

# Polarizable Simulations with Second-Order Interaction Model (POSSIM) Force Field: Developing Parameters for Alanine Peptides and Protein Backbone

Sergei Y. Ponomarev and George A. Kaminski\*

Department of Chemistry and Biochemistry, Worcester Polytechnic Institute, Worcester, Massachusetts 01609, United States

 Supporting Information

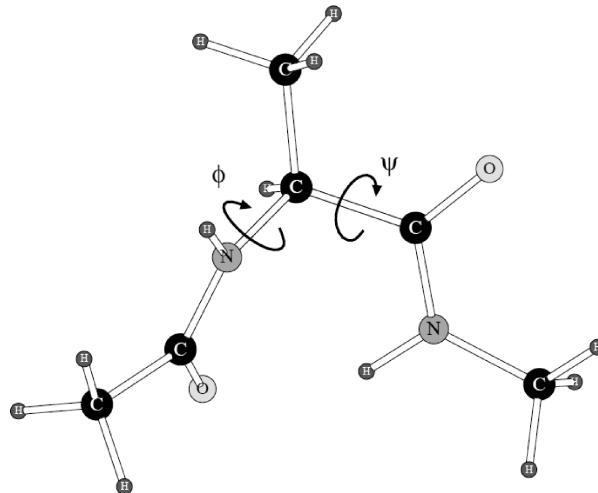
**ABSTRACT:** Polarizable simulations with second-order interaction model (POSSIM) force field has been extended to include parameters for alanine peptides and protein backbones. New features were introduced into the fitting protocol, as compared to the previous generation of the polarizable force field for proteins. A reduced amount of quantum mechanical data was employed in fitting the electrostatic parameters. Transferability of the electrostatics between our recently developed *N*-methylacetamide model and the protein backbone was confirmed. Binding energy and geometry for complexes of alanine dipeptide with a water molecule were estimated and found in a good agreement with high-level quantum mechanical results (for example, the intermolecular distances agreeing within ca. 0.06 Å). Following the previously devised procedure, we calculated average errors in alanine di- and tetrapeptide conformational energies and backbone angles and found the agreement to be adequate (for example, the alanine tetrapeptide extended globular conformational energy gap was calculated to be 3.09 kcal/mol quantum mechanically and 3.14 kcal/mol with the POSSIM force field). However, we have now also included simulation of a simple  $\alpha$  helix in both gas phase and water as the ultimate test of the backbone conformational behavior. The resulting alanine and protein backbone force field parameters are currently being employed in further development of the POSSIM fast polarizable force field for proteins.

## I. INTRODUCTION

While quantum mechanical calculations offer valuable data in a variety of biological and biomedical calculations, applications of empirical force fields remain the only way of approaching the majority of problems of interest. On one hand, they require less computer resources. On the other hand, the issue of choosing the best level of quantum theory is still a nontrivial one, and the level of quantum mechanical accuracy in a specific application is far from being guaranteed.

When empirical force fields are employed, accurate assessment of energy often requires explicit treatment of the electrostatic polarization.<sup>1</sup> The properties which depend on it include dimerization energies and acidity constants of small molecules, energies of protein–ligand interactions, protein  $pK_a$  values, or even the very thermodynamic stability of complexes in solutions. For example, we have demonstrated that that  $pK_a$  values for acidic and basic residues of the turkey ovomucoid third domain (OMTKY3) can be reproduced within 0.6 and 0.7 pH units of the experimental data with a polarizable force field. The corresponding errors with the nonpolarizable orthogonal partial least-squares (OPLS) were 3.3 and 2.2 pH units.<sup>2</sup> Formation of sugar–protein complexes represents yet another example when polarization is critical for predicting a thermodynamically stable structure.<sup>3</sup> It is generally acknowledged that polarization is an important component in many computational studies of proteins and protein–ligand complexes, although it is sometimes included in surrogate forms, such as, for example, conformation-specific protein charges.<sup>4</sup>

There are two main issues related to the empirical polarizable force field development. The first one is in the functional



**Figure 1.** Protein backbone angles  $\phi$  and  $\psi$  shown in the alanine dipeptide molecule.

form of the electrostatic polarization. Using fluctuating charges saves time and is computationally efficient in simulating uniform systems, such as pure liquid water.<sup>5</sup> However, it causes problems when out-of-plane polarization response is required or when a bifurcated hydrogen bond is formed. Therefore, the inducible dipoles approach is more adequate

**Received:** December 14, 2010

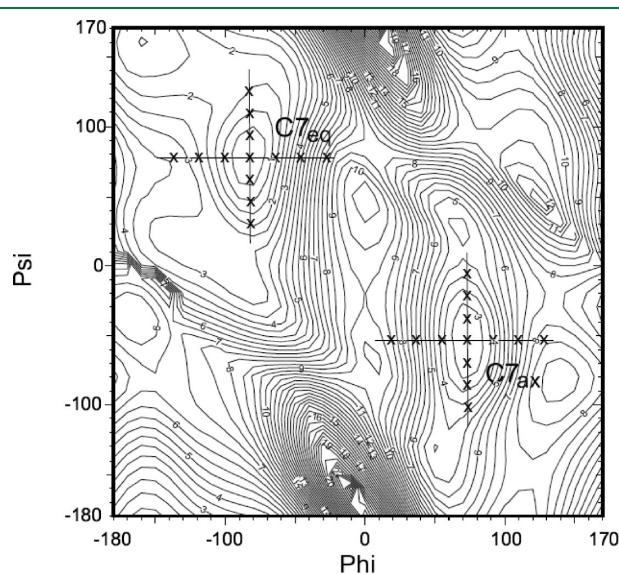
**Published:** April 14, 2011

when arbitrary systems have to be simulated with a high degree of accuracy. On the other hand, it is known that the inducible dipole technique slows down polarizable calculations significantly. In order to reduce the severity of this problem, we are applying the second-order approximation in treatment of the electrostatic polarization. It has been demonstrated to increase the speed by ca. an order of magnitude without sacrificing the accuracy.<sup>6</sup> Moreover, this approximation makes

the so-called polarization catastrophe (the resonance-like infinite growth of the induced dipole moment values) impossible. Our previous paper described development of the polarizable simulations with second order interaction model (POSSIM) software and the force field parameters for a series of small molecules, including water and *N*-methylacetamide (NMA). In this work, we describe creation of alanine and protein backbone parameter sets in the POSSIM framework.

The second issue is choosing the source of fitting data for a polarizable force field. High-level quantum mechanical results are very attractive in this respect,<sup>7,8</sup> but experimental data can be more robust. We follow the middle-of-the-road path by relying on experimental data whenever possible and by making heavy use of quantum mechanical calculations when needed. One important issue is the standard procedure of producing torsional parameters for peptides by fitting to conformational energies of di- and tetrapeptides.<sup>8</sup> We include it in our work and are describing an improved procedure for creation of the torsional parameters in the Methods Section below. At the same time, the quantum mechanical conformers employed in such calculations are created by gas-phase quantum mechanical optimizations, and they often belong to parts of the conformational space which are not found in experimental protein structures. Therefore, we have included an additional conformational test in the alanine and the backbone parameter fitting. It is known that the tridecaalanine peptide (ala-13) forms a stable  $\alpha$ -helix.<sup>9</sup> Therefore, we studied the stability of our POSSIM ala-13  $\alpha$ -helix and compared it to that of the OPLS-AA<sup>8</sup> for benchmarking. We have also discovered that the quality of the force field in reproducing the quantum mechanical di- and tetrapeptide conformational energies has a relatively weak effect on the stability of the tridecaalanine peptide in water.

Overall, the following has been derived, developed, or otherwise calculated in this work: (i) the torsional parameters for the alanine residues and the protein backbones have been produced; (ii) the binding energies of a water molecule with the alanine dipeptide as calculated with the POSSIM and OPLS-AA force fields have been compared with the quantum mechanical data to confirm transferability of the nonbonded parameters and to justify using the latter from the POSSIM NMA model in protein studies; (iii) the resulting parameters were employed in gas-phase and aqueous solution simulations of an  $\alpha$ -helix to validate the resulting POSSIM parameters as



**Figure 2.** Torsional fitting subspace, for the alanine dipeptide  $\phi/\psi$  potential energy surface. Such crosses were centered at each of the six minima, and each arm contained four fitting points (here some crosses and points are omitted for the sake of clarity).

**Table 1. Backbone Torsional Parameters, Set tors.1<sup>a</sup>**

| parameter                           | $V_1$  | $V_2$  | $V_3$  |
|-------------------------------------|--------|--------|--------|
| C-N- $\alpha$ C-C, $\phi$           | 0.667  | -0.012 | -4.003 |
| N- $\alpha$ C-C-N, $\psi$           | -2.011 | 2.528  | -4.829 |
| C-N- $\alpha$ C- $\beta$ C, $\phi'$ | -2.165 | 0.024  | 4.221  |
| $\beta$ C- $\alpha$ C-C-N, $\psi'$  | 0.594  | -0.386 | 4.378  |

<sup>a</sup> The coefficients are given in kcal/mol.

