regardless of the types of atoms present. This departs from the procedure used in developing the other force fields, although they too contain a few torsional parameters that are assigned the same ad hoc values. This constancy of torsional parameters may not cause a problem for modeling the dynamics of certain alkanes, but proteins have a highly diverse group of bonds and torsions. A proper description of protein dynamics may require more careful parametrization of the torsions associated with such bonds. The difference between the results computed with the AMBER 94 and AMBER 96 force fields is probably due to modified torsional parameters in AMBER 96, which are based upon empirical data and which have been adjusted to reproduce the energy difference between extended and constrained α-helix energies for the alanine tetrapeptide.[32] The similarity between the results of the AMBER 96 and OPLS all-atom force field simulations has also been observed by other groups.[12] Likewise, the results produced by AMBER 94 and CHARMM-27 have also been found to be quite similar.[12] The two united atom force fields discussed here also differ in their description of the van

der Waals radius of the hydrogen atoms attached to nitrogen or oxygen atoms. Whereas the OPLS united atom force field assigns a value of zero to the van der Waals radius of hydrogen atoms attached to oxygen or nitrogen atoms, CHARMM-19 assigns a nonzero radius. This difference can also lead to divergent results between the two united atom force fields. Similarly, slight differences among the explicit atom force fields also arise because the OPLS all-atom force field specifies a vanishing van der Waals radius for hydrogen atoms bonded to oxygen or nitrogen atoms, whereas AMBER (94 and 96) assigns a nonzero radius. CHARMM-27, however, specifies a nonzero van der Waals radius for hydrogen atoms bonded to either oxygen or nitrogen atoms.

We also compare and contrast a variety of time correlation functions (TCFs) for Met-enkephalin as predicted by the four explicit atom and two united atom force fields. The TCFs compared here are P_1 dipole autocorrelation functions of the interatomic position vectors, which are defined as

$$C_{ij}(t) = \frac{\langle l_{ij}(0) \cdot l_{ij}(t) \rangle}{l_{ij}^2} \tag{11}$$

where the interatom vectors l_{ij} are $l_{ij} = r_i - r_j$. The angular brackets in eq 11 denote the equilibrium average. The P_1 correlation function depicts the local or global flexibility of the molecule depending upon whether atoms i and j are distant or proximate. We compare three TCFs for the united and explicit atom force fields to sample some interesting local and global motions. Explicitly, these TCFs are those for the end-to-end vector, the $C_\gamma - C_\gamma$ vector, and the central backbone C−C vector.

The statistical error in the correlation function $C(t)$ due to the finite trajectory is estimated by the method of Zwanzig and Ailawadi[33] as

$$\sigma = \sqrt{\frac{2\tau'}{T}}[(1 - C(t)] \tag{12}$$

where $T \gg \tau'$ is the duration of the trajectory and τ' is the correlation time defined by

$$\tau' = \int_0^\infty [C(t')]^2 \, dt' \tag{13}$$

Figures 3−5 present several computed TCFs and contain error bars placed at 500, 1000, and 2000 ps with the extremum values $C_\pm(t) = C(t) \pm \sqrt{2\tau'/T}\,[(1 - C(t)]$.

Figure 3 depicts the dipole correlation function for the central backbone C−C bond that is exhibited over the curves. The short-time dynamics produced with the united atom force fields are fairly similar, but this correspondence is lost at longer times where the CHARMM-UA (i.e., CHARMM-19) force field curve decays faster than its OPLS counterpart. The explicit atom C−C TCFs (except for that from AMBER 96) are fairly similar and partially mirror the similar R_g^2 distributions from the explicit atom AMBER 94 and CHARMM-27. A larger difference between the predictions from the OPLS-UA and CHARMM-UA force fields at longer time scales is also obvious from Figure 2 for the R_g^2 from the two force fields. The TCFs for the C−C backbone vector (Figure 3) from the two united atom force fields are more similar at short (∼1 ns) time scales; however, the R_g^2 distributions are more sensitive to the dynamics over much larger (∼120 ns) time ranges. The TCFs for the two united atom force fields differ considerably and probably reflect the differences in the corresponding R_g^2 distributions. Interestingly, the TCFs from both OPLS force fields become more similar for longer times.

