Validation of an All-Atom Protein Force Field: From Dipeptides to Larger Peptides

S. Gnanakaran and Angel E. Garcia*

Theoretical Biology and Biophysics Group, T10, MS K710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received: July 2, 2003; In Final Form: September 30, 2003

New experimental techniques are capable of determining the relative population of conformations adopted by short alanine peptides in water. Most of the existing all-atom force fields used to model proteins fail to reproduce the relative population of the most relevant conformations of peptides. The calculated relative population of conformations varies significantly depending on the force field chosen, thus urging the need to check the validity and consistency of force fields over a range of peptide lengths. Here, we show how the applicability of a modified version of AMBER force field (A94/MOD) can extend from short to large peptides. It is also capable of reproducing the expected shift in conformational preference with increasing peptide length and temperature. Importantly, the consistency of the force field is judged by direct comparison to experiments rather than to the relative energies of conformations obtained from ab initio calculations. Importantly, this study illustrates that many aspects of protein force fields are already well refined and may only require minor refinements to accurately reproduce experimental observations over a range of systems.

The recent availability of two-dimensional infrared spectroscopy[1,2] (2D-IR) and polarized-Raman (PR)/FTIR[3] measurements on trialanine ($(Ala)_3$) and the NMR and CD characterization of alanine (Ala) oligopeptides[4,5] in aqueous solvent have prompted us to evaluate the extendibility of a force field over a range of peptide lengths. On the basis of 2D-IR measurements[2] it was suggested that aqueous $(Ala)_3$ primarily adopts a conformation with the $\phi-\psi$ dihedral angle of $(-60, +140)$, which lies in the poly proline like (PP_{II}) region. Subsequent analysis confirmed that $(Ala)_3$ mainly adopts PP_{II} conformation ($\geq 80\%$) and a small amount of helical (α) conformation ($\leq 20\%$).[6] A different set of measurements based on PR/FTIR[3] found that their observations can be explained by a single extended β-like structure, $(\phi, \psi) = (-123, +173)$, or by simultaneous coexistence of both PP_{II} and β.[7] This supports the existence of additional conformations, such as β, in small quantities. Published results showed that $(Ala)_{13}$ is helical at room temperature (T),[4] whereas $(Ala)_7$ adopts predominantly a PP_{II} structure at low T.[5] The content of β-strand slightly increases at high T.[5] The main question is, can a force field show consistency with the above experimental results for short as well as long lengths of peptides?

Computational studies of the conformational dynamics of peptides have shown a trend with current force fields that is very troubling. The all-atom protein force fields that predict reasonable conformational dynamics for larger peptides or proteins, fail to reproduce the measured conformational distribution for di- and tripeptide systems and vice versa. CHARMM22,[8] AMBER94,[9] and AMBER96[10] force fields have such defects. Even though many of the force fields gave rise to three main conformations for short peptides, the relative population varied significantly.[11,12] The CHARMM22 force field has a strong preference for α conformations for di- and tripeptides.[11−13] Whereas 2D-IR and NMR measurements conclusively show that these peptides adopt primarily PP_{II} conformations.[11] AMBER94 favors helical and AMBER96 favors extended structures for both small and large peptides. However, AMBER96 inadequately models the thermodynamics of helical peptides[14] as well as folding of turn structures.[15] OPLS/AA[16] does not separate the PP_{II} and β structures basins in the Ramachandran map.[11,12] GROMOS96[17,18] populates about an equal amount of PP_{II} and β and a smaller amount of α conformation.[11] Even though the total extended population (80%) agrees with experiments, they clearly oversample the β region compared to PP_{II}.