**Table 2. Conformational Energies and Angles for Alanine Dipeptide<sup>a</sup>**

| conformer  | energy |      |        | $\phi$ |        |        | $\psi$ |       |        |
|------------|--------|------|--------|--------|--------|--------|--------|-------|--------|
|            | QM     | OPLS | POSSIM | QM     | OPLS   | POSSIM | QM     | OPLS  | POSSIM |
| $C7_{eq}$  | 0.00   | 0.00 | 0.00   | -81.4  | -79.5  | -83.8  | 85.6   | 61.8  | 53.2   |
| $C5$       | 1.00   | 0.91 | 0.78   | -160.5 | -149.8 | -151.3 | 165.9  | 159.9 | 150.9  |
| $C7_{az}$  | 2.71   | 2.40 | 2.85   | 70.3   | 77.5   | 76.5   | -76.8  | -46.6 | -50.3  |
| $\beta_2$  | 2.56   | 2.82 | 2.57   | -105.1 | -105.1 | -105.1 | 10.6   | 10.6  | 10.6   |
| $\alpha_L$ | 4.21   | 5.96 | 5.41   | 68.3   | 68.3   | 68.3   | 22.4   | 22.4  | 22.4   |
| $\alpha'$  | 5.47   | 5.96 | 5.53   | -162.0 | -156.5 | -149.5 | -33.2  | -48.5 | -100.3 |
| PII        | 2.78   | 2.18 | 3.96   | -85.0  | -85.0  | -85.0  | 160.0  | 160   | 160.0  |
| $\alpha_R$ | 2.71   | 2.39 | 1.95   | -83.7  | -83.7  | -83.7  | -3.9   | -3.9  | -3.9   |
| error      | -      | 0.73 | 0.67   | -      | 3.2    | 3.8    | -      | 9.4   | 17.6   |

<sup>a</sup> Energies are in kcal/mol, and angles are in degrees. POSSIM refers to the polarizable force field with the tors.1 version of the torsional parameters. Angles  $\phi$  and  $\psi$  for conformers  $\beta_2$ ,  $\alpha_L$ , PII, and  $\alpha_R$  were fixed at their quantum mechanical values. Quantum mechanical energy minimizations were unconstrained except for PII.

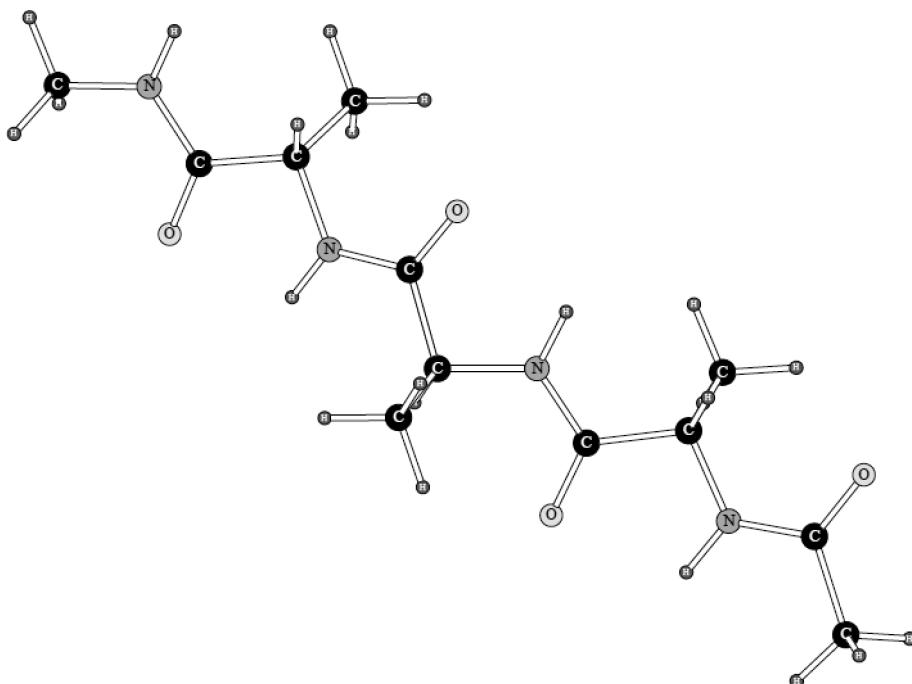


Figure 3. LMP2/cc-pVTZ(-f) geometry of the extended alanine tetrapeptide conformation.

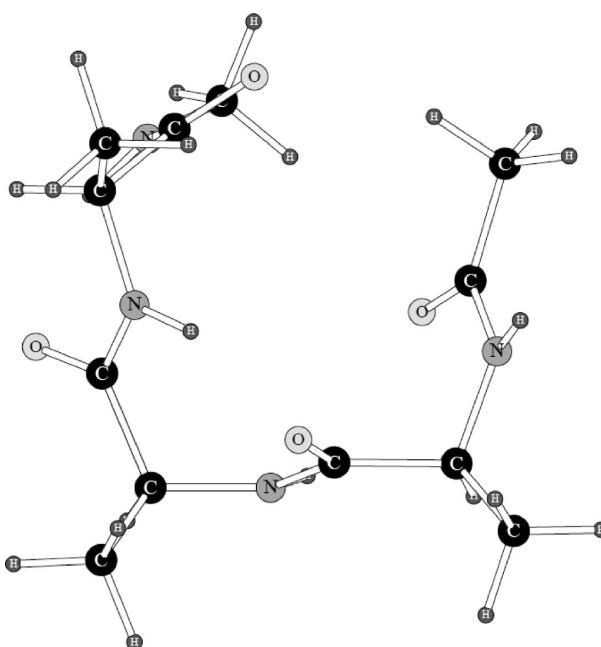


Figure 4. LMP2/cc-pVTZ(-f) geometry of the globular alanine tetrapeptide conformation.

acceptable in protein and peptide simulations. Moreover, the additional optimizations in (i) above and (ii) and (iii) altogether represent a novel development in our methodology of protein force field production, as compared to that used to create the previous version of the force field.

The rest of the paper is organized as follows: Section II is a description of the methodology involved. Section III contains results and discussion. Finally, Section IV presents the conclusions.

Table 3. Backbone Torsional Parameters, Set tors.final<sup>a</sup>

| parameter                           | $V_1$  | $V_2$  | $V_3$  |
|-------------------------------------|--------|--------|--------|
| C-N- $\alpha$ C-C, $\phi$           | 2.000  | -0.500 | -3.772 |
| N- $\alpha$ C-C-N, $\psi$           | -2.837 | 3.942  | -3.328 |
| C-N- $\alpha$ C- $\beta$ C, $\phi'$ | -2.718 | 1.757  | 5.202  |
| $\beta$ C- $\alpha$ C-C-N, $\psi'$  | 0.372  | -0.915 | 3.321  |

<sup>a</sup>The coefficients are given in kcal/mol.

## II. METHODS

**A. Force Field.** The total energy  $E_{\text{tot}}$  is a sum of the electrostatic interactions  $E_{\text{electrostatic}}$ , van der Waals energy  $E_{\text{vdW}}$ , harmonic bond stretching and angle bending  $E_{\text{stretch}}$  and  $E_{\text{bend}}$ , and the torsional term  $E_{\text{torsion}}$ :

$$E_{\text{tot}} = E_{\text{electrostatic}} + E_{\text{vdW}} + E_{\text{stretch}} + E_{\text{bend}} + E_{\text{torsion}} \quad (1)$$

**Electrostatic Energy.** The electrostatic polarization energy as calculated with inducible point dipoles  $\mu$  is

$$E_{\text{pol}} = -\frac{1}{2} \sum_i \mu_i \mathbf{E}_i^0 \quad (2)$$

where  $\mathbf{E}^0$  is the electrostatic field in the absence of the dipoles.

$$\mu_i = \alpha_i \mathbf{E}_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \mu_j \quad (3)$$

where  $\alpha$  are scalar polarizabilities, and  $T_{ij}$  is the dipole-dipole interaction tensor. The self-consistent eq 3 is usually solved iteratively. Let us explicitly write down the first two iterations:

$$\mu_i^1 = \alpha_i \mathbf{E}_i^0 \quad (4a)$$

Table 4. Conformational Energies and Angles for Alanine Dipeptide<sup>a</sup>

| conformer        | energy |      |        | $\phi$ |        |        | $\psi$ |       |        |
|------------------|--------|------|--------|--------|--------|--------|--------|-------|--------|
|                  | QM     | OPLS | POSSIM | QM     | OPLS   | POSSIM | QM     | OPLS  | POSSIM |
| C7 <sub>eq</sub> | 0.00   | 0.00 | 0.00   | -81.4  | -79.5  | -77.2  | 85.6   | 61.8  | 34.4   |
| C5               | 1.00   | 0.91 | 1.37   | -160.5 | -149.8 | -160.3 | 165.9  | 159.9 | 159.2  |
| C7 <sub>az</sub> | 2.71   | 2.40 | 2.17   | 70.3   | 77.5   | 78.1   | -76.8  | -46.6 | -36.2  |
| $\beta_2$        | 2.56   | 2.82 | 2.77   | -105.1 | -105.1 | -105.1 | 10.6   | 10.6  | 10.6   |
| $\alpha_L$       | 4.21   | 5.96 | 5.79   | 68.3   | 68.3   | 68.3   | 22.4   | 22.4  | 22.4   |
| $\alpha'$        | 5.47   | 5.96 | 5.98   | -162.0 | -156.5 | -162.9 | -33.2  | -48.5 | -38.0  |
| PII              | 2.78   | 2.18 | 3.52   | -85.0  | -85.0  | -85.0  | 160.0  | 160   | 160.0  |
| $\alpha_R$       | 2.71   | 2.39 | 0.99   | -83.7  | -83.7  | -83.7  | -3.9   | -3.9  | -3.9   |
| error            | -      | 0.73 | 0.97   | -      | 3.2    | 1.6    | -      | 9.4   | 12.9   |

<sup>a</sup> Energies are in kcal/mol, and angles are in degrees. POSSIM refers to the polarizable force field with the torsional parameters. Angles  $\phi$  and  $\psi$  for conformers  $\beta_2$ ,  $\alpha_L$ , PII,  $\alpha_R$  were fixed at their QM values. QM energy minimizations were unconstrained except for PII.

$$\boldsymbol{\mu}_i^H = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \boldsymbol{\mu}_j^I = \alpha_i E_i^0 + \alpha_i \sum_{j \neq i} T_{ij} \alpha_j E_j^0 \quad (4b)$$

We are using the second-order expression in eq 4b. It has been previously shown to yield ca. an order of magnitude increase of the computational speed with no loss of accuracy.<sup>6</sup> The electrostatic energy also includes the pairwise additive contribution from interactions of permanent charges:

$$E_{\text{additive}} = \sum_{i \neq j} \frac{q_i q_j}{R_{ij}} f_{ij} \quad (5)$$

The factor  $f_{ij}$  is set to 0 for 1,2- and 1,3-pairs (atoms which belong to the same valence bond or angle), to 0.5 for 1,4-interactions (atoms in the same dihedral angle), and to 1.0 otherwise.