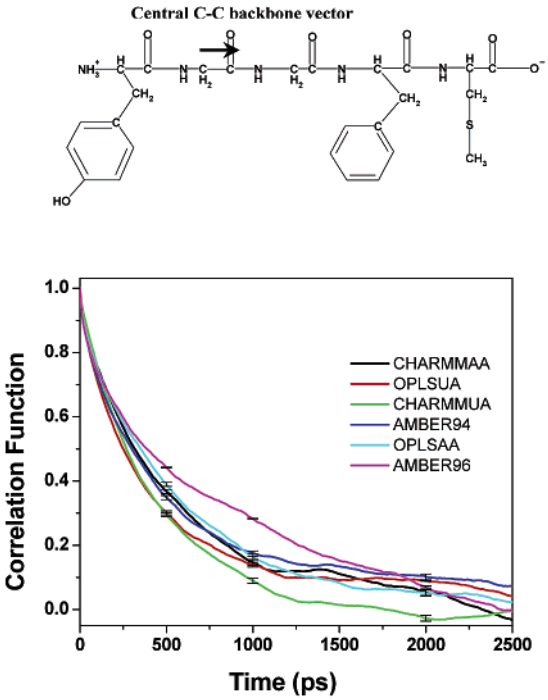

Figure 3. Central backbone C−C bond TCF from different force fields. The backbone vector is shown by an arrow in the structure of Met-enkephalin. The two united atom force fields exhibit similar behavior at very short time scales (<1000 ps), whereas the explicit atom force fields display similar dynamics.

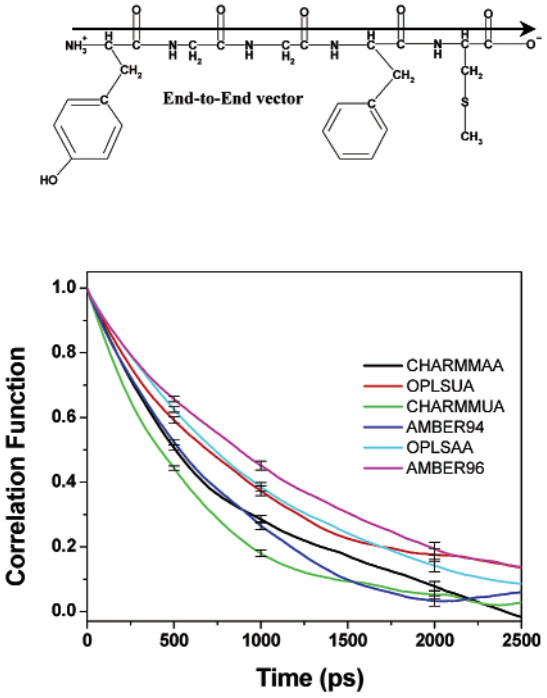

Figure 4. Peptide end-to-end [N(Tyr 1)−O(Met 5)] TCF from different force fields. The dynamics are computed using the two united atom and four explicit atom force fields. The end-to-end vector is depicted by an arrow in the structural diagram for Met-enkephalin. The TCFs from the OPLS-UA and OPLS-AA force fields are similar on this short time scale. The CHARMM-UA force field TCF decays faster than that of any other force field and the AMBER 96 force field TCF decays the slowest whereas the AMBER 94 and CHARMM-AA TCFs are again quite similar.

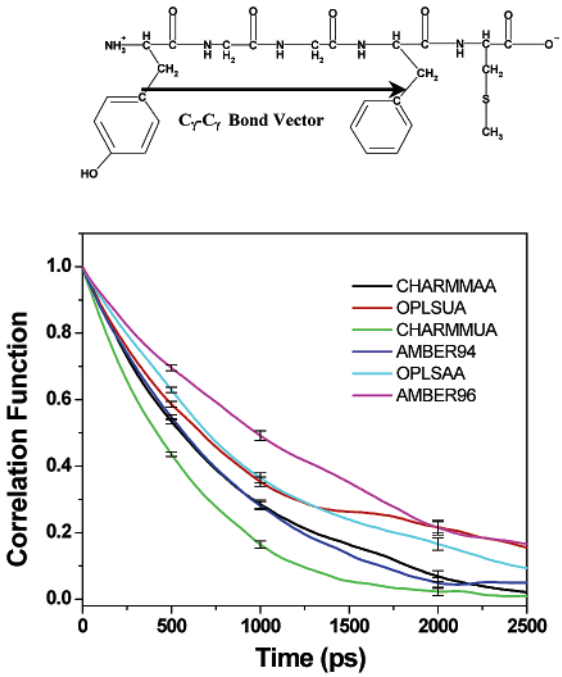

Figure 5. Phe−phe ($C_\gamma-C_\gamma$) vector. This vector represents another slowly varying variable. The TCF from the CHARMM-UA force field decays the fastest, and the AMBER 94 and CHARMM-AA force field TCFs again are very similar, whereas at shorter times (∼1200 ps) the OPLS-AA and OPLS-UA TCFs are close to each other.