A simple modification in the AMBER force field reproduces the above experimental observations on tripeptide systems and as well as on larger peptides in explicit aqueous solution. The minor modification of the force field, motivated by simulations of the helix−coil transition of α-helical peptides, eliminates the bias for the backbone dihedral terms in AMBER94.[14] In the A94/MOD force field the backbone dihedral angle energy terms for rotations around the ϕ and ψ angles are set to zero. Garcia and Sanbonmatsu[14] obtained this force field by noticing that the difference between AMBER94 and AMBER96 was only in this term. AMBER96 modified the older AMBER94 force field to reduce the large propensity of α-helix formation. However, AMBER96 favored extended conformations. To resolve this problem, Garcia and Sanbonmatsu performed free energy perturbation calculations in which a bare dihedral angle energy term (i.e., energy set to zero) was used as a reference state.[14] However, simulations of the bare potential were consistent with experiments and no further perturbations were needed. With the A94/MOD the midpoint of the helix−coil transition, the Zimm−Bragg helix nucleation, and propagation parameters for Ala based peptides were in good agreement with experiments.[14] Previous simulation studies that compare force fields to the experimental results on di-/tripeptides did not include this modified force field.[11,12] Because the validity of a comparative study based on peptide lengths depends heavily on achieving reasonable conformational sampling, we have used replica exchange molecular dynamics (REMD) over a wide range of temperatures.

* Corresponding author. E-mail: axg@lanl.gov.

10.1021/jp0359079 CCC: $25.00
Published on Web 10/29/2003

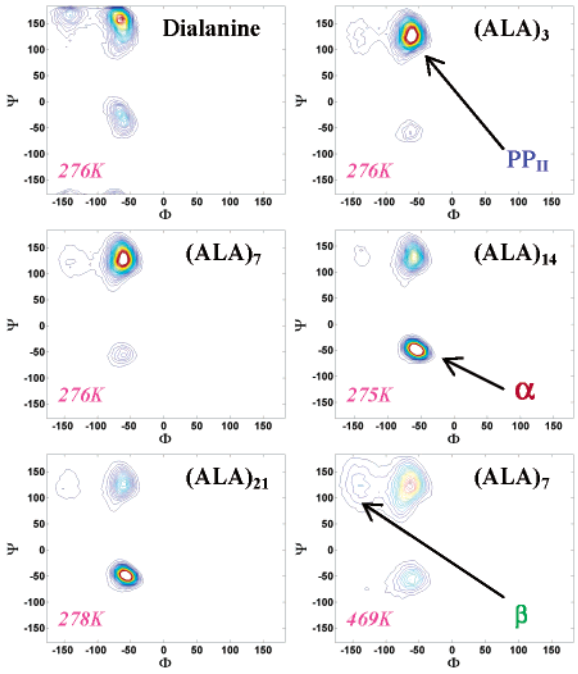

Figure 1. Conformational map of backbone dihedrals of different lengths of Ala peptides at specified temperatures. The log of the probability is plotted. All plots are on the same scale and the three main conformations sampled are marked.

We carried out REMD simulations on the following capped peptide systems in explicit TIP3P water; dialanine (Ac-L-Ala-NHMe) in 316 H_2O, $(Ala)_3$ in 523 H_2O, $(Ala)_7$ in 851 H_2O, $(Ala)_{14}$ in 1823 H_2O, and $(Ala)_{21}$ in 2640 H_2O. The cubic box lengths were chosen to reproduce correct water density at 300 K. For dialanine, $(Ala)_3$, and $(Ala)_7$, the REMD was implemented with 24 replicas with a temperature range of 276−469 K. Each replica was simulated for 10 ns, yielding a total sampling time of 240 ns. The production run was considered to be the last 6 ns/replica and error analysis was done with 1.5 ns/block averaging. The $(Ala)_{14}$ and $(Ala)_{21}$ systems were implemented with 42 (275−551 K) and 46 (278−487 K) replicas, respectively. Each replica was simulated for 8 ns. The last 4 ns were considered as the production (1 ns/block for error analysis). All other simulation details and conditions are same as previously described.[14]

The conformational (ϕ, ψ) maps for the peptides dialanine, trialanine, $(Ala)_7$, $(Ala)_{14}$, and $(Ala)_{21}$ obtained with the modified force field A94/MOD are shown in Figure 1. The three main conformational regions, α, β, and PP_{II}, are populated for all three peptides. For short peptides (≤7), the PP_{II} region is favored. For larger peptides, the helical conformations are preferred over PP_{II} and β. However, as T increases, the preference shifts toward extended β-structure. The preference for PP_{II} conformation increases slightly from dialanine to trialanine, implying water structure around the left-handed helix as a possible mechanism for PP_{II} stabilization.[19]