To avoid unphysical increase of the electrostatic interactions at short distances, each atom type has a cutoff parameter  $R_{\text{cut}}$ . When the overall distance  $R_{ij}$  is smaller than the sum of these parameters  $R_{\min}^{ij} = R_{\text{cut}}^i + R_{\text{cut}}^j$  for the atoms  $i$  and  $j$ ,  $R_{ij}$  is replaced by an effective smooth function:

$$R_{ij}^{\text{eff}} = \left( 1 - \left( \frac{R_{ij}}{R_{\min}^{ij}} \right)^2 + \left( \frac{R_{ij}}{R_{\min}^{ij}} \right)^3 \right) \cdot R_{\min}^{ij} \quad (6)$$

The following important points about the second-order approximation in eq 4b should be made: First of all, we do not fit parameters using the full-scale polarization solution to eq 3 to later employ eq 4b as an approximate technique during the simulations. For our practical purposes, eq 4b is, in fact, the representation of the many-body interactions. It does differ from the true physical point-dipole approximation, and thus we always carefully monitor whether any errors are introduced by not computing inducible dipoles with the complete iterative procedure. So far, simulations of gas-phase dimers, quantum mechanical electrostatic three-body energies, pure liquids, solutions and peptides have given us no indication that the second-order approximation leads to any deficient physical results, and we have always been able to produce fitting to quantum mechanical and experimental data which was as good as for the full-scale polarization.<sup>6,10</sup> Moreover, application of the second-order approximation given by eq 4b turns the expression for the inducible dipoles into an analytical one, thus eliminating the

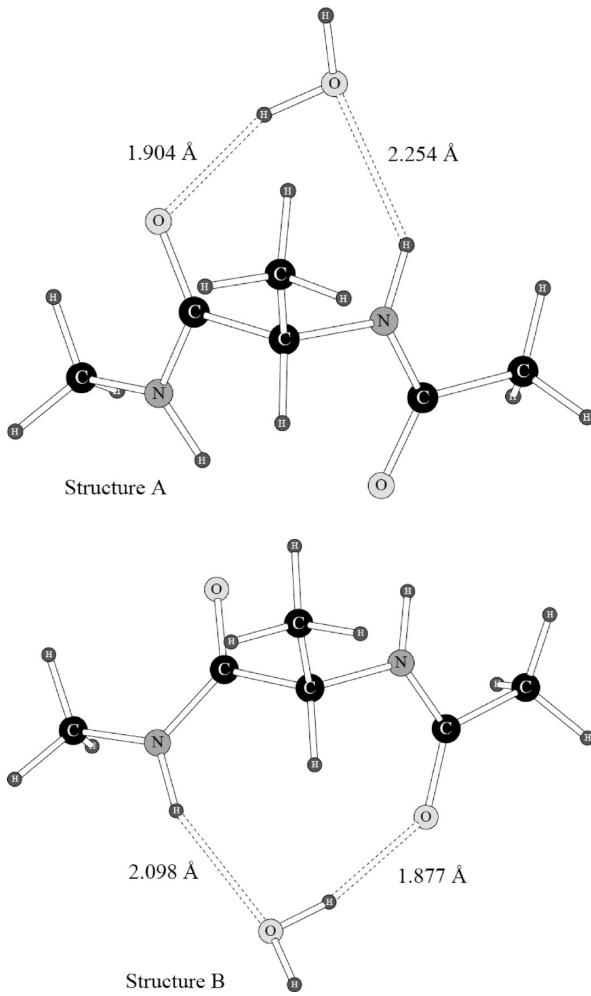


Figure 5. Two Alanine dipeptide hydrogen bonded complexes with a water molecule.

possibility of the polarization catastrophe. This can also become a very useful feature in future developments, e.g., in creating a continuum dielectric model, as convergence issues are known to be of importance for continuum solvation techniques.

**Table 5.** Results of Simulating Alanine Dipeptide Complexes with Water<sup>a</sup>

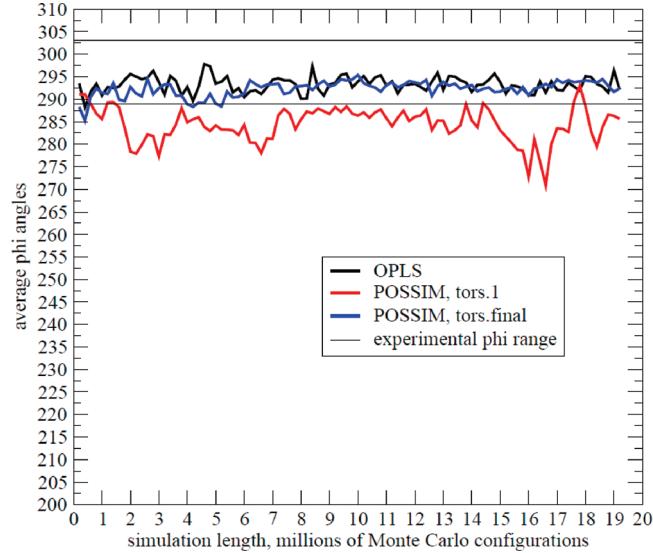
| property         | structure A |       |        | structure B |        |        |
|------------------|-------------|-------|--------|-------------|--------|--------|
|                  | QM          | OPLS  | POSSIM | QM          | OPLS   | POSSIM |
| binding energy   | -9.80       | -9.48 | -7.95  | -9.73       | -11.19 | -9.34  |
| R(O...O)         | 2.83        | 2.75  | 2.84   | 2.83        | 2.72   | 2.73   |
| R(O...N)         | 3.10        | 2.89  | 3.09   | 3.05        | 2.86   | 2.95   |
| $\phi$ , dimer   | -83.4       | -87.3 | -80.2  | -84.3       | -89.4  | -86.9  |
| $\psi$ , dimer   | 90.3        | 114.6 | 83.7   | 131.2       | 113.3  | 122.2  |
| $\phi$ , monomer | -79.7       | -79.5 | -77.2  | -79.7       | -79.5  | -77.2  |
| $\psi$ , monomer | 88.1        | 61.8  | 34.4   | 88.1        | 61.8   | 34.4   |

<sup>a</sup> Energies are in kcal/mol, distances in Å, angles in degrees.

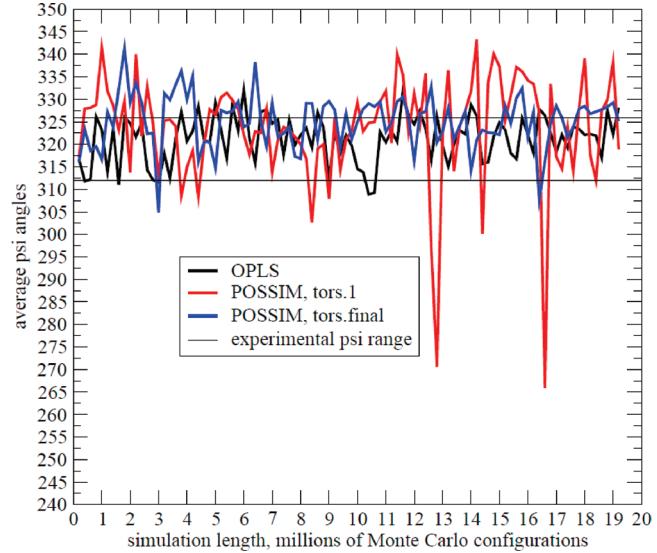
**Table 6.** Results of Simulating Alanine Dipeptide Complexes with Water<sup>a</sup>

| property         | structure A |        |        | structure B |        |        |
|------------------|-------------|--------|--------|-------------|--------|--------|
|                  | QM          | OPLS   | POSSIM | QM          | OPLS   | POSSIM |
| binding energy   | -10.71      | -10.04 | -9.75  | -11.68      | -11.79 | -12.24 |
| R(O...O)         | 2.83        | 2.81   | 2.82   | 2.83        | 2.75   | 2.74   |
| R(O...N)         | 3.10        | 2.94   | 3.06   | 3.05        | 2.94   | 2.98   |
| $\phi$ , dimer   | -83.4       | -83.4  | -83.4  | -84.3       | -84.3  | -84.3  |
| $\psi$ , dimer   | 90.3        | 90.3   | 90.3   | 131.2       | 131.2  | 131.2  |
| $\phi$ , monomer | -83.4       | -83.4  | -83.4  | -84.3       | -84.3  | -84.3  |
| $\psi$ , monomer | 90.3        | 90.3   | 90.3   | 131.2       | 131.2  | 131.2  |

<sup>a</sup>  $\phi$  and  $\psi$  of both dimers and monomers are fixed in the quantum mechanical dimer positions. Energies are in kcal/mol, distances in Å, angles in degrees.