Figure 4 presents the TCFs for the dynamics of the end-to-end (N(Tyr1) to O(Met5)) vector. (The TCF for the backbone end-to-end ($C_{1\alpha}-C_{5\alpha}$) vector behaves very much the same.) The OPLS-UA and OPLS-AA TCFs have quite similar shapes but yield very different R_g^2 distributions. Once again, this difference probably arises from the difference in time scales that are relevant to the two sets of properties. The OPLS-UA and OPLS-AA TCFs agree better at shorter time scales but begin to depart for longer times. The AMBER 94 and CHARMM-AA (CHARMM-27) force fields again produce fairly similar TCFs, but the dynamics from these two force fields differ considerably from those calculated using the OPLS-AA force field. Once again, the AMBER 96 force field correlation function decays the slowest.

Figure 5 displays the TCFs of the phenyl−phenyl $C_\gamma-C_\gamma$ vector. Once again, the CHARMM-UA TCF decays the fastest, the AMBER 96 decays the slowest, and the TCFs from the OPLS-UA and OPLS-AA force fields are quite similar, as are the AMBER 94 and CHARMM-AA TCFs.

Further analysis of the TCFs reveals several important features about the nature of the force fields. The OPLS-UA and OPLS-AA force fields yield similar TCFs at shorter time scales for the motions that vary more slowly (such as the end-to-end vectors and the phe−phe $C_\gamma-C_\gamma$ vector), whereas they depart considerably at shorter time scales for relatively rapid motions (such as the backbone C−C vector). The CHARMM-UA yields TCFs that depart substantially from those from all of the other force fields because of the reasons discussed above. The TCFs computed from the AMBER 94 and the CHARMM-AA force fields exhibit the most consistent agreement over all time scales and reflect the similarities in their R_g^2 distributions. Although their shapes differ slightly, both sample almost equally from the three sections of the peptide conformational space. The differences between the distributions of R_g^2 from the OPLS-AA force field, on one hand, and from AMBER 94 and CHARMM-27, on the other hand, are also reflected in their respective TCF plots. Except for the C−C backbone dynamics, the TCFs from OPLS-AA deviate from those computed with the other three explicit atom force fields. The similarities between the OPLS-AA and OPLS-UA force field TCFs at shorter time scales are lost in the longer-time behavior of the peptide. The AMBER 96 force field shows R_g^2 distributions that are similar to those of OPLS-AA but depart considerably from those of all of the force fields for all of the three TCFs. Once again, we believe the similarity in the results between AMBER 96 and OPLS-AA becomes more pronounced at longer time scales, whereas for shorter time scales the two force fields produce significantly different results. The R_g^2 distributions and the TCFs demonstrate that the explicit and united atom force fields predict quite disparate dynamics, probably because of the absence of partial charges on aliphatic groups and because all of these force fields have been optimized for equilibrium structures and not for large-scale conformational dynamics.

Conformational Dynamics. We have also studied the time-dependent conformational dynamics of Met-enkephalin. The study is carried out by inspecting Ramachandran maps for the various torsional angles (Φ, Ψ) of the peptide (Figure 6a). The Ramachandran map that is shown is similar to the one used by Pappu et al.[35] The map is divided into eight basins as indicated by the labels. Figure 6b−g displays the time-dependent dynamics for the central glycine residue (Gly-3) of the peptide. The other residues in the peptide exhibit very similar patterns of behavior as functions of the force field, so Gly-3 serves to illustrate the general trends. Figure 6 depicts the Ramachandran basin occupied by Gly-3 as a function of time. Table 1 summarizes the same information by presenting the time-averaged occupancy of the eight basins for Gly-3. The Figures and Table demonstrate that, apart from the CHARMM-19 force field simulations, the Gly-3 residue executes many hops between basins for all of the other force fields. The CHARMM-27, AMBER-94, and CHARMM-19 trajectories show the peptide spending maximal time in basin 2, though for CHARMM-27, basin 1 is almost equally populated. The AMBER-94 and CHARMM-27 cases exhibit very similar basin populations for all basins except 1 and 6. The OPLS-UA dynamics behaves very similar to that of CHARMM-27 apart from a much smaller presence in basin 2. The dynamical pattern produced by the AMBER-94 trajectory is strikingly different from that computed with AMBER-96, probably because of the different torsional parameters as discussed above. The AMBER-96 trajectory, in sharp contrast to that of AMBER 94, shows very little preference for any basins other than 1, 2, and 3. The Ramachandran