The population analysis for dialanine and trialanine is carried out to determine how well they reproduce the correct conformational distribution of protonated $(Ala)_3$ as obtained in 2D-IR. In Figure 2 the population probability is given as a function of T for the three well-known conformations. Only the middle dihedral angle is considered for $(Ala)_3$ in the analysis, consistently with the experiment. For $(Ala)_3$, the PP_{II} population of 85 ± 3% is very close to the observed value (≥80%) and the average (ϕ, ψ) lies at the observed value (−60, +140).[2] Small

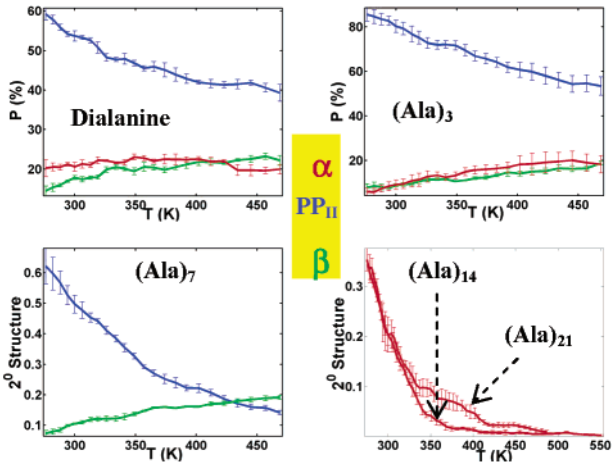

Figure 2. Percent probability of population of different conformations of dialanine and trialanine as a function of T in the top plots. The temporal profile of secondary structure fractional content of Ala oligopeptides is given in the bottom. The color coding for the secondary structural conformations; red (α), blue (PP_{II}), and green (β).

amounts of α-helical and extended β structure (<8%) are found, as predicted.[6,7]

The 2D-IR measurements are sensitive to the coupling between amide I transitions of the peptide units.[20] The first and second moments of coupling between amide I modes led to the deduction of orientations between peptide units in terms of (ϕ, ψ) dihedral angles. Even though the measurements were carried out on $(Ala)_3$, the system is comparable to the interaction of two peptide units with a common (ϕ, ψ) because the amide I transition from the unblocked terminal group is well separated. Therefore, it is imperative that we consider the conformational distribution of dialanine, which also has only two peptide units with a single set of (ϕ, ψ). The population analysis on dialanine also predominantly showed PP_{II} conformation (59 ± 2%) with small contributions from α (20 ± 2%) and β (14 ± 1%) conformations. The mean ψ of PP_{II}, however, is slightly higher than for $(Ala)_3$. The calculated average ψ is closer to the value obtained from NMR studies of dialanine in a water-based liquid crystal.[21]

The Ala peptides of length 5−10 have served as model systems to study the nature of "random-coil" or unfolded state. Here we considered $(Ala)_7$ to seek out any preference for an overall ordered structure. The A94/MOD produced predominantly PP_{II} conformation and a negligible α-helical population. Figure 2 shows the temperature profile of PP_{II}−coil transition. The fractional PP_{II} content is about 62 ± 6% at low T and decays gradually to 14 ± 1% at high T. These results agree with the measurements on $(Ala)_7$ flanked by polar groups at both ends that showed that PP_{II} is the dominant conformation at 2 °C.[5] The same measurements also a found a shift toward β structure at high T, an increases on the order of 10%. A similar increase is obtained in the simulation (7 ± 1% to 19 ± 1%), as shown in Figure 2.