**Figure 6.** Average  $\phi$  angles in the  $\alpha$ -helix gas-phase simulations vs the simulation length.



**Figure 7.** Average  $\psi$  angles in  $\alpha$ -helix gas-phase simulations vs the simulation length.

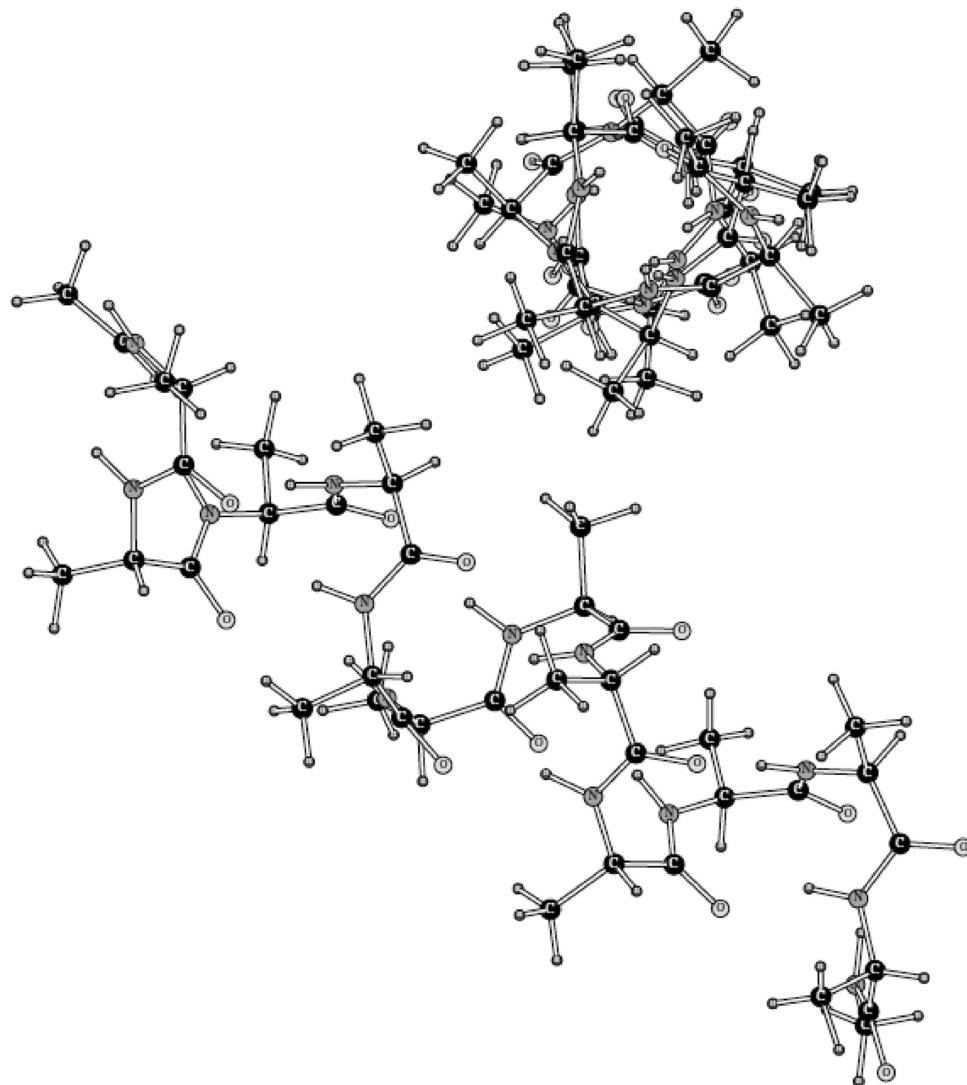
*The Rest of the Force Field.* We are using the standard Lennard-Jones formalism for the van der Waals energy:

$$E_{\text{vdW}} = \sum_{i \neq j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{R_{ij}} \right)^6 \right] f_{ij} \quad (7)$$

Geometric combining rules are applied ( $\epsilon_{ij} = (\epsilon_i \cdot \epsilon_j)^{1/2}$ ,  $\sigma_{ij} = (\sigma_i \cdot \sigma_j)^{1/2}$ ). Bond stretching and angle bending are computed

with the usual harmonic formalism, and the torsional term is calculated as

$$\begin{aligned} E_{\text{torsion}} = & \sum_i \frac{V_1^i}{2} [1 + \cos(\varphi_i)] + \frac{V_2^i}{2} [1 - \cos(2\varphi_i)] \\ & + \frac{V_3^i}{2} [1 + \cos(3\varphi_i)] \end{aligned} \quad (8)$$



**Figure 8.** Structure of the ala-13  $\alpha$ -helix simulated with OPLS in gas phase, after  $19 \times 10^6$  Monte Carlo configurations.

The fixed-charges OPLS-AA force field used for benchmarking is functionally the same, except that it lacks the polarization part of the electrostatic energy.

**B. Parameterization of the Force Field.** Whenever possible, the force field parameters for the alanine peptides were adopted directly from the previously created NMA parameter values.<sup>10</sup> The only completely new parameters were those for the backbone torsions. This is different from the previous version of the polarizable force field (PFF) for proteins in which electrostatic parameters for the alanine (and thus for the backbone) were also refitted.<sup>7</sup> Therefore, we believe that the present work demonstrates a greater degree of utilizing parameter transferability.

Fitting of torsional parameters for the protein backbone  $\phi$  and  $\psi$  angles (Figure 1) cannot be done separately from each other, as the torsions are coupled.

The initial part of our torsional fitting was the same as used before.<sup>7,8</sup> (i) The fitting was done to an ab initio data obtained previously<sup>8</sup> at the LMP2/cc-pVTZ(-f)//HF-6-31G\*\* level with Jaguar software suite.<sup>11</sup> (ii) The choice of the fitting subspace is illustrated in Figure 2. Out of the six alanine dipeptide local minima previously used,<sup>7,8</sup> only two are shown for the sake of

clarity. (iii) We used the following non-Boltzmann weighting scheme for the error at the fitting points:

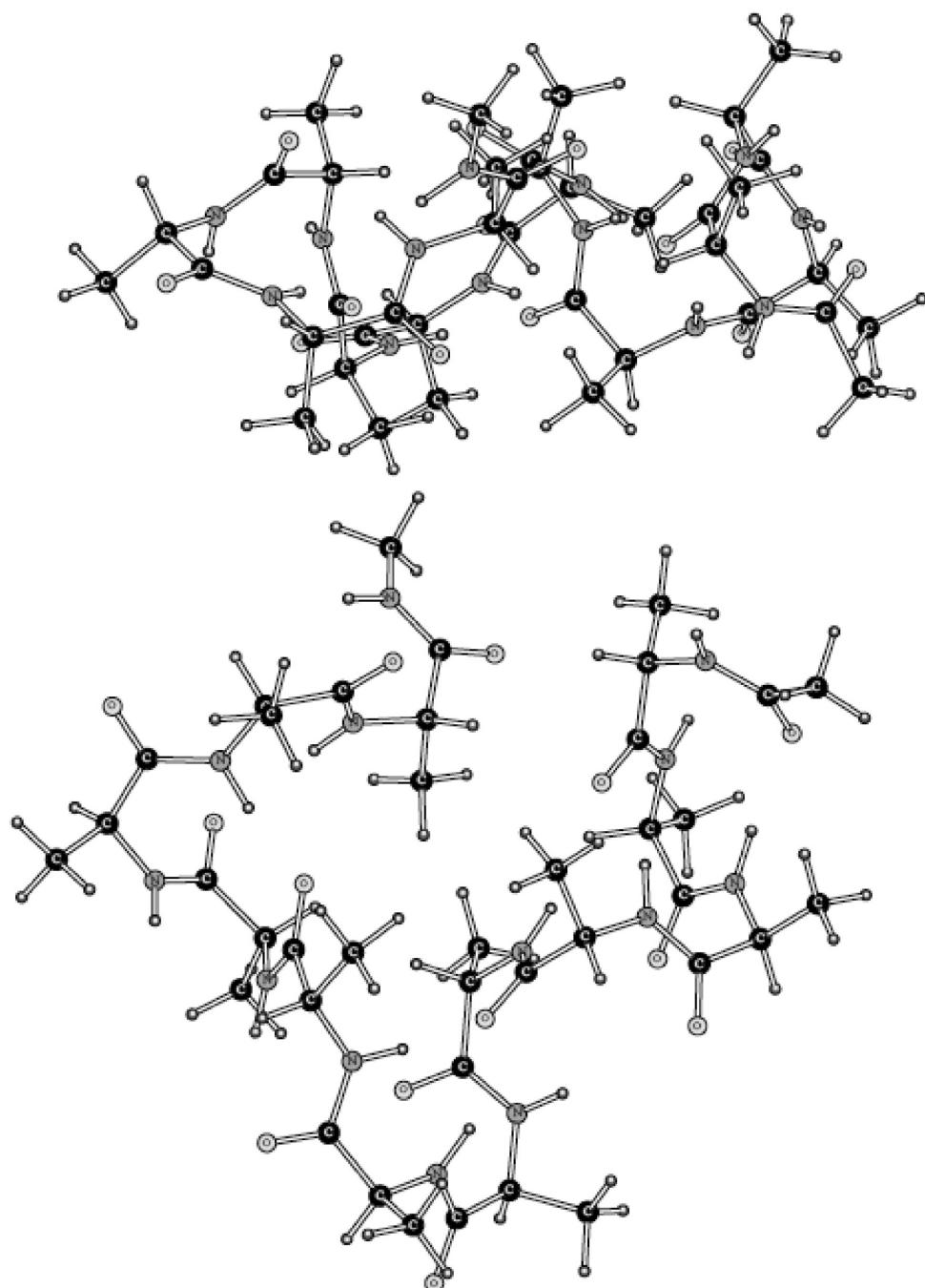
$$W_i = A \cdot \exp(-b \cdot G_i) \quad (9)$$

Here  $G_i$  is the magnitude of the torsional surface gradient at the point  $i$ , and  $W_i$  is the weight. This way more importance was given to the points with low gradients (near the minima).