TABLE 1: Percentage Occurrences of Gly-3 Individual Ramachandran Basins

	basin 1	basin 2	basin 3	basin 4	basin 5	basin 6	basin 7	basin 8
CHARMM- AA (27)	29.0	30.0	12.0	0.0	17.0	2.0	10.0	0.0
OPLS-UA	47.0	14.0	5.0	1.0	10.0	3.0	14.0	5.0
CHARMM-UA (19)	1.0	95.0	0.0	0.0	4.0	0.0	0.0	0.0
AMBER-94	3.0	43.0	9.0	1.0	4.0	23.0	15.0	2.0
OPLS-AA	51.0	1.0	4.0	8.0	12.0	0.0	22.0	2.0
AMBER-96	47.0	17.0	24.0	2.0	2.0	0.0	6.0	0.0

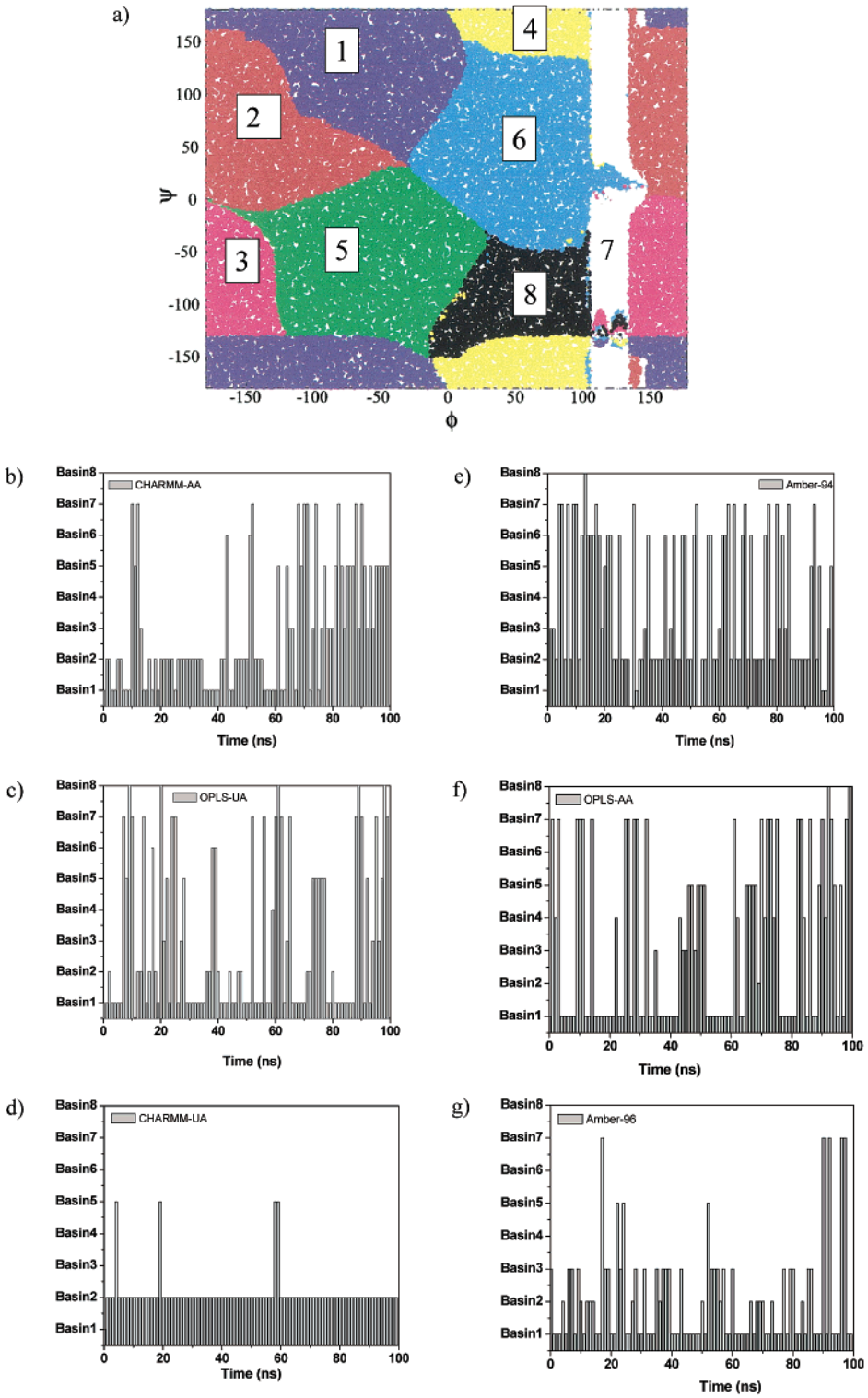

Figure 6. Time-dependent conformational dynamics of the central Gly-3 residue: (a) The Ramachandran plot of torsional angles is divided into eight basins labeled 1 through 8. The plot is similar to the one reported by Pappu et al.[35] and is reproduced with the author's permission. Basin occupations of Gly-3 are shown as functions of time for (b) CHARMM-27 (explicit atom), (c) OPLS-UA, (d) CHARMM-19 (united atom), (e) AMBER-94, (f) OPLS-AA, and (g) AMBER-96.

populations from the OPLS-UA and OPLS-AA force fields exhibit similar behaviors just as observed for the TCFs. The OPLS-AA basin populations display a preference for basins 1, 5, and 7, and the central Gly residue almost never samples basin 2, which is the predominant basin for both the AMBER-94 and CHARMM-27 trajectories. Our conformational dynamics results for AMBER 94 and AMBER 96 show good agreement with Sanbonmatsu et al.'s[36] results using explicit solvent replica−exchange MD simulations on Met-enkephalin using PARM 94 and PARM 96.

The Ramachandran basin populations indicate similar dynamics from the AMBER-94, CHARMM-27, and OPLS-UA force fields and sharp differences between AMBER-96 and AMBER-94, in accord with our observations from the R_g^2 and TCF plots. The conformational dynamics exhibit common characteristics for the OPLS-UA and OPLS-AA force fields, as is also evident

from TCF plots as well. Hence, our analysis demonstrates that a careful study of the conformational dynamics of small, very flexible peptides should provide additional information for improving the representation of current force fields in certain areas of conformational space that are not well sampled by folded protein structures. Further studies of additional dynamical properties that probe the dynamics at different time scales should provide a better overall picture of the regimes where the force fields need further improvement.

Conclusions