Longer sequences of alanine peptides have been investigated for the α helical content. Experimental measurements on $(Ala)_{13}$ flanked by two ornithine charged amino acids found an α helical content of 40% at 273 K.[4] The fractional α helical content calculated for $(Ala)_{21}$ and $(Ala)_{14}$ in aqueous solution with A94/MOD force field are shown in Figure 2. The $(Ala)_{14}$ has an α helical content of 35 ± 2% at 275 K and $(Ala)_{21}$ has 34 ± 2% at 278 K, which are comparable to the observed value. The A94/MOD force field also reproduced the transition midpoint temperature of the Fs peptide, one of commonly studied Ala

based helical peptide.[14] The PP_{II} conformations dominate in shorter fragments because (i) the $\phi-\psi$ map of a single peptide in solution favors PP_{II}, (ii) the nucleation energy to form PP_{II} is smaller than the corresponding energy for α-helices, and (iii) the propagation of PPII structure is helped by the formation of a delocalized water channel around the backbone, which favors the PP_{II} conformations.[19] However, for longer fragments the α-helix dominates because the propagation term, dominated by the cooperative hydrogen bonding in the α-helix conformation, is strong and larger than the solvent-mediated propagation free energy of the PP_{II} structure. [19]

In summary, we were able to show how the applicability of a single force field (A94/MOD) can extend from short to large peptides and show agreement with measurements. Above all, the expected shift in conformational preference is reproduced with increasing peptide length and temperature. The consistency of the force field is judged by direct comparison to experiments rather than to the relative energies of conformations obtained from ab initio calculations. A94/MOD is an example of how a minor correction in the backbone dihedral can significantly affect the conformational distribution of short peptides in aqueous solution. It illustrates that most aspects of a force field are already reasonably well refined. Therefore, achieving universality of their application may depend on minor refinement of a few terms.

Acknowledgment. This work was funded by the US DOE and LANL LDRD funds. We thank Kevin Y. Sanbonmatsu and Hugh Nymeyer for comments and suggestions.

References and Notes

(1) Woutersen, S.; Hamm, P. *J. Phys. Chem. B* **2000**, *104*, 11316−11320.

(2) Woutersen, S.; Mu, Y.; Stock, G.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11254−11258.

(3) Schweitzer-Stenner, R.; Eker, F.; Huang, Q.; Griebenow, K. *J. Am. Chem. Soc.* **2001**, *123*, 9628−9633.

(4) Spek, E.; Olson, C.; Shi, Z.; Kallenbach, N. *J. Am. Chem. Soc.* **1999**, *121*, 1, 5571−5572.

(5) Shi, Z.; Olson, C.; Rose, G.; Baldwin, R.; Kallenbach, N. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9190−9195.

(6) Woutersen, S.; Pfister, R.; Hamm, P.; Mu, Y.; Kosov, D.; Stock, G. *J. Chem. Phys.* **2002**, *117*, 6833−6840.

(7) Eker, F.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R. *J. Am. Chem. Soc.* **2002**, *124*, 14330−14341.

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

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

(10) Kollman, P. A.; Dixon, R. W.; Cornell, W. D.; Fox, T.; Chipot, C.; Pohorille, A. In *Computer simulations of biological systems*; Gunsteren, W. v., Ed.; ESCOM: Dordrecht, Netherlands, 1997.

(11) Mu, Y.; Kosov, D. S.; Stock, G. *J. Phys. Chem. B* **2003**, *107*, 5064−5073.

(12) Hu, H.; Elstner, M.; Hermans, J. *Proteins* **2003**, *50*, 451−463.

(13) Smith, P. *J. Chem. Phys.* **1999**, *111*, 5568−5579.

(14) Garcia, A.; Sanbonmatsu, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2782−2787.

(15) Gnanakaran, S.; Garcia, A. E. *Biophys. J.* **2003**, *84*, 1548−1562.

(16) Jorgensen, W.; Maxwell, D.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(17) Schuler, L.; Daura, X.; Van Gunsteren, W. *J. Comput. Chem.* **2001**, *22*, 1205−1218.

(18) Van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*; Vdf Hochschulverlag AG an der ETH Zurich: Zurich, 1996.

(19) Garcia, A. E. *Polymer*, in press.

(20) Gnanakaran, S.; Hochstrasser, R. *J. Am. Chem. Soc.* **2001**, *123*, 12886−12898.

(21) Poon, C.; Samulski, E.; Weise, C.; Weisshaar, J. *J. Am. Chem. Soc.* **2000**, *122*, 5642−5643.