In the presented work, we used the procedure described above only to produce the initial guess for the torsional parameters to be employed in eq 8. After that, the following approach was taken. The errors in the conformational energies were combined with the errors in the conformational angles  $\phi$  and  $\psi$  to produce the error function as shown in eq 10:

$$\text{erf} = \sum_i (E_i^0 - E_i)^2 + \sum_j (\phi_j^0 - \phi_j)^2 + (\psi_j^0 - \psi_j)^2 \quad (10)$$

Here  $E_i^0$  and  $E_i$  are the quantum mechanical and empirical conformational energies for all the conformers  $i$ , and the second sum contains the values of the backbone angles  $\phi$  and  $\psi$ . The



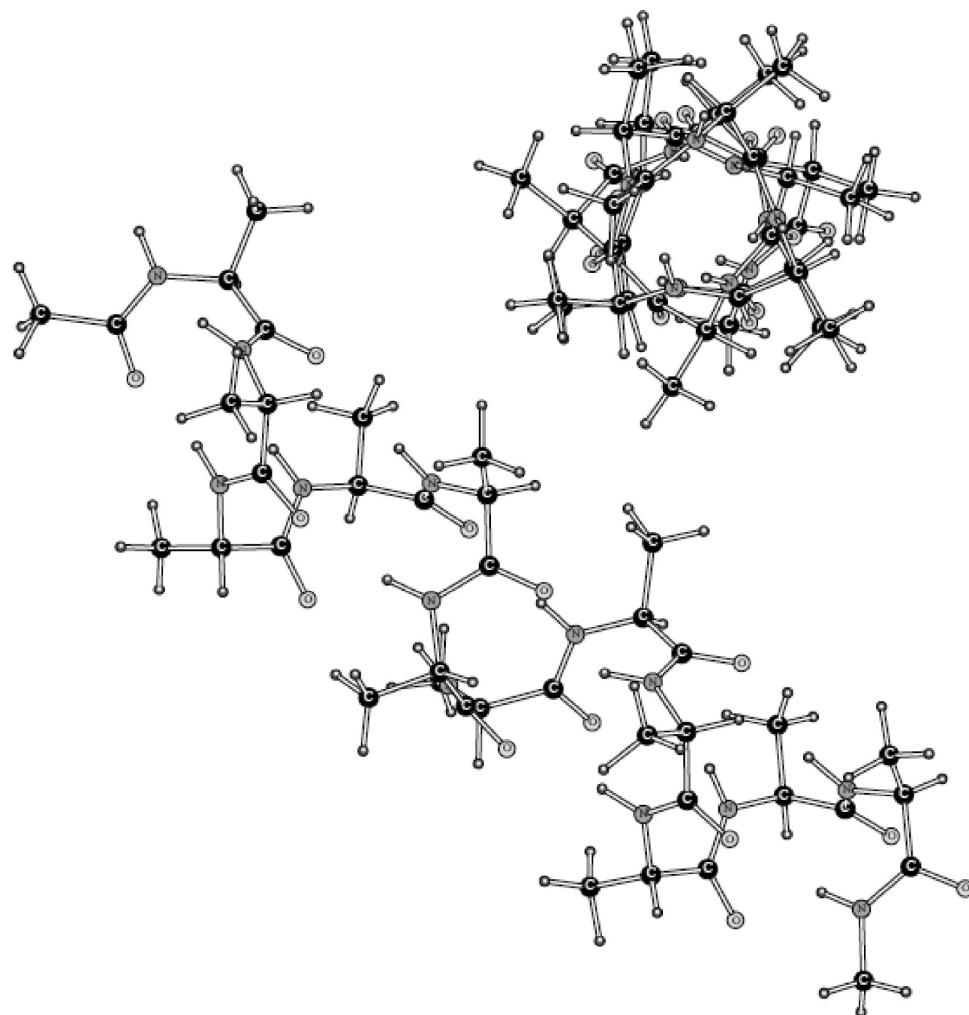
**Figure 9.** Structure of the ala-13  $\alpha$ -helix simulated with POSSIM, version tors.1, in gas phase, after  $19 \times 10^6$  Monte Carlo configurations.

error function was minimized as a function of the torsional parameters in eq 8.

**C. Calculating Dimerization Energies for the Alanine Dipeptide Complexes with Water.** In order to test the transferability of the NMA nonbonded parameters employed for our alanine and protein backbone model, we calculated energies of interaction of the alanine-dipeptide with a water molecule. Structures and energies obtained for these systems with the POSSIM program were compared to the quantum mechanical results obtained with Jaguar.<sup>11</sup> For hydrogen bonds, a good level of accuracy can be achieved via MP2 calculations extrapolated to the basis set limit, where the

contribution of higher level excitations (e.g., CCSD(T)) has been shown to be negligible (except for some cases, such as  $\pi$  stacking of aromatic rings, where the MP2 level has been shown to not be sufficient).

Briefly, dimer geometries were obtained by LMP2 optimizations with a cc-pVTZ(-f) basis set. The final quantum mechanical dimer binding energy  $E_{\text{bind}}$ , as used in this work, is a linear combination of the LMP2 binding energy for a smaller cc-pVTZ(-f) basis set ( $E_{\text{ccpvtz}}$ ) and the LMP2 binding energy with a larger cc-pVQZ(-g) basis set ( $E_{\text{ccpvq}}$ ).<sup>15</sup> This method has been previously demonstrated to produce a high-quality fitting and benchmarking data for force field development.<sup>7,8</sup>



**Figure 10.** Structure of the ala-13  $\alpha$ -helix simulated with POSSIM, version tors.final in gas phase, after  $19 \times 10^6$  Monte Carlo configurations.

**D. Gas-Phase and Liquid-State Simulations of the Tridecaalanine Peptide.** In order to give our alanine and backbone model a final test, we carried out simulations of a tridecaalanine (ala-13) peptide both in gas-phase and in aqueous solution at 25 °C and 1 atm. The initial structure was set at the  $\alpha$ -helix conformation, with  $\phi = 296^\circ$  and  $\psi = 319^\circ$ , and the simulations proceeded with all the degrees of freedom completely unconstrained. It is known experimentally that an  $\alpha$ -helix represents a stable conformation of alanine peptides, including ala-13, both in gas-phase and in aqueous solution.<sup>9</sup> We intended to show that our POSSIM force field for the alanine and backbone protein systems performs reasonably well under these conditions and is thus sufficiently robust to be successfully employed in protein and protein–ligand studies.

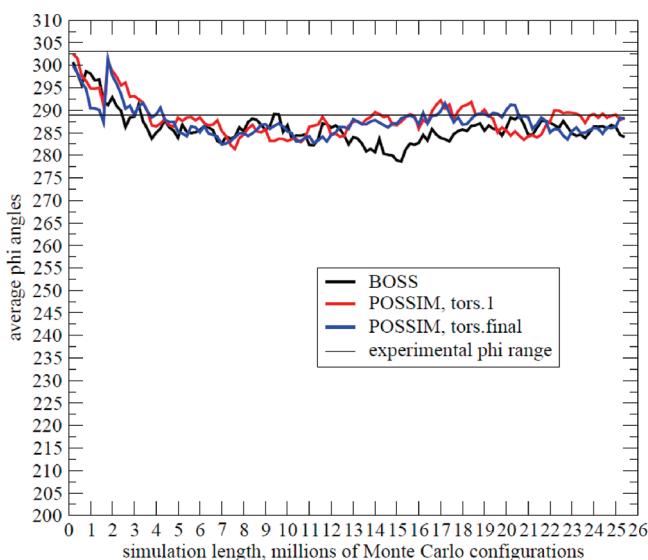
Gas-phase and hydrated simulations consisted of at least  $18 \times 10^6$  and  $25 \times 10^6$  Monte Carlo configurations, respectfully, to ensure convergence. A 7 Å dipole–dipole cutoff was used. An 8 Å cutoff was employed for the intermolecular interactions in solution (including both the solute–solvent and solvent–solvent interactions). The standard correction procedure to account for the Lennard-Jones interactions beyond the cutoff was used. The electrostatic interactions were quadratically feathered over the last 0.5 Å before the cutoff distance. A rectangular

box with periodic boundary conditions was used. The box contained 948 water molecules. The initial box setup was done to have 10 Å of water on each side of the hydrated ala-13 molecule. After that, the isobaric–isothermal (NPT) ensemble was used, with Metropolis Monte Carlo technique. In the case of the OPLS simulations, a three-site model was used with TIP3P<sup>13</sup> nonbonded parameters and flexible bond lengths and bond angles. A flexible three-site POSSIM water model<sup>10</sup> was employed in the polarizable runs.