We present the first comparison of its kind for the dynamical properties of peptides that are predicted by united atom and explicit atom force fields. Our study considers six commonly used force fields and subjects them to rigorous tests for the dynamics of a short and highly flexible peptide. The united atom and explicit atom models exhibit different dynamical descriptions for the motion of the pentapeptide, Met-enkephalin. The all-atom CHARMM-27 and AMBER 94 force fields produce rather similar results, whereas the OPLS explicit atom force field (OPLS-AA) produces slightly different dynamics that are similar to those found when using the AMBER 96 force field. There are, however, significant differences in the results produced by AMBER-94 and AMBER-96.

The united atom CHARMM-19 force field suffers from limitations in not having partial charges on certain united atom groups and in using ad hoc "wild card" parameters for certain torsions, and these deficiencies are manifest in the simulations that constrict the peptide to remain close to a given conformation. These problems are not experienced by the explicit atom force fields whose simulations display a preference for one conformation but indeed sample the whole range of conformations. The OPLS-UA case is somewhat similar to that for the explicit atom force fields in this regard.

These comparisons are meant to identify problems and develop methods for improving the current force fields that are used for studying biological molecules since both the united atom and explicit atom force fields have already shown great promise in treating alkanes.[34] Our approach represents a point of departure for further comparisons between the use of united atom and explicit atom force fields for describing the dynamics of biological molecules. With the increasing power of MD simulations for biological systems, tests such as ours will be helpful in achieving a more accurate representation of both the short- and long-scale dynamics of proteins. However, further checks for larger and structurally different peptides will be extremely useful in overcoming the shortcomings of these force fields and thereby in developing faster, more accurate, and realistic potentials. After the completion of this work, we received a related preprint[12] that contains a comparison of different explicit solvent force fields for the dynamics of trialanine using a set of properties that differ from those considered here.

Acknowledgment. We thank Gary Grest, Florian Müller-Plathe, Doros Theodorou, Tobin Sosnick, Ariel Fernandez, Andres Colubri, and members of our groups for their comments and suggestions. M.H.Z. thanks Rohit Pappu for his generosity in allowing us to use the Ramachandran plot from his paper. We also thank our reviewers for their useful suggestions. M.H.Z. thanks Peter Rossky for his hospitality and for the use of the computer facilities in his group. M.H.Z. is a Burroughs-Wellcome Interfaces Fellow. This research is supported, in part, by NIH grant GM56678.