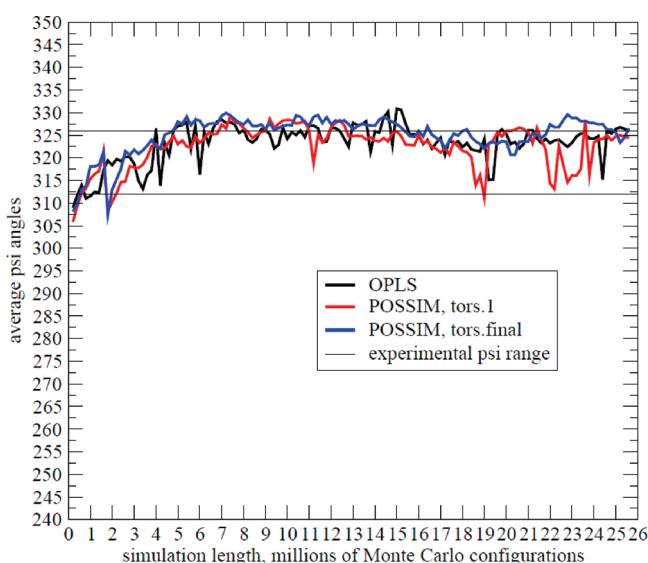
All the calculations which did not involve quantum mechanics (i.e., geometry optimizations and Monte Carlo runs) were performed with our previously introduced POSSIM software suite.<sup>10</sup> Whenever possible, comparison with the fixed charges OPLS-AA force field was done, and the OPLS-AA results were also calculated with the POSSIM program.

### III. RESULTS AND DISCUSSION

**A. Alanine Dipeptide and Tetrapeptide Conformational Energies and Angles.** We have followed the previously established procedure of calculating the alanine di- and tetrapeptide conformational energies and  $\phi$  and  $\psi$  values as the initial assessment of the quality of the parameters for the alanine and protein backbones. The same set of the



**Figure 11.** Average  $\phi$  angles in  $\alpha$ -helix simulations in aqueous solution vs the simulation length, in millions of Monte Carlo configurations.



**Figure 12.** Average  $\psi$  angles in  $\alpha$ -helix simulations in aqueous solution vs the simulation length, in millions of Monte Carlo configurations.

conformers that was employed in the previous studies was used.<sup>7,8</sup> The production of the torsional parameters proceeded as described in the Methods Section. First, weighted fitting to rotamer energies was carried out. The resulting parameters are shown in Table 1 (torsional parameters which are not listed were the same as in the NMA model).<sup>10</sup> We denote this set of parameters as tors.1, as opposed to the final set tors.final. Given in Table 2 are conformational energies and  $\phi$  and  $\psi$  values, as computed with the quantum mechanics, POSSIM and OPLS. In addition to the six conformers used in previous studies, we have also added PII and  $\alpha_R$  which are more relevant in aqueous solution.<sup>14</sup> Quantum mechanical optimizations were done at the LMP2/cc-pVTZ-(f) level. In both OPLS and POSSIM calculations, conformers  $\beta_2$ ,  $\alpha_L$ , PII, and  $\alpha_R$  had the backbone dihedral angles

fixed at the quantum mechanical values. It is known that molecular mechanics usually does not reproduce these conformers well. Overall, the performance of both POSSIM and OPLS is satisfactory. The POSSIM results have a slightly lower error in the conformational energies, while the OPLS results are closer to the quantum mechanics in terms of the geometries.

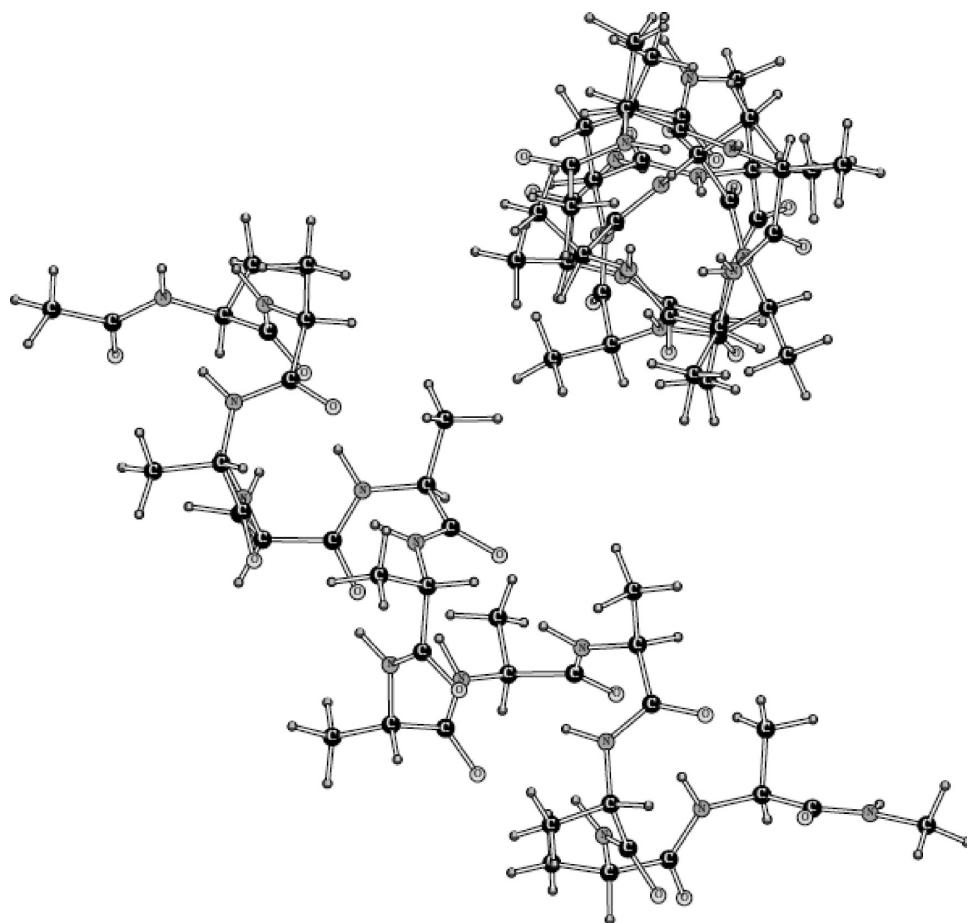
We have also calculated relative energies of the extended and globular conformations of the alanine tetrapeptide (shown on Figures 3 and 4, respectively). We determined the quantum mechanical energy difference for these conformers to be 3.09 kcal/mol, the globular form being the global energy minimum. At the same time, this quantity is known to have a relatively large range of calculated quantum mechanical energies. For example, ref 15 lists the globular–extended energy gap for the alanine tetrapeptide to be between 2.88 and 4.99 kcal/mol. The POSSIM result with the tors.1 torsional parameters set was 2.53 kcal/mol, and the OPLS result was 3.51 kcal/mol.<sup>7,8</sup>

We then further refined the backbone torsional parameters as described in the Methods section. The resulting values of the torsional Fourier coefficients and the conformational energies and angles are given in Tables 3 and 4, respectively. This set of the torsional parameters is termed tors.final, and this is the final set for the POSSIM protein backbone  $\phi$  and  $\psi$ . The average dipeptide conformational energy error is now slightly higher at 0.97 kcal/mol, but the average errors in the backbone angles  $\phi$  and  $\psi$  are reduced to  $1.6^\circ$  and  $12.9^\circ$ , respectively. Moreover, the globular–extended energy gap in the tetrapeptide is 3.14 kcal/mol, in a better agreement with the quantum mechanical results (3.09 kcal/mol with our calculations and 2.88–4.99 kcal/mol from the data ref 15). The value of the  $\psi$  for the C7<sub>eq</sub> conformer is lower now, but this part of the conformational space is not relevant in practical protein applications. The overall average error in both backbone angles was reduced.

**B. Alanine Dipeptide–Water Dimerization Energies and Distances.** There are four possible water hydrogen bonding sites in the alanine dipeptide—two NH hydrogens and two carbonyl oxygen atoms. However, our quantum mechanical energy minimizations have demonstrated that water molecules prefer to make two hydrogen bonds at the same time, one with the H and one with the O atoms. Therefore, there are only two water–alanine dipeptide heterodimer structures, as shown on Figure 5.

The quantum mechanical structures were used as the initial guesses for the POSSIM optimizations. Both POSSIM and OPLS-AA were utilized. We compared the binding energies, as well as the geometries of the complexes. Both hydrogen bonding distances ( $O \cdots H-N$ ) and  $H \cdots O=C$ ) and the  $\phi$  and  $\psi$  angles of the alanine dipeptide backbone were used for the comparison. The results of these calculations are presented in Table 5. The quantum mechanical energy of the dimerization is reproduced slightly better with the OPLS, the average error being 0.89 kcal/mol vs 1.12 kcal/mol with POSSIM. The latter tends to underestimate the magnitude of the binding energy. This is not unexpected. The nonbonded parameters for the alanine dipeptide have been adopted from NMA fitting.

And the same tendency was also present in the NMA case, with the POSSIM underestimating the NMA–water binding energy by an average of 0.89 kcal/mol.<sup>10</sup> The overall performance



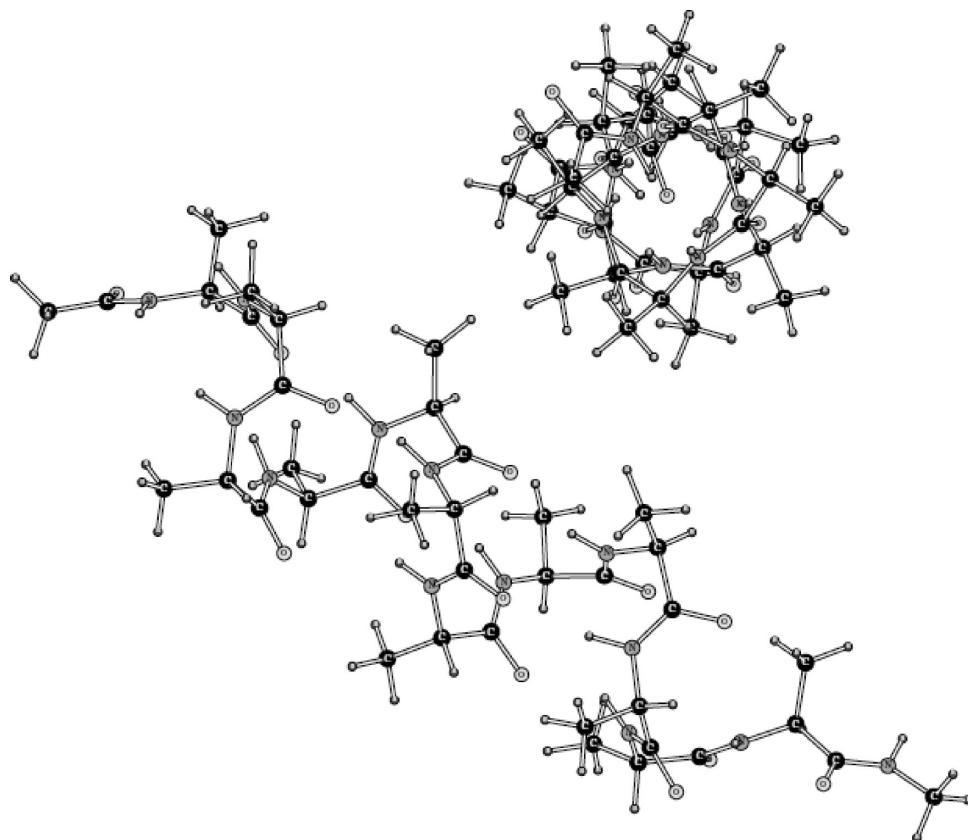
**Figure 13.** Structure of the ala-13  $\alpha$ -helix simulated with OPLS, in aqueous solution, after  $25 \times 10^6$  Monte Carlo configurations. Water molecules are not shown for the sake of clarity.

of the NMA parameters was very good. This included reproducing liquid NMA heat of vaporization and density. Which lead us to the conclusion that our quantum mechanical NMA–water binding energies are probably somewhat overestimated. Therefore, a similar trend in the alanine dipeptide complex formation with water could have been expected and is not at all an indication of problems with the protein POSSIM force field. Moreover, it can be easily seen from the data in Table 5 that the POSSIM performed noticeably better than the OPLS in reproducing the hydrogen-bond lengths, which are probably given much more accurate than the energies by the quantum mechanics. The average errors in these lengths are 0.15 and 0.06 Å with the OPLS and POSSIM calculations, respectively.

It is also worth noting that the values of the  $\phi$  and  $\psi$  backbone angles in this complex, as computed with the POSSIM, are much closer to the resulting quantum mechanical values of these angles than their OPLS counterparts, with the average error of only  $5.3^\circ$  vs  $12.8^\circ$ . This is so even though the POSSIM gives the lowest-energy monomer conformer (C7eq)  $\psi$  angle of only  $34.4^\circ$  vs the quantum mechanical  $88.1^\circ$  and the OPLS  $61.8^\circ$ . We believe that this fact confirms that: (i) the conformational energy surface is rather flat at that region, and so the precise location of the minimum is not entirely crucial; and (ii) the POSSIM force field is robust and adequate in reproducing important binding geometries.

We have further investigated the alanine dipeptide–water binding properties by running calculations, in which the values of  $\phi$  and  $\psi$  were kept the same as in the fully optimized quantum mechanical dimers in all the cases (quantum mechanical, OPLS, and POSSIM monomers and also the OPLS and POSSIM dimers). The results are presented in Table 6. The structure B dimerization energy as computed with the POSSIM is slightly greater than the quantum mechanical one in this case ( $-12.2$  vs  $-11.7$  kca/mol), otherwise the trends are the same as in the fully relaxed geometry optimizations. The average errors in the dimerization energies with the POSSIM and OPLS are 0.76 and 0.39 kcal/mol, respectively. The POSSIM and OPLS errors in the hydrogen-bonding distances are 0.09 and 0.05 Å. Interestingly, the improvement in geometry achieved by fixing the backbone angles is greater with the OPLS than it is with the POSSIM. Once again, we believe this indicates that, even though the C7eq conformational geometry is better reproduced with the OPLS, the more important binding properties are better assessed with the POSSIM force field.

**C. Gas-Phase and Hydrated Simulations of the Tridecaalanine Peptide (Ala-13).** We have carried out Monte Carlo simulations of the ala-13 in order to test the robustness of the POSSIM force field by assessing stability of this experimentally known  $\alpha$ -helical peptide. While quantum mechanical gas-phase alanine dipeptide conformational energies and



**Figure 14.** Structure of the ala-13  $\alpha$ -helix simulated with POSSIM, version tors.1, in aqueous solution, after  $25 \times 10^6$  Monte Carlo configurations. Water molecules are not shown for the sake of clarity.

geometries are important in fitting, these simulations provided a direct comparison with the available experimental observations. In particular, we were assessing the general stability of the helix and the average values of the backbone  $\phi$  and  $\psi$  angles. Figures 6 and 7 show graphs of the average values of these angles as a function of the simulation length (in millions of Monte Carlo configurations) for the OPLS force field, as well as with POSSIM, using both the tors.1 and tors.final torsional parameters. Each angle value represents averaging over the last 200 000 configurations before the indicated simulation length.

The experimental values of the backbone  $\phi$  and  $\psi$  in an  $\alpha$ -helix are  $296^\circ$  and  $319^\circ$ , respectively, with a  $7^\circ$  uncertainty.<sup>16</sup> In finding the average values of the backbone angles, we disregarded one residue on each end of the helix.

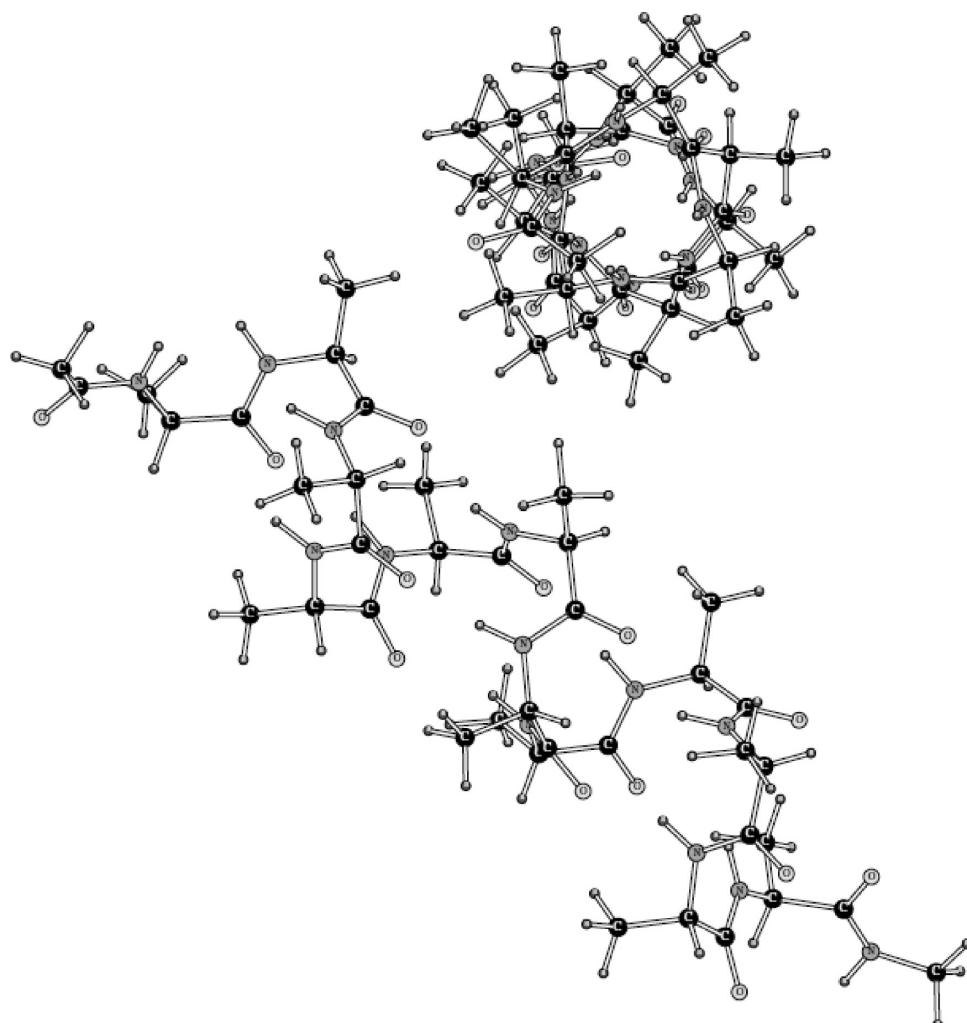
Two conclusions can be made from the presented results. First, the final versions of the POSSIM as well as the OPLS force field yield better agreement with the experimental data than the POSSIM version with the tors.1 parameters. Second, the  $\phi$  values are more stable than those of the angle  $\psi$  with all the force fields tested.