References and Notes

(1) Williams, D. E. *J. Chem. Phys.* **1967**, *47*, 4680.

(2) Ryckaert, J. P.; Bellemans, A. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 95.

(3) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1998**, *102*, 2569.

(4) Zagrovic, B.; Sorin, E. J.; Pande, V. S. *J. Mol. Biol.* **2001**, *313*, 151.

(5) Shirts, M. R.; Pande, V. S. *Phys. Rev. Lett.* **2001**, *86*, 4983.

(6) Yoon, D. Y.; Smith, G. D.; Matsuda, T. *J. Chem. Phys.* **1993**, *98*, 10037.

(7) Smith, G. D.; Yoon, D. Y. *J. Chem. Phys.* **1994**, *100*, 649.

(8) Martin, M. G.; Siepmann, J. I. *J. Phys. Chem. B* **1999**, *103*, 4508.

(9) Karayiannis, N. C.; Mavrantzas, V. G.; Theodorou, D. N. *Phys. Rev. Lett.* **2002**, *88*, 105503-1.

(10) Sides, S. W.; Curro, J.; Grest, G. S.; Stevens, M. J.; Soddenmann, T.; Habenschuss, A.; Londono, J. D. *Macromolecules* **2002**, *35*, 6455.

(11) Mu, Y.; Stock, G. *J. Phys. Chem. B* **2002**, *106*, 5294.

(12) Mu, Y.; Kosov, D. S.; Stock, G. *J. Phys. Chem. B*, submitted for publication, 2002.

(13) Bryant Z.; Pande, V. S.; Rokshar, D. S. *Biophys. J.* **2000**, *78*, 584.

(14) Shen, M. Y.; Freed, K. F. *Biophys. J.* **2002**, *82*, 1791.

(15) Shen, M. Y.; Freed, K. F. *Proteins: Struct., Funct., Genet.* **2002**, *49*, 439.

(16) Hughes, J.; Smith, T. W.; Kosterlitz, H. W.; Fothergill, L. A.; Moran, B. A.; Morris, H. R. *Nature (London)* **1975**, *258*, 577.

(17) Smith, G. D.; Griffin, J. F. *Science (Washington, D.C.)* **1978**, *199*, 1214.

(18) Roques, B. P.; Garbay-Jaureguiberry, C.; Oberlin, R.; Anteunis, M.; Lala, A. K. *Nature (London)* **1976**, *262*, 778.

(19) Deber, C. M.; Behnam, B. A. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 61.

(20) Wang, Y.; Kuczera, K. *J. Phys. Chem.* **1996**, *100*, 2555.

(21) Graham, W. H.; Carter, E. S.; Hicks, R. P. *Biopolymers* **1992**, *32*, 1755.

(22) Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. In *Computer Simulations of Biomolecular System* 3; van Gunstern, W. F., Wiener, P. K., Wilson, A. J., Eds.; Kluwer Academic: Dordrecht, The Netherlands, 1997; p 83.

(23) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230.

(24) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, J.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(25) MacKerrell, A. D.; Bashford, D.; Bellot, M.; Dunbrack, R. L.; Evvanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, K.; Kuczera, F.; Lau, T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W, E.; Rounx, 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.

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

(27) Neria, E.; Fischer, S.; Karplus, M. J. *J. Chem. Phys.* **1996**, *105*, 1902.

(28) Ooi, T.; Oobatake, G.; Nemethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086.

(29) Pastor, R. W.; Karplus, M. *J. Phys. Chem.* **1988**, *92*, 2636.

(30) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U.K., 1987.

(31) Ponder, J. W.; Rubenstein, S.; Kundrot, C.; Huston, S.; Dudek, M.; Kong, Y.; Hart, R.; Hodson, M.; Pappu, R.; Mooiji, W.; Loeffler, G. *TINKER: Software Tools for Molecular Design*, version 3.7; Washington University: St. Louis, MO, 1999.

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

(33) Zwanzig, R.; Aliwadi, N. K. *Phys. Rev.* **1969**, *182*, 280.

(34) Pütz, M.; Curro, J. G.; Grest, G. S. *J. Chem. Phys.* **2001**, *114*, 2847.

(35) Pappu, R. V.; Rose, G, D. *Protein Sci.* **2002**, *11*, 2437.

(36) Sanbonmatsu, K. Y.; Garcia, A. E. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 225.