But one should keep in mind that the experimental data represent crystallographic results, and thus the thermal motion allowed in the Monte Carlo calculations can cause oscillations beyond the  $\pm 7^\circ$  experimental lines. Overall, we can conclude that the gas-phase simulations confirm that the newly developed POSSIM force field is stable and robust. They reproduce the experimentally observed  $\alpha$ -helix

gas-phase stability (see also Supporting Information). The stability of the simulated helices can also be evaluated by studying the final structure of the system shown in Figures 8–10. One can see that, while the OPLS and POSSIM with tors.final produce a stable  $\alpha$ -helix, the POSSIM/tors.1 helix denatures. At the same time, the average  $\phi$  and  $\psi$  angles in the tors.1 version of POSSIM are not extremely far from the experimental data, therefore the helicity of the structure is at least partially conserved.

Average values of the  $\phi$  and  $\psi$  angles as a function of the simulation length for the ala-13 peptide in water are shown on Figures 11 and 12. In this case, as can be expected, the stability of the both angles is greater, and the deviations are smaller. It should be noted that the angle  $\phi$  tends to be too low compared to the experimental crystallographic values, while the angle  $\psi$  is somewhat too high, thus their sum stays roughly at the same spot as the experimental one ( $255^\circ$  or  $-105^\circ$ ), and the  $\alpha$ -helicity of the structure for all the force fields employed is good.

Structures of the ala-13 peptide after  $25 \times 10^6$  Monte Carlo configurations in water are given on Figures 13–15. Water molecules are not removed for clarity. It can be seen from the figures, in combination with the graphs and the table for the liquid-state simulations, that in this case (hydrated ala-13) all three force fields (OPLS and the two versions of POSSIM) perform adequately, and no denaturation of the tridecaalanine  $\alpha$ -helix is observed.



**Figure 15.** Structure of the ala-13  $\alpha$ -helix simulated with POSSIM, version tors.final, in aqueous solution, after  $25 \times 10^6$  Monte Carlo configurations. Water molecules are removed for clarity.

#### IV. CONCLUSIONS

We have presented results of developing a fast polarizable POSSIM force field for alanine and protein backbones. The quantum mechanical data set used for fitting was streamlined and simplified as compared to the previous version of the complete polarizable force field for proteins, and a high degree of transferability of the potential energy parameters has been demonstrated.

We have included a previously unused step of calculating dipeptide dimerization energies with a water molecule as an additional proof of validity of the technique and the resulting force field. The POSSIM force field performs well in this test.

The torsional fitting procedure has been augmented by a new step, a direct optimization-type fitting of the torsional parameters to the quantum mechanical conformational energies and structures.

At the same time, we believe that quantum mechanical dipeptide conformers in themselves are not a sufficient tool in validation of a force field. One of the reasons for this assumption is that most of these conformers belong to parts of the total conformational space which are rarely found in experimentally known proteins. Therefore, we have included an additional step to further test the robustness of the POSSIM force field. We have

simulated the tridecaalanine peptide (ala-13) in both gas phase and aqueous solution with the Monte Carlo technique. This peptide is experimentally known to form an  $\alpha$ -helix under these conditions. The POSSIM ala-13 (and the OPLS used for benchmarking) was found to maintain a stable  $\alpha$ -helical conformation as well.

We conclude that the resulting polarizable POSSIM force field is adequately accurate, and we will use this model for the alanine and protein backbones as the basis for further development of a complete polarizable POSSIM force field for proteins.

#### ■ ASSOCIATED CONTENT

**S Supporting Information.** Tabulated values of  $\phi$  and  $\psi$  angles of the  $\alpha$ -helix in gas phase and solution as a function of simulation length. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### ■ AUTHOR INFORMATION

##### Corresponding Author

\*E-mail: gkaminski@wpi.edu.

## ■ ACKNOWLEDGMENT

This project was supported by grant no. R01GM074624 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. The authors express gratitude to Schrödinger, LLC, for the Jaguar and Impact software.

## ■ REFERENCES

- (1) (1) See, for example: (a) Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4177–4178. (b) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comput. Chem.* **2001**, *22*, 1048–1057. (c) Kaminski, G. A. *J. Phys. Chem. B* **2005**, *119*, 5884–5890. (d) Jiao, D.; Zhang, J. J.; Duke, R. E.; Li, G. H.; Schneiders, M. J.; Ren, P. Y. *J. Comput. Chem.* **2009**, *30*, 1701–1711. (e) Hernandez, G.; Anderson, J. S.; LeMaster, D. M. *Biochemistry* **2009**, *48*, 6482–6494. (f) Wang, X. Y.; Zhang, J. Z. H. *Chem. Phys. Lett.* **2011**, *501*, 508–512.
- (2) (a) MacDermaid, C. M.; Kaminski, G. A. *J. Phys. Chem. B* **2007**, *111*, 9036–9044. (b) Click, T. H.; Kaminski, G. A. *J. Phys. Chem. B* **2009**, *113*, 7844–7850.
- (3) Veluraja, K.; Margulis, C. J. *J. Biomol. Struct. Dyn.* **2005**, *23*, 101–111.
- (4) (a) Ji, C.; Mei, Y.; Zhang, J. Z. H. *Biophys. J.* **2008**, *95*, 1080–1088. (b) Ji, C. G.; Zhang, J. Z. H. *J. Phys. Chem. B* **2009**, *113*, 16059–16064.
- (5) For representative publications see: (a) Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141–6156. (b) Liu, Y. P.; Kim, K.; Berne, B. J.; Friesner, R. A.; Rick, S. W. *J. Chem. Phys.* **1998**, *108*, 4739–4755. (c) Ramon, J. M. H.; Rios, M. A. *Chem. Phys.* **1999**, *250*, 155–169. (d) Gonzalez, M. A.; Enciso, E.; Bermejo, F. J.; Bee, M. J. *Chem. Phys.* **1999**, *110*, 8045–8059. (e) Soetens, J. C.; Jansen, G.; Millot, C. *Mol. Phys.* **1999**, *96*, 1003–1012. (f) Dang, L. X. *J. Chem. Phys.* **2000**, *113*, 266–273. (g) Chen, B.; Xing, J. H.; Siepmann, J. I. *J. Phys. Chem. B* **2000**, *104*, 2391–2401. (h) Jedlovszky, P.; Vallauri, R. *J. Chem. Phys.* **2001**, *115*, 3750–3762. (i) Ribeiro, M. C. C. *Phys. Rev. B* **2001**, *63*(09), 4205. (j) Rinker, S.; Gunsteren, W. F. *J. Chem. Phys.* **2011**, *134*, 084110. (k) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerrel, A. D.; Schulten, K.; Roux, B. *J. Phys. Chem. Lett.* **2011**, *2*, 87–92.
- (6) Kaminski, G. A.; Zhou, R.; Friesner, R. A. *J. Comput. Chem.* **2003**, *24*, 267–276.
- (7) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. *J. Comput. Chem.* **2002**, *23*, 1515–1531.
- (8) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (9) (a) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286–5290. (b) Scholtz, J. M.; York, E. J.; Steward, J. M.; Baldwin, R. L. *J. Am. Chem. Soc.* **1991**, *113*, 5102–5104. (c) Scholtz, J. M.; Baldwin, R. L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 95–119. (d) Kinnear, B. S.; Kaleta, D. T.; Kohtani, M.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **2000**, *122*, 9243–9256. (e) Wei, Y.; Nader, W.; Hansmann, U. H. E. *J. Chem. Phys.* **2007**, *126*, 204307.
- (10) Kaminski, G. A.; Ponomarev, S. Y.; Liu, A. B. *J. Chem. Theory Comput.* **2009**, *5*, 2935–2943.
- (11) (a) *Jaguar*, v3.5, Schrödinger, Inc.: Portland, OR, 1998; (b) *Jaguar*, v4.2, Schrödinger, Inc.: Portland, OR, 2000.
- (12) Kaminski, G. A.; Maple, J. R.; Murphy, R. B.; Braden, D.; Friesner, R. A. *J. Chem. Theory Comput.* **2005**, *1*, 248–254.
- (13) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (14) Takekiyo, T.; Imai, T.; Kato, M.; Taniguchi, Y. *Biopolymers* **2004**, *73*, 283–290.
- (15) Distasio, R. A.; Steele, R. P.; Rhee, Y. M.; Shao, Y.; Head-Gordon, M. *J. Comput. Chem.* **2007**, *28*, 839–856.
- (16) Berndt, K. D. *Protein Secondary Structure*, Birkbeck College, University of London: London; [http://www.cryst.bbk.ac.uk/PPS2/course/section8/ss-960531\\_6.html](http://www.cryst.bbk.ac.uk/PPS2/course/section8/ss-960531_6.html). Accessed on December 10, 2010).